ISRAEL JOURNAL OF MATHEMATICS 78 (1992), 33–49

AN OPERATOR SOLUTION OF STOCHASTIC GAMES

BY

A. MAITRA AND W. SUDDERTH*

School of Statistics
University of Minnesota, Minneapolis, MN 55455, USA

ABSTRACT

A class of zero-sum, two-person stochastic games is shown to have a value which can be calculated by transfinite iteration of an operator. The games considered have a countable state space, finite action spaces for each player, and a payoff sufficiently general to include classical stochastic games as well as Blackwell's infinite G_δ games of imperfect information.

1. Introduction

Let X be a countable, non-empty set of states, and let A and B be finite, non-empty sets of actions for players I and II, respectively. Let u be a bounded, real-valued utility function defined on X and let q be a function which assigns to each triple (x, a, b) in $X \times A \times B$ a probability distribution $q(\bullet|x, a, b)$ on X.

The game starts at some initial state x. Player I chooses an action $a_1 \in A$ and, simultaneously, player II chooses $b_1 \in B$. (The players may choose their actions at random.) The next state x_1 has distribution $q(\bullet|x, a_1, b_1)$ and is announced to the players along with their chosen actions. The procedure is iterated so as to generate a random sequence $x_1, x_2, \ldots$ and the payoff from player II to player I is

$$u^* = \limsup_n u(x_n). \tag{1.1}$$

* Research supported by National Science Foundation Grants DMS-8801085 and DMS-8911548.
Received August 16, 1990 and in revised form March 23, 1992

This payoff function is quite general and includes, for example, the classical payoff

$$\limsup_n \left(\sum_{i=1}^{n} r(x_i) \right) / n$$

where r is a bounded, real-valued function on X. To see this, redefine the state space to be the set of finite sequences $(x_1, \ldots, x_n)$ of elements of X and set

$$u(x_1, \ldots, x_n) = \left(\sum_{i=1}^{n} r(x_i) \right) / n.$$

It is clear how to redefine the law of motion q. One can also redefine the state space to allow the payoff to depend on actions as well as states.

Similar, but slightly more intricate, transformations can be used to show that our formulation includes the G_δ games of Blackwell [2,3]. Indeed the operators defined below are analogous to his.

The techniques which we use to show our game has a value are from the Dubins and Savage theory of gambling [7]. If the action set B for player II is a singleton, then the game is a nonleavable gambling problem for player I. An operator solution to such problems was given by Dubins et al. [6]. Our approach is an extension of their methods which involve the approximation of nonleavable problems by leavable problems in which a gambler is allowed to stop play at any time.

Similarly, we introduce a "leavable game" in which player I (but not player II) can stop play at any time n and receive $u(x_n)$ from player II. It is shown in section 3 that the leavable game starting from state x has a value $U(x)$ which can be calculated by backward induction as follows: To each bounded, real-valued function φ defined on X, let $S\varphi(x)$ be the value of an **auxiliary** one-day game $\mathcal{A}(\varphi)(x)$ starting from x with payoff φ; i.e.

$$S\varphi(x) = \sup_\mu \inf_\nu \int \int \int \varphi(x_1) q(dx_1 | x, a, b) \mu(da) \nu(db)$$

where μ and ν range over the sets of probability measures on A and B, respectively. Define

$$U_0 = u \tag{1.3}$$

and, for $n = 0, 1, \ldots,$

$$U_{n+1} = u \vee SU_n. \tag{1.4}$$

Here $a \vee b$ is the maximum of a with b. Let

$$U = \sup_n U_n.$$

Next define an operator T by the rule

$$Tu = SU. \tag{1.5}$$

Equation (1.5) may be a little confusing at first glance since u does not appear explicitly on the right side. We remind the reader that U is derived from u in accordance with the equations above. For each initial state x, $Tu(x)$ is the value of a game in which player I can stop at any time $n \geq 1$ and receive $u(x_n)$ as will be shown in section 3.

