ZOR – Mathematical Methods of Operations Research (1994) 40:145–162

Nonzero-Sum Stochastic Games with Unbounded Costs: Discounted and Average Cost Cases

LINN I. SENNOTT

Department of Mathematics, 4520 Illinois State University, Normal, Illinois 61790-4520 USA

Abstract: We treat non-cooperative stochastic games with countable state space and with finitely many players each having finitely many moves available in a given state. As a function of the current state and move vector, each player incurs a nonnegative cost. Assumptions are given for the expected discounted cost game to have a Nash equilibrium randomized stationary strategy. These conditions hold for bounded costs, thereby generalizing Parthasarathy (1973) and Federgruen (1978). Assumptions are given for the long-run average expected cost game to have a Nash equilibrium randomized stationary strategy, under which each player has constant average cost. A flow control example illustrates the results. This paper complements the treatment of the zero-sum case in Sennott (1993a).

Key Words: Nonzero-sum stochastic games, discounted and average cost stochastic games, flow control queueing models

1 Introduction

Nonzero-sum non-cooperative stochastic games on a countable state space are treated. In each state, each of the K ($<\infty$) players has finitely many moves from which to choose. As a function of the current state and move vector, a nonnegative cost is incurred by each player; the costs may be unbounded above on the state space. The game then transitions to another state, where the transition probabilities are a function of the current state and move vector; termination of the game is not allowed.

The existence of a randomized stationary strategy that is a Nash equilibrium is of interest. Such a strategy allows randomization among the allowable moves, where the randomization is a function only of the given state. A Nash equilibrium strategy has the property that no single player has an incentive to unilaterally deviate from the strategy, given that the other players continue to follow the strategy.

In Section 2 assumptions are given for the existence of a Nash equilibrium in the expected discounted case. These assumptions hold when the costs are

0340–9422/94/40:2/145–162 $2.50

bounded, thus generalizing results of Parthasarathy (1973) and Federgruen (1978). Section 3 illustrates this result with a flow control model.

In Section 4 we treat the long-run average expected cost case and give assumptions that guarantee the existence of a randomized stationary strategy that is a Nash equilibrium. In Section 5, the verification of these assumptions is discussed. As a corollary, a theorem of Rogers (1969) and Sobel (1971) for the finite state case is obtained. In Section 6, the average cost results are applied to the flow control model.

The notion of a (zero-sum) stochastic game was introduced by Shapely (1953). Independently, Fink (1964), Takashashi (1964), Rogers (1969), and Sobel (1971) proved the existence of a randomized stationary Nash equilibrium for finite state discounted stochastic games. For a countable state space, finite move sets, and bounded cost functions, Parthasarathy (1973) proves the existence, while Federgruen (1978) proves the existence in the case of a countable state space with compact metric action spaces and bounded costs. The discounted case is also treated by Sobel (1973) under the assumption that the state and move spaces are compact metric and the set of players may be infinite. Useful survey papers include Parathasarathy and Stern (1977) and Raghavan and Filar (1991).

In the average cost case, it has been proved by Rogers (1969) and Sobel (1971) that if the state space is finite and every stationary strategy is unichain, then a Nash equilibrium randomized stationary strategy exists. Federgruen (1978) has treated the average cost case under the assumption of bounded costs. The approach taken is to examine the relative value functions. The assumptions made are rather strong. Our analysis is also based on the relative value functions and builds on work of Sennott (1989) and (1993b) on the existence of average cost optimal stationary policies in Markov decision chains. Other recent treatments of the average cost case are Borkar and Ghosh (1992), Ghosh and Bagchi (1992), and Nowak (1992).

2 The Discounted Case

Consider a non-cooperative stochastic game with players 1, 2, ..., K and a countable state space. When the game is in state i, player k has a finite nonempty set $A_k(i)$ of moves available. The players each non-cooperatively choose a move, which results in a vector $m = (a_1, a_2, \ldots, a_K)$ of moves, such that $a_k \in A_k(i)$. Given the current state i and move vector m, the game transitions to state j with probability $P_{ij}(m)$, where $\sum_j P_{ij}(m) = 1$.

The cost vector $C(i, m)$ associates with each allowable pair (i, m) a K-tuple of costs; player k incurs nonnegative cost $C_k(i, m)$. Costs may be unbounded above in i.

A strategy for a particular player is a rule for choosing moves that may depend on the history of the game (all states, through the current state, and moves of all palyers), and may be randomized. A strategy is a vector $\theta = (\theta_1, \theta_2, \ldots, \theta_K)$, where θ_k is a strategy for player k.

Fix a number $\alpha \in (0, 1)$. Since the discount factor α will be fixed throughout Sections 2 and 3, we suppress it in our notation. Given initial state i and strategy vector θ, define the expected discounted cost to player k under θ by

$$V_k(i, \theta) = E_\theta\left(\sum_{n=0}^{\infty} \alpha^n (C_k(X_n, M_n) | X_0 = i)\right), \tag{1}$$

where X_n is the state, and M_n the move vector, at time n.

A (pure) stationary strategy is a vector $f = (f_1, f_2, \ldots, f_K)$. If the game is in state i, then player k will choose move $f_k(i) \in A_k(i)$. A randomized stationary strategy is a vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_K)$. If the game is in state i, then player k will choose move $a_k \in A_k(i)$ with probability $\lambda_k(i)(a_k)$. Let $P(A_k(i))$ be the set of all probability distributions on $A_k(i)$. Note that

$$\lambda \in \Lambda = \prod_i \prod_{k=1}^{K} P(A_k(i)), \tag{2}$$

which is a compact metric space. For the duration of the paper, f will refer to a generic stationary policy and λ to a generic randomized stationary policy.

Let θ be a strategy, k a fixed player, and assume that the game is in state i. The notation $\theta \backslash_k a$ indicates that the other players operate under θ, whereas player k chooses $a \in A_k(i)$. The notation may easily become quite cumbersome. Other terms in an expression will usually indicate that player k is the one referred to. Hence, if no misunderstanding can occur, this will be simplified to $\theta \backslash a$.

