PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 60, October 1976

EXISTENCE OF p-EQUILIBRIUM AND OPTIMAL STATIONARY STRATEGIES IN STOCHASTIC GAMES

C. J. HIMMELBERG, T. PARTHASARATHY, T. E. S. RAGHAVAN
AND F. S. VAN VLECK

ABSTRACT. In this paper we prove the existence of p-equilibrium stationary strategies for non-zero-sum stochastic games when the reward functions and transitions satisfy certain separability conditions. We also prove some results for positive and discounted zero-sum stochastic games when the state space is infinite.

Introduction. A stochastic game is determined by five objects: S, A, B, q, r. Here S is a nonempty Borel subset of a Polish space, the set of states of the system. A is a nonempty Borel subset of a Polish space, the set of actions available to player I; B is the set of actions for player II. The law of motion q associates Borel measurably with each $(s, a, b) \in S \times A \times B$ a probability measure on the Borel subsets of S. Let $r_i(s, a, b)$, $i = 1, 2$, be the reward functions for I and II, respectively, when s is the state and a, b are the actions of I and II. As a consequence of the actions chosen by the players, two things happen: players I and II receive $r_1(s, a, b)$, $r_2(s, a, b)$ and the system moves to a new state s' according to $q(\cdot|s, a, b)$. Then the whole process is repeated from the new state s'. The problem is to find whether they have suitable Nash equilibrium strategies.

A strategy Π for I is a sequence $(\Pi_1, \Pi_2, \ldots)$ where Π_n specifies the action to be chosen on the nth day depending on the past history. A strategy Π is called stationary if there is a Borel map $f\colon S \to P_A$ (the class of all probability distributions on A) such that $\Pi_n \equiv f$ for all n. Similarly, strategies and stationary strategies are defined for II.

Let β be a fixed number with $0 \leqslant \beta < 1$. A pair (Π, Γ) of strategies for I and II associates with each initial state s an nth day expected income $r_i^{(n)}(\Pi, \Gamma)(s)$ for player i and a total expected discounted income for player i:

$$I_i(\Pi, \Gamma)(s) = \sum_{n=1}^{\infty} \beta^{n-1} r_i^{(n)}(\Pi, \Gamma)(s).$$

In case $r_1 = -r_2 = r$, we will call such games discounted zero-sum stochastic games. In case $r \geqslant 0$ and $\beta = 1$, we will simply call them *positive stochastic games*.

Let p be a fixed probability distribution on S. We call (Π^*, Γ^*) a p-equilibrium pair if

Received by the editors April 8, 1974 and, in revised form, December 18, 1974.
AMS (MOS) subject classifications (1970). Primary 90D15, 90D10.
Key words and phrases. Noncooperative-zero-sum-positive stochastic games, p-equilibrium-optimal-stationary strategies.

$$p\{s: I_1(\Pi^*, \Gamma^*)(s) \geqslant I_1(\Pi, \Gamma^*)(s) \quad \text{for all } \Pi,$$
$$I_2(\Pi^*, \Gamma^*)(s) \geqslant I_2(\Pi^*, \Gamma)(s) \quad \text{for all } \Gamma\} = 1.$$

[Here, in general, the set in braces need not be Borel measurable, but it will be universally measurable.]

For the case $r_1 = -r_2$, Π^* is optimal for I if

$$I(\Pi^*, \Gamma)(s) \geqslant \inf_{\Gamma} \sup_{\Pi} I(\Pi, \Gamma)(s) \quad \text{for all } \Gamma \text{ and } s,$$

and Γ^* is optimal for II if

$$I(\Pi, \Gamma^*)(s) \leqslant \sup_{\Pi} \inf_{\Gamma} I(\Pi, \Gamma)(s) \quad \text{for all } \Pi \text{ and } s.$$

If $\inf \sup I(\Pi, \Gamma)(s) = \sup \inf I(\Pi, \Gamma)(s)$, we call this function the value function for the stochastic zero-sum game.

Now we shall prove the following theorems.

