W. POŁOWCZUK (Wrocław)

NONZERO-SUM SEMI-MARKOV GAMES WITH
COUNTABLE STATE SPACES

Abstract. We consider nonzero-sum semi-Markov games with a countable state space and compact metric action spaces. We assume that the payoff, mean holding time and transition probability functions are continuous on the action spaces. The main results concern the existence of Nash equilibria for nonzero-sum discounted semi-Markov games and a class of ergodic semi-Markov games with the expected average payoff criterion.

1. Introduction. Nonzero-sum Markov games with the expected average payoff criterion and countably many states were first studied by Federgruen [6] under a uniform ergodicity assumption. His results were generalized to Markov games with unbounded daily payoff functions satisfying some weaker ergodicity or recurrence conditions [3, 5, 14, 20]. Lal and Sinha [11] considered a semi-Markov generalization of Federgruen's model [6]. However, they solved two-person zero-sum games only. They proved the existence of a value and optimal stationary strategies for the players by the "vanishing discount factor approach".

In this paper, we generalize the results by Lal and Sinha [11] to n-person nonzero-sum semi-Markov games with countable state spaces. We adopt the same uniform geometric ergodicity assumption as in [6] and [11]. Our paper is also a generalization of Federgruen's work [6]. We apply a more direct argument based on the optimality equation for corresponding semi-Markov control chains. A similar method was used in the proof of the existence of a stationary Nash equilibrium for Markov games with the expected average payoff criterion in [14]. For a broad discussion of many results concerning nonzero-sum stochastic (mainly Markov) games with finite or Borel state

space the reader is referred to [15]. Some results on correlated equilibria in nonzero-sum semi-Markov games with general state spaces are stated in [13]. Semi-Markov decision models and stochastic games have recently been applied by many authors to studying queueing networks and replacement problems [1, 2, 3, 10, 16, 19, 20].

An *N-person nonzero-sum semi-Markov game* is defined by the following objects:

(i) $\mathbf{S}$ is a countable *state space* for the game.

(ii) $X_k(s)$ is a nonempty compact metric *space of actions* for player k at state $s \in \mathbf{S}$. We put $X(s) = X_1(s) \times \ldots \times X_N(s)$. Let A_k be the union of the sets $X_k(s)$ over all $s \in \mathbf{S}$, and

$$\mathrm{D} := \{(s, x) : s \in \mathbf{S} \text{ and } x \in X(s)\}.$$

(iii) $r_k : \mathrm{D} \to \mathbb{R}$ is a *bounded payoff function* for player k such that $r_k(s, \cdot)$ is *continuous* on $X(s)$ for each $s \in \mathbf{S}$.

(iv) q is a transition probability from D to $\mathbf{S}$, called the *law of motion* among states. If s is a state at some stage of the game and the players select an $x \in X$, then $q(\cdot \,|\, s, x)$ is the probability distribution of the next state of the game. It is assumed that $q(s' \,|\, s, \cdot)$ is *continuous* on $X(s)$ for every s, $s' \in \mathbf{S}$.

(v) $F(t \,|\, s, s', x)$ is a Borel measurable distribution function of the transition time from s to s' when an action vector $x \in X(s)$ is selected. For any $s \in \mathbf{S}$ and $x \in X(s)$, the *mean holding time* in state s is

$$\tau(s, x) = \sum_{s' \in \mathbf{S}} \int_0^\infty t \, F(dt \,|\, s, s', x) q(s' \,|\, s, x).$$

In this paper, we make the following assumptions:

C1. For each $s \in \mathbf{S}$ the function $\tau(s, \cdot)$ is continuous on $X(s)$.
C2. (i) There exist $\delta > 0$ and $\varepsilon > 0$ such that

$$1 - \sum_{s' \in \mathbf{S}} F(\delta \,|\, s, s', x) q(s' \,|\, s, x) \geq \varepsilon \quad \text{for every } (s, x) \in \mathrm{D}.$$

(ii) There exists $M > 0$ such that $\tau(s, x) \leq M$ for every $(s, x) \in \mathrm{D}$.