The quantity $C_k(i, \lambda \backslash a)$ denotes the expected one-step cost to player k, if $\lambda \backslash_k a$ is employed, and it is a convex combination of costs associated with the vectors $m \backslash a = (a_1, \ldots, a_{k-1}, a, a_{k+1}, \ldots, a_K)$. The quantity $P_{ij}(\lambda \backslash_k a)$ denotes a transition probability under $\lambda \backslash_k a$.

The strategy in which all the players except k follow λ, whereas k follows the randomized stationary strategy δ, will be denoted $\lambda \backslash_k \delta$ (or $\lambda \backslash \delta$). The quantities $C_k(i, \lambda \backslash \delta)$ and $P_{ij}(\lambda \backslash_k \delta)$ are defined in the obvious way.

Now assume that every other player follows λ, but that player k is free to follow any strategy. Then the best that k can do is denoted $V_k(i, \lambda \backslash)$. Note that no misunderstanding can result since the subscript tells us that we are dealing with player k. We then have

$$V_k(i, \lambda \backslash) = \inf_\theta V_k(i, \lambda \backslash \theta), \tag{3}$$

where θ refers to any strategy that player k may follow.

If the other players follow λ, then player k is faced with a Markov decision chain. Then $V_k(i, \lambda\backslash)$ is the minimal nonnegative solution of the discount optimality equation

$$V_k(i, \lambda\backslash) = \min_a \left\{ C_k(i, \lambda\backslash a) + \alpha \sum_j P_{ij}(\lambda\backslash a) V_k(j, \lambda\backslash) \right\}$$

$$= \inf_\delta \left\{ C_k(i, \lambda\backslash\delta) + \alpha \sum_j P_{ij}(\lambda\backslash\delta) V_k(j, \lambda\backslash) \right\}, \qquad \text{for all } i \,. \tag{4}$$

Moreover, any stationary policy that realizes the right side of (4) is optimal for player k (Bertsekas (1987)).

The optimization criterion to be used in this paper is a generalization of the concept of a Nash equilibrium.

2.1 Definition: A randomized stationary strategy λ is an α-discounted equilibrium point (α-DEP), for initial state i, if $V_k(i, \lambda\backslash) = V_k(i, \lambda)$, for $1 \leq k \leq K$.

This implies that no single player has an incentive to unilaterally deviate from λ. The following assumptions will guarantee the existence of a randomized stationary α-DEP with finite value for all initial states. The postulated functions in the assumptions are assumed finite. To avoid confusion, superscripts are sometimes used to denote sequences.

Assumption D1: Assume that $\lambda^n \to \lambda$. Then there exist functions $R_k(i)$ such that $V_k(i, \lambda^n\backslash) \leq R_k(i)$, for all i, k, and n. Moreover, $\sum_j P_{ij}(m) R_k(j) < \infty$, for all i, k, and m.

Assumption D2: Assume that there exists a nonnegative function $U_k(i) \leq R_k(i)$ satisfying the discount optimality equation (4) for $\lambda\backslash$ and all k. Then $U_k(i) = V_k(i, \lambda\backslash)$, for all i and k.

If there exists an upper bound C on the components of all the cost vectors, then we may take $R_k(i) \equiv C/(1 - \alpha)$, and Assumptions D1 is satisfied. Within the class of bounded functions, it is known that the solution to the discount optimality equation is unique (Ross (1983)), and hence Assumption D2 also holds. In this case, there exists a randomized stationary α-DEP (Parthasarathy (1973) or Federgruen (1978)).

2.2 Theorem: Assume that Assumptions D1 and D2 hold. Then there exists a randomized stationary strategy that is an α-DEP.

Proof: As in Federgruen (1978), the following theorem is employed: Let S be a compact convex nonvoid subset of a Hausdorff linear locally convex topological space. Let φ be an upper semi-continuous (usc) set-valued map taking $s \in S$ to a closed convex nonvoid subset of S. Then there exists a point $s^* \in \varphi(s^*)$. (The map φ is usc if given $s_n \to s$, $x_n \to x$, such that $x_n \in \varphi(s_n)$ for all n, it follows that $x \in \varphi(s)$.) This theorem was proved independently by Fan (1952) and Glicksberg (1952).

The map φ will be defined on Λ, which is a compact (hence closed) convex subset of

$$\prod_i \prod_{k=1}^{K} R^{|A_k(i)|} \ . \tag{5}$$

This is a product of locally convex linear topological spaces, and hence is a locally convex linear topological space. Given λ, i and k, let $B_k(i, \lambda\backslash)$ be the subset of $A_k(i)$ consisting of moves that realize the minimum on the right of (4). Define

$$\varphi(\lambda) = \prod_i \prod_{k=1}^{K} P(B_k(i, \lambda\backslash)) \ . \tag{6}$$

Clearly $\varphi(\lambda)$ is closed and convex; we show that the map is usc. By p. 124 of Fan (1952), it is sufficient to show that the coordinate maps $\varphi(\lambda)_k$ are usc. So fix a player k and assume that $\lambda^n \to \lambda$ and $\delta^n \to \delta$ such that $\delta^n \in \varphi(\lambda^n)$ for all n. We must show that $\delta_k \in \varphi(\lambda)_k$.