THEOREM 1. *Let* $S = [0, 1]$, $A = \{1, 2, \ldots, k\}$, $B = \{1, 2, \ldots, l\}$. *Let* $\gamma_\alpha(s, i, j) = g_\alpha(s, i) + k_\alpha(s, j)$, $\alpha = 1, 2$, *where* g_α, k_α *are bounded measurable functions in* s, *for all* $i \in A$, $j \in B$. *Let* $q(\cdot|s, i, j) = \frac{1}{2}[q'(\cdot|s, i) + q''(\cdot|s, j)]$ *where* q', q'' *are probability measures and further are measurable in* s *for each* $i \in A, j \in B$. *Then for any probability distribution* p *on* $[0, 1]$ *with*

$$q(\cdot|s, i, j) \ll p \quad \text{for all } s, i, j$$

and for any $0 \leqslant \beta < 1$ *there exists a* p*-equilibrium stationary pair* (Π^*, Γ^*) *for the two players.*

THEOREM 2. *Let* S *be any Borel set and* A, B *be finite sets. Let* $q(\cdot|s, i, j)$ *be measurable and* $r_1 = -r_2 = r$ *be a nonnegative bounded measurable function on* S. *Further suppose* $I(\Pi, \Gamma)(s) \leqslant k$ *for all* Π, Γ, s. *Then the positive stochastic game has a measurable value function and player* II *(minimizer) has an optimal stationary strategy.*

THEOREM 3. *Let* S *be complete separable and* A, B *be separable metric spaces. Let* $s \to A(s)$, $s \to B(s)$ *be compact valued multifunctions from* $S \to A$, $S \to B$ *respectively. Here* $A(s)$, $B(s)$ *is the set of actions available to* I *and* II *at state* s. *Let* r: $S \times A \times B \to R'$ *be measurable in* s *and continuous in* (a, b). *Assume that* $q(\cdot|s, a, b)$ *is measurable in* s *and continuous in* (a, b) *in the sense that* $q(\cdot|s, a_n, b_n) \to q(\cdot|s, a_0, b_0)$ *weakly whenever* $(a_n, b_n) \to (a_0, b_0)$. *Finally, suppose that the multifunctions* $s \to F(s) = P_{A(s)}$ *and* $s \to G(s) = P_{B(s)}$ *are measurable. Then the discounted zero-sum stochastic game has a measurable value function and the two players have optimal stationary strategies.*

REMARKS. When S, A, B are finite, Theorem 1 is true without any restriction on γ_1, γ_2 and q [10], [13]. Theorem 2 is also known when S, A, B are finite [12], [9]. A particular case of Theorem 3 is contained in [8]. The real problem in Theorem 1 is to topologize the space of strategies so that it becomes a compact metric space and so that sequential arguments and fixed point theorems could be applied. For the proof of Theorem 1 we need the following facts. Let M_1 and M_2 be the space of all measurable functions from $S \to P_A$ and $S \to P_B$ respectively. Following Warga [15] we shall regard

M_1–after identifying functions coinciding almost everywhere–as a closed convex subset of the dual space of R^k valued integrable functions. With the weak* topology, M_1 and, in a similar fashion, M_2 are compact metric.

The proof of Theorem 1 follows from the following lemmas.

LEMMA 1. *For every stationary strategy g for player* II *the operator T_g defined by*

$$T_g\colon u \to \max_{\mu}\left[r_1(s, \mu, g(s)) + \beta \int u(s')\, dq(s'|s, \mu, g(s))\right]$$

is a contraction operator on the space of bounded measurable functions on [0, 1].

LEMMA 2. *Let u_g be the fixed point of the operator T_g. There exists a measurable function f: $S \to P_A$ such that*

$$u_g(s) = r_1(s, f(s), g(s)) + \beta \int u_g(s')\, dq(s'|s, f(s), g(s)).$$

This follows from a selection theorem due to Olech [7].

We can similarly define operators L_f and fixed points $v_f(s)$ for the function r_2.

LEMMA 3. *Let τ: $M_1 \times M_2 \to 2^{M_1 \times M_2}$ (all nonempty ω^*-closed convex subsets of $M_1 \times M_2$).*

$$\tau\colon (f, g) \to \Big\{(f', g')\colon u_g(s) = r_1(s, f'(s), g(s)) + \beta \int u_g(\cdot)\, dq(\cdot|s, f'(s), g(s))\ a.e.\ \text{and}$$

$$v_f(s) = r_2(s, f(s), g'(s)) + \beta \int v_f(\cdot)\, dq(\cdot|s, f(s), g'(s))\ a.e.\Big\}.$$

The map τ is upper-semicontinuous.