C2 is a natural assumption. It assures that there are only finitely many transitions in any bounded time interval. From **C2** it follows that the mean holding time in state s is bounded:

$$\exists \varepsilon > 0 \; \exists M > 0 \; \forall (s, x) \in \mathrm{D} \quad \varepsilon \leq \tau(s, \pi) \leq M.$$

The game proceeds over the infinite future with past history as common knowledge for all the players. If s is a state at some decision epoch in the game and the players (independently of one another) select (possibly at random) some $x \in X(s)$, then player k receives a reward $r_k(s, x)$ and a new

state s' for the game is generated according to the probability distribution $q(s' \,|\, s, x)$. Conditionally on the next state s', the time until the transition from s to s' actually occurs is a random variable having the distribution function $F(t \,|\, s, s', x)$. The information available to all the players before their nth choice of actions is the history vector $h^n = (s^1, x^1, \ldots, s^{n-1}, x^{n-1}, s^n)$ where $s^i \in \mathbf{S}$ and $x^i \in X(s^i)$. We denote the set of such vectors by H^n and assume that H^n is endowed with the product σ-algebra.

A *strategy* for player k is a sequence $\pi_k = (\pi_k^1, \pi_k^2, \ldots)$, where every π_k^n is a transition probability from H^n into A_k such that $\pi_k(X_k(s^n) \,|\, h^n) = 1$. A *stationary strategy* for player k is a strategy $\pi_k = (\pi_k^1, \pi_k^2, \ldots)$ such that each π_k^n depends on the current state s^n only. In other words, a strategy π_k of player k is called stationary if there exists a transition probability f_k from $\mathbf{S}$ into A_k such that for every decision epoch n and each history $h^n \in H^n$, we have $\pi_k^n(\cdot \,|\, h^n) = f_k(\cdot \,|\, s^n)$. We let Π_k (resp. F_k) denote the set of all strategies (resp. stationary strategies) for player k.

Let $F = F_1 \times \ldots \times F_N$ be the set of all stationary multi-strategies for the players. We introduce some helpful notation. For any $f = (f_1, \ldots, f_N) \in F$ and any *bounded* function $u : \mathrm{D} \to \mathbb{R}$, define

$$u(s, f) = \int_{X_1(s)} \ldots \int_{X_N(s)} u(s, x_1, \ldots, x_N) \, f_1(dx_1 \,|\, s) \ldots f_N(dx_N \,|\, s),$$

and, for any set $B \subset \mathbf{S}$, put

$$q(B \,|\, s, f) = \int_{X_1(s)} \ldots \int_{X_N(s)} q(B \,|\, s, x_1, \ldots, x_N) \, f_1(dx_1 \,|\, s) \ldots f_N(dx_N \,|\, s).$$

2. Semi-Markov games with limiting average payoffs. Let $H = \mathbf{S} \times X \times \mathbf{S} \times \ldots$ be the space of all infinite histories of the game, endowed with the product σ-algebra. For any multi-strategy $\pi = (\pi_1, \ldots, \pi_N)$ and every initial state $s^1 = s \in \mathbf{S}$, a probability measure P_s^π and a stochastic process $\{\sigma_n, \alpha_n\}$ are defined on H in a canonical way, where the random variables σ_n and α_n describe the state and the actions chosen by the players, respectively, at the nth decision epoch (cf. Chapter 7 of [4]). Thus, for each multi-strategy $\pi = (\pi_1, \ldots, \pi_N)$ and every initial state $s \in \mathbf{S}$, the expected *average payoff* to player k is

$$J_k(s, \pi) = \liminf_{n \to \infty} \frac{E_s^\pi \left(\sum_{i=1}^n r_k(\sigma_i, \alpha_i) \right)}{E_s^\pi \left(\sum_{i=1}^n \tau(\sigma_i, \alpha_i) \right)}.$$

Here E_s^π stands for the expectation operator with respect to the probability measure P_s^π.

For any multi-strategy π and any strategy $\varphi \in \Pi_k$ let (π_{-k}, φ) be the multi-strategy π with π_k replaced by φ.

DEFINITION 1. A multi-strategy $f^* \in F$ is a *stationary Nash equilibrium* for the semi-Markov game with the expected average payoff criterion if for every player k, $\pi_k \in \Pi_k$ and $s \in \mathbf{S}$, we have

$$J_k(s, f^*) \geq J_k(s, (f^*_{-k}, \pi_k)).$$

In this section, we make the following uniform geometric ergodicity assumption.

C3. There exist $L > 0$ and $\alpha \in (0, 1)$ such that

$$\sup_{B \subset \mathbf{S}} |q^n(B \,|\, s, f) - \pi_f(B)| \leq L\alpha^n$$

for any $f \in F$ and all $s \in \mathbf{S}$, $n \geq 1$. Here $q^n(\cdot \,|\, s, f)$ is the n-step transition probability of the Markov chain induced by q and f, and π_f is the unique invariant probability distribution for this Markov chain.