Since $\delta_k^n \in \varphi(\lambda^n)_k$, it follows that

$$\begin{aligned} V_k(i, \lambda^n\backslash) &= C_k(i, \lambda^n\backslash\delta^n) + \alpha \sum_j P_{ij}(\lambda^n\backslash\delta^n) V_k(j, \lambda^n\backslash) \\ &\leq C_k(i, \lambda^n\backslash a) + \alpha \sum_j P_{ij}(\lambda^n\backslash a) V_k(j, \lambda^n\backslash) \ , \qquad \text{for } a \in A_k(i) \ . \end{aligned} \tag{7}$$

It follows from Assumption D1 that $V_k(i, \lambda^n\backslash)$ is a sequence in the compact metric space $\prod_i [0, R_k(i)]$. Hence there exist a nonnegative function $U_k(i) \leq R_k(i)$ and a subsequence n_m such that

$$\lim_{m \to \infty} V_k(i, \lambda^{n_m}\backslash) = U_k(i) \ , \qquad \text{for all } i \ . \tag{8}$$

We now take the limit in (7) through values n_m. Assumption D1 and the dominated convergence theorem yield

$$U_k(i) = C_k(i, \lambda\backslash\delta) + \alpha \sum_j P_{ij}(\lambda\backslash\delta) U_k(j)$$

$$\leq C_k(i, \lambda\backslash a) + \alpha \sum_j P_{ij}(\lambda\backslash a) U_k(j) \ , \qquad \text{for } a \in A_k(i) \ . \tag{9}$$

Hence $U_k(i)$ satisfies the discount optimality equation for $\lambda\backslash$. By Assumption D2, it follows that $U_k(i) = V_k(i, \lambda\backslash)$, and equality is achieved at δ_k.

We must show that $\delta_k(i) \in P(B_k(i, \lambda\backslash))$. From (9), it follows that

$$V_k(i, \lambda\backslash) = \sum_a \delta_k(i)(a) \left\{ C_k(i, \lambda\backslash a) + \alpha \sum_j P_{ij}(\lambda\backslash a) V_k(j, \lambda\backslash) \right\} \ , \quad \text{for all } i \ . \tag{10}$$

Consider the bracketed terms. If $a \in A_k(i) - B_k(i, \lambda\backslash)$, then the term is greater than $V_k(i, \lambda\backslash)$, whereas if $a \in B_k(i, \lambda\backslash)$, then the term is equal to $V_k(i, \lambda\backslash)$. This implies that if $\delta_k(i)(a) > 0$, then we must have $a \in B_k(i, \lambda\backslash)$, and hence $\delta_k(i) \in P(B_k(i, \lambda\backslash))$.

This completes the proof that φ is usc, and hence there exists $\lambda^* \in \varphi(\lambda^*)$. For a player k, this means that $\lambda_k^*(i)$ employs only actions in the optimal set $B_k(i, \lambda^*\backslash)$. In the Appendix, we have proved that any policy with this property (including history dependent and/or randomized policies), is optimal. This implies that $V_k(i, \lambda^*\backslash) = V_k(i, \lambda^*)$, and hence λ^* is an α-DEP.

3 An Example for the Discounted Case

Consider a distributed packet communication system with K sources, each with its own infinite capacity buffer. The sources generate packets independently of each other; the number of packets generated by source k in any slot is governed by a probability distribution (p_j^k), $j \geq 0$. The state of the system is given by a vector $i = (i_1, i_2, \ldots, i_K)$, where i_k is the number of packets currently in buffer k.

The system is served by a single server, and the service time of a packet is one slot. In each slot the server chooses, at random, the buffer to serve from among the nonempty buffers. Observe that on average, each source will receive service at least onece very K slots. If there are empty buffers, then on average, the nonempty ones will receive service at a higher rate.

We may think of the sources as playing a non-cooperative game. Each source may make decision $a = 1$ to allow its generated packets into its buffer, or $a = 0$ to reject (and lose) those packets. Thus for move vector m, it follows that $m_k = 1$, if k chooses to admit packets, and $m_k = 0$, if k chooses to reject packets.

Source k incurs a nonnegative holding cost $H_k(i_k)$ on its current buffer contents and a positive cost D_k for rejecting packets. Hence $C_k(i, m) = H_k(i_k) + D_k(1 - m_k)$. We will assume that $H_k(0_k) = 0$. For $i_k \geq 1$, let $L_k(i_k) = H_k(i_k) + H_k(i_k - 1) + \cdots + H_k(1)$, and let $L_k(0_k) = 0$.

3.1 Proposition: Assume the Following:

i) $\max_k \sup_j \{j | p_j^k > 0\} = J < \infty$.

ii) $\lim_{n\to\infty} \alpha^n L_k(i_k + nJ) = 0$, for all i and k.

Then Assumptions D1 and D2 hold, and hence there exists a randomized stationary strategy that is an α-DEP.

Proof: Let e be the strategy for k that always rejects packets. Then $V_k(i, \lambda\backslash) \leq V_k(i, \lambda\backslash e)$. The latter is bounded above by the cost of emptying buffer k, plus the additional discounted cost of keeping it empty. Since k receives service, on average, at least once every K slots, the expected time in each state of the first passage will not exceed K. Hence

$$V_k(i, \lambda\backslash) \leq K(L_k(i_k) + D_k i_k) + \frac{D_k}{1 - \alpha} =: R_k(i) \ . \tag{11}$$

Since no more than J packets may be generated by any source in a given slot, it follows that the summation is over a finite set, and hence Assumption D1 holds.

To verify Assumption D2, assume that $U_k(i) \leq R_k(i)$ satisfies (4). It is known that $V_k(i, \lambda\backslash) \leq U_k(i)$. Hence it is sufficient to prove the reverse inequality. From reasoning similar to Theorem 2.2 of Ross (1983), it is sufficient to prove that

$$\liminf_{n\to\infty} \alpha^n E_{(\lambda\backslash f)}(R_k(X_n) | X_0 = i) = 0 \ , \tag{12}$$

where f is the discounted optimal stationary policy from (4).

Now $X_n \leq i + nJ$, and hence $E_{(\lambda\backslash f)}(R_k(X_n) | X_0 = i) \leq R_k(i + nJ)$. It then follows from (11) and (ii) that (12) holds.

4 The Average Cost Case

Given strategy vector θ and initial state i, define the average expected cost to player k under θ by

$$j_k(i, \theta) = \limsup_{n\to\infty} \frac{1}{n} E_\theta\left(\sum_{t=0}^{n-1} (C_k(X_t, M_t)|X_0 = i)\right), \tag{13}$$

where X_t is the state, and M_t the move vector, at time t. Given λ, let $j_k(i, \lambda\backslash) = \inf_\theta j_k(i, \lambda\backslash\theta)$, where θ refers to any strategy that k may follow.