PROOF. Since M_1 is metrizable we can restrict ourselves to sequential arguments. Let $(f_n, g_n) \to (f^0, g^0)$ and $(f_n^*, g_n^*) \in \tau(f_n, g_n)$ with $(f_n^*, g_n^*) \to (f^*, g^*)$. We have to show that $(f^*, g^*) \in \tau(f^0, g^0)$. We have

$$u_{g_n} = r_1(s, f_n^*(s), g_n(s)) + \beta \int u_{g_n}(s')\, dq(s'|s, f_n^*(s), g_n(s)) \quad \text{a.e.}$$

$$v_{f_n} = r_2(s, f_n(s), g_n^*(s)) + \beta \int v_{f_n}(s')\, dq(s'|s, f_n(s), g_n^*(s)) \quad \text{a.e.}$$

Since $\{u_{g_n}\}$ is a uniformly bounded subset of L, it has a convergent subsequence–without loss of generality $\{u_{g_n}\}$ itself–converging in the w^* sense to some u_0. Let

$$f_n^*(s) = (\xi_1^{(n)}(s), \ldots, \xi_k^{(n)}(s)), \quad f^*(s) = (\xi_1(s), \ldots, \xi_k(s)).$$

$$g_n(s) = (\eta_1^{(n)}(s), \ldots, \eta_l^{(n)}(s)), \quad g^0(s) = (\eta_1(s), \ldots, \eta_l(s)).$$

$$W_n'(s, i) = \frac{1}{2}\int u_{g_n}(\cdot)\, dq'(\cdot|s, i), \quad W_0'(s, i) = \frac{1}{2}\int u_0(\cdot)\, dq'(\cdot|s, i).$$

$$W_n''(s, j) = \frac{1}{2}\int u_{g_n}(\cdot)\, dq''(\cdot|s, j), \quad W_0''(s, j) = \frac{1}{2}\int u_0(\cdot)\, dq''(\cdot|s, j).$$

Since $q = \frac{1}{2}(q' + q'') \ll p$ it follows that $q' \ll p$, $q'' \ll p$. Further $u_{g_n} \to u_0$ in

the weak* sense and hence $W_n'(s, i) \to W_0'(s, i)$ and $W_n''(s, j) \to W_0''(s, j)$ pointwise. Since $0 \leqslant \xi_i^n(s) \leqslant 1$, for any integrable function h and for every fixed i,

$$\left|\int h(s) W_n'(s, i)\xi_i^n(s)\, dp(s) - \int h(s) W_0'(s, i)\xi_i(s)\, dp(s)\right|$$
$$\leqslant \int |h(s) W_n'(s, i) - h(s) W_0'(s, i)|\, dp$$
$$+ \left|\int h(s) W_0'(s, i)(\xi_i^n(s) - \xi_i(s))\, dp(s)\right|.$$

The first expression on the right goes to zero by the dominated convergence theorem. The second expression goes to zero since each $\xi_i^{(n)}(s)$ itself converges to $\xi_i(s)$ in the weak* sense. Hence we can conclude that

$$\int u_{g_n}(\cdot)\, dq(\cdot|s, f_n^*(s), g_n(s)) \to \int u_0(\cdot)\, dq\big(\cdot|s, f^*(s), g^0(s)\big)$$

in the weak* sense. Similarly $r_1(s, f_n^*(s), g_n(s)) \to r_1(s, f^*(s), g^0(s))$ in the weak* sense. Hence we can conclude that

$$u_0(s) = r_1\big(s, f^*(s), g^0(s)\big) + \beta \int u_0(\cdot)\, dq\big(\cdot|s, f^*(s), g^0(s)\big) \quad \text{a.e.}$$

Now we will prove that

$$u_0(s) = \max_{\mu}\left[r_1\big(s, \mu, g^0(s)\big) + \beta \int u_0(\cdot)\, dq\big(\cdot|s, \mu, g^0(s)\big)\right] \quad \text{a.e.}$$

Observe that

$$u_{g_n}(s) \geqslant r_1(s, i, g_n(s)) + \beta \int u_{g_n}(\cdot)\, dq(\cdot|s, i, g_n(s)) \quad \text{for all } i \in A, s \in S.$$

Hence we can conclude that

$$u_0(s) \geqslant r_1\big(s, i, g^0(s)\big) + \beta \int u_0(\cdot)\, dq\big(\cdot|s, i, g^0(s)\big) \quad \text{a.e.}$$

Hence $u_0(s)$ satisfies the above functional equation a.e. Using a similar argument for v_{f_n} one can prove

$$v_0(s) = r_2\big(s, f^0(s), g^*(s)\big) + \beta \int v_0(s')\, dq\big(s'|s, f^0(s), g^*(s)\big) \quad \text{a.e.}$$
$$= \max\left[r_2\big(s, f^0(s), \lambda\big) + \beta \int v_0(s')\, dq\big(s'|f^0(s), \lambda\big)\right] \quad \text{a.e.}$$

This shows that $(f^*, g^*) \in \tau(f^0, g^0)$.