Condition **C3** is rather difficult to check. It is equivalent to the simultaneous Doeblin assumption as formulated by Hordijk [9]. (For a discussion of **C3** in the context of Markov control processes and stochastic games consult [6, 7, 9]). As noted by Meyn and Tweedie (see Theorems 2.1 and 2.2 in [12]), **C3** is implied by the following two assumptions:

C3(a). There exists $\delta > 0$ and a state $z \in \mathbf{S}$ such that

$$q(z \,|\, z, x) > \delta \quad \text{for all } x \in X(z).$$

C3(b). There exists a bounded function $w : \mathbf{S} \to [1, \infty)$ such that for some $\lambda \in (0, 1)$ and $\eta > 0$ we have

$$\sum_{s' \in \mathbf{S}} w(s')q(s' \,|\, s, x) \leq \lambda w(s) + \eta \mathbf{1}_{\{z\}}(s)$$

for each $(s, x) \in \mathrm{D}$. Here $\mathbf{1}_{\{z\}}$ is the characteristic function of the set $\{z\}$.

We are ready to state our first main result.

THEOREM 1. *Any nonzero-sum semi-Markov game with the expected average payoff criterion satisfying conditions* (i)–(iv) *and* **C1**–**C3** *has a stationary Nash equilibrium.*

It is well known that the space F_k of all stationary strategies for player k can be recognized as a compact convex subset of a metrizable topological vector space [6]. A sequence $\{f^n\}$ converges to some f in F_k if and only if $\{f^n(\cdot \,|\, s)\}$ converges weakly to $f(\cdot \,|\, s)$ for each $s \in \mathbf{S}$ in the space of all probability measures on $X_k(s)$, denoted by $P(X_k(s))$.

The following fact follows from page 232 in [18].

LEMMA 1. *Let* $\{v^n\}$ *be a sequence of functions such that* $|v^n(s)| \leq D$ *for some* $D > 0$ *and for all* $s \in \mathbf{S}$. *Assume that* $\{v^n\}$ *converges pointwise to*

some function v and $f^n \to f$ in F (endowed with the product topology) as $n \to \infty$. Then

$$\sum_{s' \in \mathbf{S}} v^n(s')q(s' \mid s, f^n) \to \sum_{s' \in \mathbf{S}} v(s')q(s' \mid s, f).$$

Our proof of Theorem 1 is based on the following result which follows from Theorem 2.1 of [7] and standard iteration arguments (see Theorem 7.6 in [17]).

LEMMA 2. *Suppose that our assumptions* (i)–(iv) *and* **C1**–**C3** *are satisfied. Let $s^* \in \mathbf{S}$ be a fixed state. Then for every player k and any stationary multi-strategy f_{-k} of the other players, there exist a unique bounded function $v_{f_{-k}} : \mathbf{S} \to \mathbb{R}$ and a constant $g_{f_{-k}}$ such that $v_{f_{-k}}(s^*) = 0$ and*

$$(1) \qquad v_{f_{-k}}(s) = \max_{\mu \in P(X_k(s))} \left\{ r_k\big(s, (f_{-k}, \mu)\big) - g_{f_{-k}} \tau\big(s, (f_{-k}, \mu)\big) \right.$$

$$\left. + \sum_{s' \in \mathbf{S}} v_{f_{-k}}(s')q\big(s' \mid s, (f_{-k}, \mu)\big) \right\}.$$

Moreover, if $f_k^0 \in F_k$ is a stationary strategy for player k which realizes the maximum on the right side of (1), *then*

$$g_{f_{-k}} = J_k\big(s, (f_{-k}, f_k^0)\big) = \sup_{\pi_k \in \Pi_k} J_k\big(s, (f_{-k}, \pi_k)\big),$$

i.e., f_k^0 is average payoff optimal for player k.

Proof of Theorem 1. For any $f \in F$ and every player k, define

$$g_{f_{-k}} = \sup_{\pi_k \in \Pi_k} J_k\big((f_{-k}, \pi_k)\big).$$

Let s^* be a fixed state in $\mathbf{S}$. By Lemma 2, there exist unique functions $v_{f_{-k}}$ such that $v_{f_{-k}}(s^*) = 0$ $(k = 1, \ldots, N)$ and $v_{f_{-k}}$ is the solution of the optimality equation (1) for all $s \in \mathbf{S}$.

For each $s \in \mathbf{S}$, define $Y_{f_{-k}}(s)$ as the set of all $x \in X_k(s)$ for which the maximum in the optimality equation (1) is attained.

Our continuity assumption implies that $Y_{f_{-k}}(s)$ are nonempty and compact. Define $\Phi_k(f_{-k})$ as the set of all $\varphi_k \in F_k$ such that $\varphi_k(s)(Y_{f_{-k}}(s)) = 1$ for each $s \in \mathbf{S}$. Clearly, $\Phi_k(f_{-k})$ is a compact convex subset of F_k $(k = 1, \ldots, N)$.