4.1 Definition: A randomized stationary strategy λ is an average expected cost equilibrium point (AEP), for initial state i, if $j_k(i, \lambda\backslash) = j_k(i, \lambda)$, for all k.

This implies that no player has an incentive to unilaterally deviate from the strategy λ. The following assumptions will guarantee the existence of a randomized stationary AEP under which each player has constant average cost. In referring to the discounted case, recall that the discount factor was suppressed in the notation. The postulated functions and constants are finite.

Assumption A1: Assume that for each $\alpha \in (0, 1)$, there exists an α-DEP λ_α.

Let $h_{k,\alpha}(i) = V_{k,\alpha}(i, \lambda_\alpha) - V_{k,\alpha}(0, \lambda_\alpha)$, where 0 is a distinguished state. Let $\alpha(n)$ be a sequence of discount factors converging to 1 such that $\lambda_{\alpha(n)} \to \eta$. (Recall that randomized stationary strategies lie in the compact metric space Λ. Hence any sequence of α-DEP has a convergent subsequence.)

Assumption A2: There exist nonnegative functions $M_k(i)$ such that $h_{k,\alpha(n)}(i) \leq M_k(i)$, for all i, k, and n. Moreover, $\sum_j P_{ij}(m)M_k(j) < \infty$, for all i, m, and k.

Assumption A3: There exist nonnegative functions $L_k(i)$ such that $-L_k(i) \leq h_{k,\alpha(n)}(i)$, for all i, k, and n. Moreover, $\sum_j P_{ij}(m)L_k(j) < \infty$, for all i, m, and k.

Assumption A4: Assume that for all k, there exist a constant j_k and function $h_k(i)$, with $-L_k(i) \leq h_k(i) \leq M_k(i)$, satisfying the average cost optimality equation

$$\begin{aligned} j_k + h_k(i) &= C_k(i, \eta) + \sum_j P_{ij}(\eta)h_k(j) \\ &= \min_a \left\{C_k(i, \eta\backslash a) + \sum_j P_{ij}(\eta\backslash a)h_k(j)\right\}, \qquad \text{for all } i . \end{aligned} \tag{14}$$

Then $j_k \equiv j_k(i, \eta) = j_k(i, \eta\backslash)$, for all i and k.

The major result may now be stated.

4.2 Theorem: Assume that Assumptions A1–A4 hold. Then the randomized stationary strategy η is an AEP. For player k, the value of the average expected cost under η is j_k.

Proof: We may write (4) as

$$(1-\alpha)V_{k,\alpha}(0,\lambda_\alpha)+h_{k,\alpha}(i)=C_k(i,\lambda_\alpha)+\alpha\sum_j P_{ij}(\lambda_\alpha)h_{k,\alpha}(j)$$

$$=\min_a\left\{C_k(i,\lambda_\alpha\backslash a)+\alpha\sum_j P_{ij}(\lambda_\alpha\backslash a)h_{k,\alpha}(j)\right\}. \tag{15}$$

Let $i=0$ in (15). It follows easily from Assumption A2 that $(1-\alpha(n))V_{k,\alpha(n)}(0,\lambda_{\alpha(n)})$ is bounded. Hence there exist a subsequence of $\alpha(n)$ (call it $\beta(n)$ for notational convenience) and numbers j_k such that

$$\lim_{n\to\infty}(1-\beta(n))V_{k,\beta(n)}(0,\lambda_{\beta(n)})=j_k\ , \qquad \text{for } 1\le k\le K\ . \tag{16}$$

Now $h_{k,\beta(n)}(i)$ is a sequence in the compact metric space $\prod_k\prod_i[-L_k(i),M_k(i)]$. Hence there exist a subsequence of $\beta(n)$ (call it $\varepsilon(n)$) and functions $h_k(i)$ such that

$$\lim_{n\to\infty}h_{k,\varepsilon(n)}(i)=h_k(i)\ , \qquad \text{for all } i \text{ and } k\ . \tag{17}$$

Now take the limit in (15) through values $\alpha=\varepsilon(n)$. Using (16), (17), and the dominated convergence theorem yields (14). It then follows immediately from Assumption A4 that η is an AEP with constant average value j_k for player k.

4.3 Corollary: Assume that the costs are bounded and that $|h_{k,\alpha}(i)|\le N$, for all i, k, and α. Then Assumptions A2–A4 hold for all α, and hence any limit point of α-DEP is an AEP.

Proof: If the costs are bounded, then Assumption A1 holds. Clearly, Assumptions A2 and A3 hold. Finally, it follows as in Ross (1983, p. 93) that Assumption A4 holds.

The requirement that $h_{k,\alpha}(i)$ be uniformly bounded is very strong when the state space is denumerable.

5 Verification of the Assumptions for the Average Cost Case

The major result of this section is a set of recurrence-type conditions that are sufficient for the validity of Assumptions A1–A4. We first make some general comments. Any strategy λ induces a Markov chain on the state space, and $m_{ij}(\lambda)$ is the expected time of a first passage from i to j, under λ. If $m_{i0}(\lambda) < \infty$, for all i, then this implies that λ induces a chain with a single positive recurrent class $R(\lambda)$ containing 0. Moreover, if the chain begins in $i \notin R(\lambda)$, then it will reach $R(\lambda)$ in finite expected time.

Assume that there is a cost $E(i)$ associated with state i, such that the expected cost $e_{i0}(\lambda)$ of a first passage is finite, for all i. If the process starts in state i, then by the delayed renewal reward theorem, the expected average E cost is obtained as a limit and equals $\sum_{j \in R(\lambda)} E(j)\pi_j(\lambda) = e_{00}(\lambda)/m_{00}(\lambda)$, where $(\pi_i(\lambda))_{i \in R(\lambda)}$ is the steady-state distribution on $R(\lambda)$.