We can imitate the same proof to show that $\tau(f, g)$ is a closed set for each (f, g).

LEMMA 4. *There exists a p-equilibrium stationary pair* (f^0, g^0) *for the two players.*

PROOF. The conditions of Kakutani-Glicksberg's fixed point theorem are satisfied for the map τ in Lemma 3 [4]. Hence there exists an $(f^0, g^0) \in \tau(f^0, g^0)$. Namely

$$u_0(s) = \max_{\mu}\left[r_1\big(s, \mu, g^0(s)\big) + \beta\int u_0(s')\, dq\big(s'|s, \mu, g^0(s)\big)\right] \quad \text{a.e.}$$

$$= r_1\big(s, f^0(s), g^0(s)\big) + \beta\int u_0(s')\, dq\big(s'|s, f^0(s), g^0(s)\big) \quad \text{a.e.},$$

$$v_0(s) = \max_{\lambda}\left[r_2\big(s, f^0(s), \lambda\big) + \beta\int v_0(s')\, dq\big(s'|s, f^0(s), \lambda\big)\right] \quad \text{a.e.}$$

$$= r_2\big(s, f^0(s), g^0(s)\big) + \beta\int v_0(s')\, dq\big(s'|s, f^0(s), g^0(s)\big) \quad \text{a.e.}$$

Now we can assume the above equations to be exact over a set S_1 of p measure 1. Since $q(\cdot|s, i, j) \ll p$, $q(S_1|s, i, j) = 1$ for all s, i, j. We can view the problem as a dynamic programming problem on S_1 and hence we can conclude from Blackwell's Theorem (6f) in [2] that

$$u_0(s) = \max_{\Pi} I_1(\Pi, g^0)(s) = I_1(f^0, g^0)(s) \quad \text{for all } s \in S_1,$$

$$v_0(s) = \max_{\Gamma} I_2(f^0, \Gamma)(s) = I_2(f^0, g^0)(s) \quad \text{for all } s \in S_1.$$

The equalities asserted above have in them maxima taken over plans in the dynamic programming problem and they are still true even if we allow behaviour strategies of the game problem. This can be done as in [6, Theorem 3.1]. This establishes that (f^0, g^0) is a p-equilibrium pair. This completes the proof of Theorem 1.

REMARK 1. We are unable to prove the theorem when r_1, r_2 and q do not satisfy the separability conditions.

REMARK 2. The notion of p-optimality as formulated in this paper is due to R. Strauch [14].

PROOF OF THEOREM 2. Let $0 < \beta_n < 1$ be any sequence increasing to 1. From [8, Theorem 3.2], it follows that

$$v_n(s) = \min_{\lambda}\max_{\mu}\left[r(s, \mu, \lambda) + \beta_n\int v_n(s')\, dq(s'|s, \mu, \lambda)\right]$$

$$= \max_{\mu}\min_{\lambda}\left[r(s, u, \lambda) + \beta_n\int v_n(s')\, dq(s'|s, \mu, \lambda)\right].$$

Since $I(\Pi, \Gamma)(s) \leqslant K$ for all Π, Γ, s, the v_n's are bounded. Also the v_n's are monotone nondecreasing. Let $v_n \to v$. We will show that v is the value of the positive stochastic game. Let f_n be optimal for I for the game corresponding to β_n. We have

$$I_n(f_n, \Gamma)(s) \geqslant v_n(s) \quad \text{and} \quad I(f_n, \Gamma)(s) \geqslant I_n(f_n, \Gamma)(s).$$

Here the income I_n corresponds to the case β_n and I refers to the case $\beta = 1$.

Thus $\sup \inf I(\Pi, \Gamma)(s) \geqslant v(s)$. We will now show that

$$\inf \sup I(\Pi, \Gamma)(s) \leqslant v(s).$$

Since A, B are finite,

$$v(s) = \min_{\lambda}\max_{\mu}\left[r(s, \mu, \lambda) + \int v(s')\, dq(s'|s, \mu, \lambda)\right]$$

$$= \max_{\mu}\min_{\lambda}\left[r(s, \mu, \lambda) + \int v(s')\, dq(s'|s, \mu, \lambda)\right]$$

$$= \max_{\mu}\left[r(s, \mu, g(s)) + \int v(s')\, dq(s'|s, \mu, g(s))\right].$$

Here the existence of such a Borel measurable g follows from the theorem of Olech [7]. From a result of Blackwell [1] on positive dynamic programming it follows that

$$v(s) = \sup_{\Pi} I(\Pi, g)(s).$$

This equation is valid even for behaviour strategies of the game problem [6]. Hence we have

$$v(s) = \inf_{\Gamma} \sup_{\Pi} I(\Pi, \Gamma)(s) = \sup_{\Pi} I(\Pi, g)(s) = \sup_{\Pi} \inf_{\Gamma} I(\Pi, g)(s).$$

This proves that the game has a value and player II has an optimal stationary strategy.