For any $f \in F$, put

$$\Phi(f) = \Phi_1(f_{-1}) \times \Phi_2(f_{-2}) \times \ldots \times \Phi_N(f_{-N}).$$

We will show that Φ has a closed graph. Let $f^n \to f^0$ in F. Assume that $\varphi^n = (\varphi_1^n, \ldots, \varphi_N^n) \in \Phi(f^n)$ for each n and $\varphi^n \to \varphi^0$ in F as $n \to \infty$. We have to show that $\varphi^0 \in \Phi(f^0)$. From our ergodicity assumption **C3**, (iii)

and **C2**, it follows that there exist constants c_1 and c_2 such that

$$|v_{f^n_{-k}}(s)| \le c_1, \quad |g_{f^n_{-k}}| \le c_2 \quad \text{for all } s \in \mathbf{S}, \ n \ge 1.$$

For the details see [7]. Therefore we can construct a subsequence $\{n'\}$ of positive integers such that all the sequences $\{v_{f^{n'}_{-k}}\}$ are pointwise convergent to some bounded functions v_k. Moreover, $g_{f^{n'}_{-k}} \to g_k$ for some real numbers g_k. To simplify notation, we assume that $v_{f^n_{-k}} \to v_k$ pointwise and $g_{f^n_{-k}} \to g_k$ $(k = 1, \ldots, N)$ as $n \to \infty$. The inclusions $\varphi^n \in \Phi(f^n)$ for every n imply that in the limiting case we have

$$(2) \qquad v_k(s) = \max_{\mu \in P(X_k(s))} \left\{ r_k(s, (f^0_{-k}, \mu)) - g_k \tau(s, (f^0_{-k}, \mu)) \right.$$
$$\left. + \sum_{s' \in \mathbf{S}} v_k(s') q(s' \mid s, (f^0_{-k}, \mu)) \right\}$$
$$= r_k(s, (f^0_{-k}, \varphi^0_k)) - g_k \tau(s, (f^0_{-k}, \varphi^0_k))$$
$$+ \sum_{s' \in \mathbf{S}} v_k(s') q(s' \mid s, (f^0_{-k}, \varphi^0_k))$$

for all $s \in \mathbf{S}$. Moreover, $v_k(s^*) = 0$ $(k = 1, \ldots, N)$. By (2) and the uniqueness part of Lemma 2, we have $g_k = g_{f^0_{-k}}$ and $v_k(s) = v_{f^0_{-k}}(s)$. This implies that $\varphi^0 \in \Phi(f^0)$. By Glicksberg's theorem [8], there exists $f^* \in \Phi(f^*)$. This fact and Lemma 2 imply that f^* is a Nash equilibrium for the semi-Markov game in the class of all strategies. ∎

3. Semi-Markov games with the discounted payoff criterion. In this model payoffs are continuously discounted, that is, a payoff r at time t is equivalent to a payoff $re^{-\alpha t}$ at time 0, where $\alpha \in (0, 1)$ is a discount factor. We define $\beta(s, s', x)$ to be the one-step discount function

$$\beta(s, s', x) = \int_0^\infty e^{-\alpha t} F(dt \mid s, s', x).$$

For every $(s, x) \in \mathrm{D}$ and $s' \in \mathbf{S}$, let

$$\eta(s, s', x) = \beta(s, s', x) q(s' \mid s, x).$$

The following fact follows from Lemma 3.4 of [11].

LEMMA 3. *Assume* (i)–(iv) *and* **C1**–**C2**. *Then there exists* $\delta > 0$ *such that for every* $(s, x) \in \mathrm{D}$,

$$\sum_{s' \in \mathbf{S}} \eta(s, s', x) = \sum_{s' \in \mathbf{S}} \int_0^\infty e^{-\alpha t} F(dt \mid s, s', x) q(s' \mid s, x) \le 1 - \delta < 1.$$

The stochastic process $\{\sigma_n, \alpha_n\}$ and the probability measure P_s^π were defined at the beginning of the previous section. For each multi-strategy

$\pi \in \Pi$ and every initial state $s \in \mathbf{S}$, the expected *discounted payoff* to player k is

$$J_k^\beta(s, \pi) = E_s^\pi \Big\{ r_k(\sigma_1, \alpha_1)$$
$$+ \sum_{n=2}^\infty r_k(\sigma_n, \alpha_n) \beta(\sigma_1, \sigma_2, \alpha_1) \dots \beta(\sigma_{n-1}, \sigma_n, \alpha_{n-1}) \Big\}.$$

From Lemma 3, it follows that $J_k^\beta(s, \pi)$ is uniformly bounded in $s \in \mathbf{S}$ and $\pi \in \Pi$.