5.1 Proposition: Assume that the following conditions hold, for all i and λ:

i) There exists a function $B(i)$ such that $m_{i0}(\lambda) \leq B(i)$.
ii) The B cost $b_{i0}(\lambda)$ of a first passage is finite.
iii) There exist functions $M_k(i)$ such that $c_{k,i0}(\lambda) \leq M_k(i)$, for all k.
iv) The M_k cost $d_{k,i0}(\lambda)$ of a first passage is finite, for all k.

Then Assumptions A1–A4 hold for all $\alpha \in (0, 1)$. Hence any limit point of α-DEP is an AEP.

Proof: We verify that Assumption D1 holds. Fix α, k, and λ and let f be a discounted optimal stationary policy for $\lambda\backslash$. Then for $i \neq 0$, a standard argument gives

$$V_{k,\alpha}(i, \lambda\backslash) \leq c_{k,i0}(\lambda\backslash f) + V_{k,\alpha}(0, \lambda\backslash) \leq M_k(i) + V_{k,\alpha}(0, \lambda\backslash) \ . \tag{18}$$

(See p. 97 of Ross (1983) or p. 19 of Sennott (1986)).

From (4), it follows that

$$V_{k,\alpha}(0, \lambda\backslash) = C_k(0, \lambda\backslash f) + \alpha \sum_j P_{0j}(\lambda\backslash f) V_{k,\alpha}(j, \lambda\backslash) \ , \tag{19}$$

and using the first inequality in (18) yields

$$(1 - \alpha) V_{k,\alpha}(0, \lambda\backslash) \leq C_k(0, \lambda\backslash f) + \sum_{j \neq 0} P_{0j}(\lambda\backslash f) c_{k,j0}(\lambda\backslash f) \ . \tag{20}$$

The quantity on the right of (20) equals $c_{k,00}(\lambda \backslash f)$, which is bounded above by $M_k(0)$. We then define $R_{k,\alpha}(i) = M_k(i) + M_k(0)/(1-\alpha)$, for $i \neq 0$, and $R_{k,\alpha}(0) = M_k(0)/(1-\alpha)$.

We now verify the second statement of Assumption D1. Fix i, k, and m, and let e be a stationary strategy that chooses m in state i. Then

$$d_{k,i0}(e) = M_k(i) + \sum_{j \neq 0} P_{ij}(m) d_{k,j0}(e) \geq \sum_{j \neq 0} P_{ij}(m) M_k(j) \ . \tag{21}$$

By (iv), the left side of (21) is finite, hence so is the right side.

To verify that Assumption D2 holds, it is sufficient to verify (12) for $M_k(\cdot)$. Now

$$\liminf_{n \to \infty} E_{(\lambda \backslash f)}(M_k(X_n) | X_0 = i) \leq \lim_{n \to \infty} \left(\frac{1}{n} E_{(\lambda \backslash f)} \left(\sum_{t=0}^{n-1} M_k(X_t) | X_0 = i \right) \right) , \tag{22}$$

where the limit exists, and equals $\sum \pi_j(\lambda \backslash f) M_k(j) < \infty$, by the delayed renewal reward theorem. From this we see that (12) holds.

It follows from Theorem 2.2 that there exists an α-DEP λ_α, for $\alpha \in (0, 1)$, and hence Assumption A1 holds. It follows from (18) and (21) that Assumption A2 holds.

For $i \neq 0$, let T be the first passage time from i to 0, under λ_α. By a standard argument (p. 97–98 of Ross (1983)), we have

$$V_{k,\alpha}(i, \lambda_\alpha) \geq E_{\lambda_\alpha}(\alpha^T) V_{k,\alpha}(0, \lambda_\alpha) \ , \tag{23}$$

and hence

$$h_{k,\alpha}(i) \geq -E_{\lambda_\alpha} \left(\frac{1 - \alpha^T}{1 - \alpha} \right) (1 - \alpha) V_{k,\alpha}(0, \lambda_\alpha) \ . \tag{24}$$

Since $(1 - \alpha^T)/(1 - \alpha) \uparrow T$, it follows that the first term is bounded above by $B(i)$. From (20) it follows that the second term is bounded above by $M_k(0)$. Hence $h_{k,\alpha}(i) \geq -M_k(0) B(i)$, and we may define $L_k(i) =: M_k(0) B(i)$. Set $L_k(0) = 0$. The verification that $\sum_j P_{ij}(m) B(j) < \infty$ is similar to the reasoning for Assumption D1 and uses (ii). This proves that Assumption A3 holds.

We will now verify that Assumption A4 holds. First consider an arbitrary randomized stationary strategy δ. Similarly to (22), we have, for all i

$$\liminf_{n \to \infty} E_\delta(B(X_n) | X_0 = i) \leq \lim_{n \to \infty} \left(\frac{1}{n} E_\delta \left(\sum_{t=0}^{n-1} B(X_t) | X_0 = i \right) \right) , \tag{25}$$

and by (ii) and the delayed renewal reward theorem, it follows that the limit on the right exists and equals the finite constant average B-cost. This implies that the following two statements hold:

$$E_\delta(B(X_n)|X_0 = i) < \infty \ , \qquad \text{for all } n \ ; \tag{26}$$

$$\liminf_{n\to\infty} \frac{1}{n} E_\delta(B(X_n)|X_0 = i) = 0 \ . \tag{27}$$

Reasoning similarly, we have, for all i and k,

$$E_\delta(M_k(X_n)|X_0 = i) < \infty \ , \qquad \text{for all } n \ ; \tag{28}$$

$$\liminf_{n\to\infty} \frac{1}{n} E_\delta(M_k(X_n)|X_0 = i) = 0 \ . \tag{29}$$

Now assume that (14) holds and consider the first equality. This may be iterated as in Lemma A1 of Sennott (1989), where we use (26) to guarantee that the expectation of $h_k(\cdot)$ is never $-\infty$. For an initial state i, this yields

$$\begin{aligned} \frac{1}{n} E_\eta\left[\sum_{t=0}^{n-1} C_k(X_t, M_t)\right] &= j_k + \frac{h_k(i)}{n} - \frac{1}{n} E_\eta[h_k(X_n)] \\ &\le j_k + \frac{h_k(i)}{n} + \frac{1}{n} E_\eta[L_k(X_n)] \ . \end{aligned} \tag{30}$$

Again by the delayed renewal reward theorem, the limit of the left side exists. Taking the limit infimum of both sides, and using (27), yields $j_k(i, \eta) \le j_k$.