REMARK. Player I (maximizer) need not have an optimal stationary strategy. For an example see [8], [14].

PROOF OF THEOREM 3. It follows along similar lines as in [8]. However one has to rely on the following selection theorem proved recently [5].

SELECTION THEOREM [5]. *Let $(S, \mathcal{A})$ be a measurable space, X a separable metric space and Y a separable metric space. Let $u\colon S \times X \to Y$ be a function measurable in s and continuous in x, $\Gamma\colon S \to X$, a measurable multifunction with compact values and $g\colon S \to Y$ a measurable function such that $g(s) \in u(s \times \Gamma(s))$ for all $s \in S$. Then there exists a measurable selector $r\colon S \to X$ for Γ such that $g(s) = u(s, r(s))$ for all s in S.*

ACKNOWLEDGEMENT. We are grateful to the referee for several useful suggestions and for pointing out a serious error in the proof of Lemma 3 in an earlier draft of this paper, where we attempted to prove Lemma 3 without any separability conditions on r_1, r_2 and q.

REFERENCES

1. D. Blackwell, *Positive dynamic programming*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), vol. 1: Statistics, Univ. of California Press, Berkeley, Calif., 1967, pp. 415–418. MR **36** #1193.

2. ———, *Discounted dynamic programming*, Ann. Math. Statist. **36** (1965), 226–235. MR **30** #3749.

3. N. Dunford and J. T. Schwartz, *Linear operators*. 1: *General theory*, Pure and Appl. Math., vol. 7, Interscience, New York, 1958. MR **22** #8302.

4. I. Glicksberg, *A further generalization of the Kakutani fixed point theorem with application to Nash equilibrium points*, Proc. Amer. Math. Soc. **3** (1952), 170–174. MR **13**, 764.

5. C. J. Himmelberg, *Measurable relations*, Fund. Math. **87** (1975), 53–72.

6. A. Maitra and T. Parthasarathy, *On stochastic games*, J. Optimization Theory Appl. **5** (1970), 289–300. MR **41** #8043.

7. C. Olech, *A note concerning set-valued measurable functions*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **13** (1965), 317–321. MR **33** #7486.

8. T. Parthasarathy, *Discounted positive and noncooperative stochastic games*, Internat. J. Game Theory **2** (1973), 25–37.

9. T. Parthasarathy and T. E. S. Raghavan, *Some topics in two-person games*, Modern Analytic and Computational Methods in Sci. and Math., no. 22, American Elsevier, New York, 1971. MR **43** #2996.

10. P. D. Rogers, *Non-zero sum stochastic games*, Ph. D. Thesis, University of California, Berkeley, 1969.

11. L. Schwartz, *Séminaire Schwartz de la Faculté des Sciences de Paris*, 1953/54, *Produits*

tensoriels topologiques d'espaces vectoriels topologiques. Espaces vectoriels topologiques nucléaires. Applications, Secrétariat mathématique, Paris, 1954. MR **17**, 764.

12. L. S. Shapley, *Stochastic games*, Proc. Nat. Acad. Sci. U.S.A. **39** (1953), 1095–1100. MR **15**, 887.

13. M. J. Sobel, *Noncooperative stochastic games*, Ann. Math. Statist. **42** (1971), 1930–1935. MR **46** #8672.

14. R. E. Strauch, *Negative dynamic programming*, Ann. Math. Statist. **37** (1966), 871–890. MR **33** #2456.

15. J. Warga, *Functions of relaxed controls*, SIAM J. Control **5** (1967), 628–641. MR **37** #2064a.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF KANSAS, LAWRENCE, KANSAS 66044 (Current address of C. J. Himmelberg and F. S. Van Vleck)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, CIRCLE CAMPUS, CHICAGO, ILLINOIS 60680 (Current address of T. Parthasarathy)

Current address (T. E. S. Raghavan): Department of Mathematics, Indian Statistical Institute 7, SJS Sansanwal Marg, New Delhi-110029, India