DEFINITION 2. A multi-strategy $f^* \in F$ is a *stationary Nash equilibrium* for the discounted semi-Markov game if for every player k, $\pi_k \in \Pi_k$ and $s \in \mathbf{S}$, we have
$$J_k^\beta(s, f^*) \geq J_k^\beta(s, (f_{-k}^*, \pi_k)).$$

THEOREM 2. *Any nonzero-sum semi-Markov game with the expected discounted payoff criterion satisfying conditions* (i)–(iv) *and* **C1**–**C2** *has a stationary Nash equilibrium.*

P r o o f. The proof follows the lines of that of Theorem 1. We just replace (1) in the definition of Φ by the following discounted optimality equation:

$$V_{f_{-k}}(s) = \max_{\mu \in P(X_k(s))} \Big\{ r_k(s, (f_{-k}, \mu)) + \sum_{s' \in \mathbf{S}} \eta(s, s', (f_{-k}, \mu)) V_{f_{-k}}(s') \Big\}.$$

Using similar arguments, we prove that there exists $f^* \in \Phi(f^*)$, and to conclude that f^* is a Nash equilibrium in the class of all strategies we refer to the dynamic programming paper by Federgruen and Tijms [7]. ∎

Acknowledgements. We are grateful to Prof. Andrzej S. Nowak for many discussions on the subject of stochastic games and also for his helpful comments.

References

[1]   E. Altman, *Non zero-sum stochastic games in admission, service and routing control in queueing systems*, Queueing Systems Theory Appl. 23 (1996), 259–279.

[2]   E. Altman and A. Hordijk, *Zero-sum Markov games and worst-case optimal control of queueing systems*, ibid. 21 (1995), 415–447.

[3]   E. Altman, A. Hordijk and F. M. Spieksma, *Contraction conditions for average and α-discount optimality in countable state Markov games with unbounded rewards*, Math. Oper. Res. 22 (1997), 588–618.

[4]   D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1979.

[5]   V. S. Borkar and M. K. Ghosh, *Denumerable state stochastic games with limiting average payoff*, J. Optim. Theory Appl. 76 (1993), 539–560.

[6] A. Federgruen, *On n-person stochastic games with denumerable state space*, Adv. Appl. Probab. 10 (1978), 452–471.

[7] A. Federgruen and H. C. Tijms, *The optimality equation in average cost denumerable state semi-Markov decision problems, recurrency conditions and algorithms*, J. Appl. Probab. 15 (1978), 356–373.

[8] I. L. Glicksberg, *A further generalization of the Kakutani fixed point theorem with application to Nash equilibrium points*, Proc. Amer. Math. Soc. 3 (1952), 170–174.

[9] A. Hordijk, *Dynamic Programming and Markov Potential Theory*, Math. Centrum, Amsterdam, 1977.

[10] M. Kurano, *Semi-Markov decision processes and their applications in replacement models*, J. Oper. Res. Soc. Japan 28 (1985), 18–30.

[11] A. K. Lal and S. Sinha, *Zero-sum two-person semi-Markov games*, J. Appl. Probab. 29 (1992), 56–72.

[12] S. P. Meyn and R. L. Tweedie, *Computable bounds for geometric convergence rates of Markov chains*, Ann. Appl. Probab. 4 (1994), 981–1011.

[13] A. S. Nowak, *Some remarks on equilibria in semi-Markov games*, this issue, 385–394.

[14] —, *Sensitive equilibria for ergodic stochastic games with countable state spaces*, Math. Methods Oper. Res. 50 (1999), 65–76.

[15] A. S. Nowak and K. Szajowski, *Nonzero-sum stochastic games*, Ann. Internat. Soc. Dynamic Games 4 (1999), 297–342.

[16] O. Passchier, *The Theory of Markov Games and Queueing Control*, Ph.D. thesis, Dept. Math. and Computer Sci., Leiden Univ., 1996.

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

[18] H. Royden, *Real Analysis*, MacMillan, New York, 1968.

[19] L. I. Sennott, *Average cost semi-Markov decision processes and the control of queueing systems*, Probab. Engnrg. Inform. Sci. 3 (1989), 247–272.

[20] —, *Nonzero-sum stochastic games with unbounded costs*: *discounted and average cost cases*, Z. Oper. Res. 40 (1994), 145–162.

Wojciech Połowczuk
Institute of Mathematics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: polowcz@ulam.im.pwr.wroc.pl