If the other players play η, but player k is free to play any strategy, then player k faces an average cost minimization Markov decision chain. We claim that there exists an average cost optimal stationary policy for this problem. This may be shown by verifying that the hypotheses of Proposition 3.1 of Sennott (1993b) hold, for all strategies $\eta \backslash e$. These conditions follow immediately from our assumptions, and hence there exists a stationary policy f for k such that $j_k(\eta \backslash f) \le j_k(i, \eta \backslash \theta)$, for any strategy θ for k and for all i.

Therefore to verify that Assumption A4 holds, it is sufficient to prove that $j_k \le j_k(\eta \backslash f)$. From (14), it follows that

$$j_k + h_k(i) \le C_k(i, \eta \backslash f) + \sum_j P_{ij}(\eta \backslash f) h_k(j) \ , \qquad \text{for all } i \ . \tag{31}$$

One may use (26) and (28) to show that $-\infty < E_{(\eta\setminus f)}(h_k(X_t)|X_0 = i) < \infty$, for all t. Iterating (31) yields, for initial state i,

$$\frac{1}{n}E_{(\eta\setminus f)}\left[\sum_{t=0}^{n-1} C_k(X_t, M_t)\right] \geq j_k + \frac{h_k(i)}{n} - \frac{1}{n}E_{(\eta\setminus f)}[h_k(X_n)]$$

$$\geq j_k + \frac{h_k(i)}{n} - \frac{1}{n}E_{(\eta\setminus f)}[M_k(X_n)] \ . \tag{32}$$

Taking the limit supremum of both sides, the desired result follows from (29).

5.2 Corollary: Assume that the costs are bounded and that (i) and (ii) of Proposition 5.1 hold. Then Assumptions A1–A4 hold for all $\alpha \in (0, 1)$.

Proof: Assume that $C_k(i, m) \leq C$, for all i, k, and m. We will verify that (iii) and (iv) of Proposition 5.1 hold. By (i), it follows that $c_{k,i0}(\lambda) \leq Cm_{i0}(\lambda) \leq CB(i)$, and hence (iii) holds. Condition (iv) follows from (ii).

The result may be used to prove a result of Rogers (1969) and Sobel (1971) (see also Federgruen (1978) and Stern (1975)).

5.3 Corollary: Assume that the state space is finite. Assume that given any stationary strategy f and initial state $i \neq 0$, there is an f path from i to 0. Then Assumptions A1–A4 hold for all $\alpha \in (0, 1)$.

Proof: It is well known that the assumption implies that $m_{i0}(f) < \infty$, for all i. Let $B =: \max_{i\neq 0} \max_f m_{i0}(f)$. Since there are finitely many states and finitely many stationary policies, B is well-defined and finite. Moreover, $m_{00}(f) = 1 + \sum_{j\neq 0} P_{0j}(f)m_{j0}(f) \leq 1 + B$. We will show that these bounds also hold for randomized stationary policies. By Corollary 5.2, this will complete the proof.

Given any λ and $i \neq 0$, we first show that there is a λ path from i to 0. For each $j \neq 0$, select a move vector m_j such that $\lambda(j)(m_j) > 0$. Select m_0 arbitrarily. Define $f(j) = m_j$. By assumption, there exists n such that $P_{i0}^{(n)}(f) > 0$. Then it is easy to see that $P_{i0}^{(n)}(\lambda) > 0$.

This implies that $m_{i0}(\lambda) < \infty$, for all i. Now suppose the claimed result is true for $i \neq 0$. Since $m_{00}(\lambda) = 1 + \sum_m \lambda(0)(m) \sum_{j\neq 0} P_{0j}(m)m_{j0}(\lambda) \leq 1 + B$, we are finished. Now define $r(0) = 0$, and for $i \neq 0$, define $r(i) = m_{i0}(\lambda)$. For each $i \neq 0$, select a move vector m_i^* that realizes $\max_m \sum_j P_{ij}(m)r(j)$. Select m_0^* arbitrarily. This defines a stationary strategy f with the property that $r(i) \leq 1 + \sum_j P_{ij}(f)r(j)$, for all $i \neq 0$. A simple modification of the proof in Seneta and Tweedie (1985, p. 150) proves that $r(i) \leq m_{i0}(f)$, and this completes the proof.

6 An Example for the Average Cost Case

This is the example of Section 3, with the following changes. When the system is in state 0 (i.e. all components are 0), then each source must accept new packets. Moreover, we assume that $\sup\{j|p_j^k > 0\} =: J_k \leq \infty$. The costs are the same as in Section 3. Let ω_k be the mean of the packet arrivial distribution for source k, and let $\omega_k^{(n)}$ be its nth moment.

6.1 Proposition: Assume that the following conditions hold, for all k:

i) We have $p_0^k > 0$ and $\omega_k < 1/K$.
ii) The holding cost $H_k(i_k)$ is bounded above by a polynomial of degree n (≥ 0), and $\omega_k^{(n+2)} < \infty$.

Then Assumptions A1–A4 hold, and hence there exists a randomized stationary strategy that is an AEP.

We will verify that the hypotheses of Proposition 5.1 hold. We first set up some notation and prove some lemmas. Let $j = (j_1, \ldots, j_K)$ denote the numbers of newly generated packets in each slot, and let $Q(j)$ be the probability of vector j being generated. If move vector m is chosen, then the dot product $j \bullet m$ is the total number of generated packets that are accepted into the system. The notation jm denotes the vector with kth component equal to $j_k m_k$.

If $i \neq 0$, then $|i|$ denotes the number of nonzero coordinates of i. The notation e denotes a vector with 1 in a nonzero coordinate of i, and 0 elsewhere; there are $|i|$ of these vectors. If the system is operating under λ then the next state will be $i + jm - e$ with probability $|i|^{-1}\lambda(i)(m)Q(j)$.

For the lemmas, we assume that the hypotheses of 6.1 hold, even though not every hypothesis is required in each lemma.

6.2 Lemma: Let $\varepsilon =: 1 - \sum_k \omega_k$. Then (i) of 5.1 holds with $B(i) = (\sum_k i_k)\varepsilon^{-1}$, for $i \neq 0$ and $B(0) = \varepsilon^{-1}$.

Proof: It is easily seen that any λ induces a Markov chain with a communicating class containing $\prod_k [0_k, J_k]$. The policy that always rejects will have this as its communicating class and other policies may have a larger class.

We apply the result on p. 752 of Tweedie (1976) with test function $y(i) = \sum_k i_k$. Let i be a state with at least one nonzero component. Then

$$\sum_r P_{ir}(\lambda)y(r) - y(i) = \sum_m \lambda(i)(m) \sum_j Q(j)(j \bullet m) - 1$$

$$\leq \sum_m \lambda(i)(m) \sum_j Q(j)\left(\sum_k j_k\right) - 1 = \sum_k \omega_k - 1 = -\varepsilon \ . \quad (33)$$

It then follows that $m_{i0}(\lambda) \leq (\sum_k i_k)\varepsilon^{-1}$, for $i \neq 0$. Moreover, $m_{00}(\lambda) = 1 + \sum_{i \neq 0} Q(i)m_{i0}(\lambda) \leq 1 + (1 - \varepsilon)\varepsilon^{-1} = \varepsilon^{-1}$.

6.3 Lemma: Fix λ and an integer $q \geq 2$. Then $\sum_r P_{0r}(\lambda)r_k^q = \omega_k^{(q)}$, and for $i \neq 0$,

$$\sum_r P_{ir}(\lambda)r_k^q \leq i_k^q + q\left(\omega_k - \frac{1}{K}\right)i_k^{q-1} + q^2(\omega_k^{(q)} + 1)(i_k + 1)^{q-2} \ . \quad (34)$$

Proof: The first statement is clear. Now let $i \neq 0$. The next state is $i + jm - e$ with probability $|i|^{-1}\lambda(i)(m)Q(j)$, and hence

$$\sum_r P_{ir}(\lambda)r_k^q = \sum_{u=0}^{q} \binom{q}{u} i_k^u \left[\sum_e \frac{1}{|i|} \sum_m \lambda(i)(m) \sum_j Q(j)(j_k m_k - e_k)^{q-u}\right] . \quad (35)$$

Fix $0 \leq u \leq q - 2$, and consider $(j_k m_k - e_k)^{q-u}$. If $j_k = 0$, then this expression is bounded above by 1, whereas if $j_k \geq 1$, it is bounded above by j_k^q. Hence in either case, it is bounded above by $j_k^q + 1$, and thus the expression in brackets is bounded above by $\omega_k^{(q)} + 1$. For $u = q - 1$, the expression in brackets is bounded above by $\omega_k - |i|^{-1}(\sum_e e_k)$. If $i_k \neq 0$, this is $\omega_k - |i|^{-1} \leq \omega_k - K^{-1}$. Hence

$$\sum_r P_{ir}(\lambda)r_k^q \leq i_k^q + q\left(\omega_k - \frac{1}{K}\right)i_k^{q-1} + \left(\omega_k^{(q)} + 1\right)\sum_{u=0}^{q-2} \binom{q}{u} i_k^u$$

$$= i_k^q + q\left(\omega_k - \frac{1}{K}\right)i_k^{q-1} + (\omega_k^{(q)} + 1)\sum_{u=0}^{q-2} \binom{q-2}{u} i_k^u \left[\frac{\binom{q}{u}}{\binom{q-2}{u}}\right] . \quad (36)$$

Since the quantity in brackets is bounded above by q^2, the result follows.

6.4 Lemma: Fix k and let $F(i_k)$ be a nonnegative polynomial of degree q, where $1 \leq q \leq n+1$. Let $d_{i0}(\lambda)$ be the expected F cost of a first passage. Then there exist constants A and A^* such that $d_{i0}(\lambda) \leq Ai_k^{q+1} + A^*(\sum_s i_s)\varepsilon^{-1}$, for $i \neq 0$ and all λ. Finally, $d_{00}(\lambda) \leq F(0_k) + A\omega_k^{(q+1)} + A^*\varepsilon^{-1}(1-\varepsilon) < \infty$.

Proof: For $i \neq 0$, we set up a test function $y(i) = Ai_k^{q+1}$, where A is to be specified. It follows from (34) that

$$\sum_r P_{ir}(\lambda)y(r) - y(i)$$

$$\leq A(q+1)\left(\left(\omega_k - \frac{1}{K}\right)i_k^q + (q+1)(\omega_k^{(q+1)}+1)(i_k+1)^{q-1}\right). \tag{37}$$

Let $S(i_k)$ denote the right side of (37), and consider $S(i_k) + F(i_k)$. This is a polynomial of degree q, and A may be chosen so that the leading coefficient is negative. Then there exists a number N^* such that $i_k > N^*$ implies that $\sum_r P_{ir}(\lambda)y(r) - y(i) \leq -F(i_k)$.

For $0 \leq i_k \leq N^*$, it follows easily from (34) that $\sum_r P_{ir}(\lambda)y(r)$ is bounded. Let G be an upper bound. Let F be the maximum value of $F(i_k)$, for $0 \leq i_k \leq N^*$. It follows from the proof of Proposition 4 of Sennott (1989) that $d_{i0}(\lambda) \leq Ai_k^{q+1} + (G+F)m_{i0}(\lambda)$. Then the first statement follows from 6.2. The second statement follows using the standard first passage equation for 0. By assumption $\omega_k^{(q+1)} < \infty$.

Proof of 6.1: Part (i) of 5.1 follows from 6.2. To verify (ii), apply 6.4 with $F(i_k) = i_k$. To verify (iii), observe that the cost to k in state i is bounded above by $H_k(i_k) + D_k \leq F(i_k)$, some polynomial of degree n. Hence by 6.4, there exist constants A and A^* such that $c_{k,i0}(\lambda) \leq Ai_k^{n+1} + A^*(\sum i_s)\varepsilon^{-1} =: M_k(i)$, for $i \neq 0$, and $c_{k,00}(\lambda) \leq F(0_k) + A\omega_k^{(n+1)} + A^*\varepsilon^{-1}(1-\varepsilon) =: M_k(0)$.

It remains to verify that (iv) holds. This may be seen by an application of 6.4 to the function i_k^{n+1}.

Appendix

Assume that we have a Markov decision chain with countable state space, finite nonempty action sets, and nonnegative costs. The discount factor $\alpha \in (0, 1)$ is assumed fixed and will be suppressed in our notation.

The value function $V(i)$ is the minimal nonnegative solution of the discount optimality equation

$$V(i) = \min_{a \in A(i)} \left\{ C(i, a) + \alpha \sum_j P_{ij}(a) V(j) \right\}, \qquad \text{for all } i \,. \tag{A1}$$

Let $B(i) = \{a \in A(i) | a \text{ realizes the minimum on the right of (A1)}\}$. Any stationary policy f with the property that $f(i) \in B(i)$, for all i, is α-discount optimal (Bertsekas (1987)).

Now let θ be any policy that restricts itself to actions in $B(i)$, for all i. That is, θ may be history dependent and/or randomized, but when the process is in state i, then θ only employs actions in the set $B(i)$.

Proposition: The policy θ is α-discount optimal.

Proof: Define $V_0(i, \theta) \equiv 0$, and for $n \geq 1$, let $V_n(i, \theta)$ be the expected n-step discounted cost under θ. It will be proved by induction that for any policy θ defined as above, it follows that $V_n(i, \theta) \leq V(i)$, for all i and $n \geq 0$. This is clearly true for $n = 0$.

Now assume that the claim holds for $n - 1$. If the process begins in state i, let $v(i)(a)$ be the initial probability that θ chooses action $a \in B(i)$, and let $\theta(i, j)$ be the policy that applies, if the process begins in state i and then transitions to state j. Then

$$\begin{aligned} V_n(i, \theta) &= \sum_{a \in B(i)} v(i)(a) \left\{ C(i, a) + \alpha \sum_j P_{ij}(a) V_{n-1}(j, \theta(i, j)) \right\} \\ &\leq \sum_{a \in B(i)} v(i)(a) \left\{ C(i, a) + \alpha \sum_j P_{ij}(a) V(j) \right\} = V(i) \,, \qquad \text{for all } i \,. \end{aligned} \tag{A2}$$

The second line follows from the induction hypothesis, and the final equality follows since every term in the brackets is equal to $V(i)$.

Since $\lim_{n \to \infty} V_n(i, \theta) = V(i, \theta)$, it follows that $V(i, \theta) \leq V(i)$, and hence that $V(i, \theta) = V(i)$.

References

Bertsekas DP (1987) Dynamic programming: Deterministic and stochastic models. Prentice-Hall Englewood Cliffs New Jersey

Borkar VS, Ghosh MK (1992) Denumerable state stochastic games with limiting average payoff. Preprint
Fan K (1952) Fixed-point and minimax theorems in locally convex topological linear spaces. Proc Nat Acad Sci USA 38:121–126
Federgruen A (1978) On n-person stochastic games with denumerable state space. Adv Appl Probab 10:452–471
Fink AM (1964) Equilibrium in a stochastic N-person game. J Sci Hiroshima Univ, Series A-I 28:89–93
Glicksberg II (1952) A further generalization of the Kakutani fixed-point theorem, with application to Nash equilibrium points. Proc Amer Math Soc 3:170–174
Ghosh MK, Bagchi A (1992) Stochastic games with average payoff criterion. Preprint
Nowak AS (1992) Stationary equilibria for nonzero-sum average payoff ergodic stochastic games with general state space. Preprint
Parthasarathy T (1973) Discounted, positive, and noncooperative stochastic games. Int J Game Theory 2:25–37
Parathasarathy T, Stern M (1977) Markov games – a survey. Differential Games and Control Theory II, Roxin EO, Liu PI, Sternberg RL (Eds) Dekker 1–46
Raghavan TES, Filar JA (1991) Algorithms for stochastic games – a survey. Zeitschrift fur Oper Res 35:437–472
Rogers PD (1969) Nonzero-sum stochastic games. PhD Thesis Univ of California at Berkeley Berkeley California
Ross SM (1983) Introduction to stochastic dynamic programming. Academic Press New York
Seneta E, Tweedie RL (1985) Moments for stationary and quasi-stationary distributions of Markov chains. J Appl Probab 22:148–155
Sennott LI (1986) A new condition for the existence of optimal stationary policies in average cost Markov decision prodesses. Oper Res Lett 5:17–23
Sennott LI (1989) Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs. Oper Res 37:626–633
Sennott LI (1994) Zero-sum stochastic games with unbounded costs: Discounted and average cost cases. ZOR 39:209–225
Sennott LI (1993b) Another set of conditions for average optimality in Markov control processes. To appear, Systems and Control Letters 23
Shapley LS (1953) Stochastic games. Proc Nat Acad Sci USA 39:1095–1100
Sobel MJ (1971) Non-cooperative stochastic games. Ann Math Stat 42:1930–1935
Sobel MJ (1973) Continuous stochastic games. J Appl Probab 10:597–604
Stern MA (1975) On stochastic games with limiting average payoff. Thesis Univ of Illinois Chicago
Takashashi M (1964) Equilibrium points of stochastic non-cooperative n-person games. J Sci Hiroshima Univ Series A-I, 28:95–99
Tweedie RL (1976) Criteria for classifying general Markov chains. Adv Appl Probab 8:737–771

Received: February 1993
Revised version received: September 1993