Now let

$$Q_0 = Tu \tag{1.6}$$

and, for every countable ordinal ξ, let

$$Q_\xi = T(u \wedge (\inf_{\eta<\xi} Q_\eta)),$$

where $a \wedge b$ is the minimum of a and b. Set

$$Q = \inf Q_\xi \tag{1.8}$$

Because X is countable and the Q_ξ's are decreasing, there is a countable ordinal ξ^* such that $Q = Q_{\xi^*}$ and $T(u \wedge Q) = Q_{\xi^*+1} = Q$. It is shown in section 4 that $Q(x) = V(x)$, the value of the game with payoff u^* starting from x.

Stochastic games were formulated by Shapley [14], with state and action spaces finite and payoff function equal to the total discounted reward. Shapley proved that his game had a value and that both players had optimal stationary strategies. Thereafter, a number of authors considered the problem when the payoff function is the average reward pay day. Notable contributors to the average reward problem include Gillette [8], Hoffman and Karp [9], Blackwell and Ferguson [4] and Kohlberg [10], who solved different special cases of the problem. The definitive solution of the problem was provided by Mertens and Neyman [12], who used a difficult result of Bewley and Kohlberg [1] on the asymptotic

behavior of the value of the discounted reward game as the discount factor tends to one.

Blackwell [2] proposed a variant of Shapley's game in which the law of motion was eliminated but which allowed for payoff functions more general than either the total discounted reward or the average reward per day. He proved that a win-lose game, where the winning set for player I is a G_δ subset of the set of histories, has a value. In [3], he gave an operator solution of the game problem. This paper, together with [6], forms the basis of the present article.

The next section has some definitions and preliminary results on strategies and stop rules. Leavable games are treated in section 3 and our main result is in section 4. The final section has three simple examples.

2. Preliminaries

Let $Z = A \times B \times X$ and define the **space of histories** to be $H = Z \times Z \times \cdots$. An element $h = (z_1, z_2, \ldots)$ of H will be written as $h = ((a_1, b_1, x_1), (a_2, b_2, x_2), \ldots)$ where $z_n = (a_n, b_n, x_n)$ for every n. We use $p_n(h)$ or, more briefly, p_n to denote the **partial history** $(z_1, \ldots, z_n)$.

Let $P(A)$ and $P(B)$ be the sets of probability measures on A and B, respectively. Given $x \in X$, $\mu \in P(A)$, and $\nu \in P(B)$, write $m = m(x, \mu, \nu)$ for the probability on Z given by

$$m\{(a, b, x_1)\} = \mu\{a\}\nu\{b\}q(\{x_1\}|x, a, b)$$

for $(a, b, x_1) \in Z$.

A **strategy** σ for player I is a sequence $\sigma_0, \sigma_1, \ldots$ where $\sigma_0 \in P(A)$ and, for $n \geq 1$, σ_n is a mapping from Z^n into $P(A)$. A **strategy** τ for player II is defined similarly with $P(B)$ in place of $P(A)$. Strategies σ and τ together with an initial state x determine a probability measure $P_{\sigma,\tau} = P_{x,\sigma,\tau}$ on the Borel subsets of H. (The initial state x will usually be clear from the context and we will usually suppress it.) Namely, the $P_{\sigma,\tau}$ distribution of the first coordinate $z_1 = (a_1, b_1, x_1)$ is $P_{\sigma_0,\tau_0} = m(x, \sigma_0, \tau_0)$ and the $P_{\sigma,\tau}$ conditional distribution of $z_{n+1} = (a_{n+1}, b_{n+1}, x_{n+1})$ given $z_1, \ldots, z_n$ is

$$P_{\sigma,\tau}(\bullet|z_1, \ldots, z_n) = m(x_n, \sigma_n(z_1, \ldots, z_n), \tau_n(z_1, \ldots, z_n)).$$

If g is a bounded, Borel measurable function from H to the reals, we will write its expectation under $P_{\sigma,\tau}$ as $\int g dP_{\sigma,\tau}$ or $E_{\sigma,\tau}g$.

If σ is a strategy and $p = (z_1, \ldots, z_n)$ is a partial history, the **conditional strategy** $\sigma[p]$ is defined by

$$\sigma[p]_0 = \sigma_n(p),$$
$$\sigma[p]_m(z'_1, \ldots, z'_m) = \sigma_{n+m}(z_1, \ldots, z_n, z'_1, \ldots, z'_m)$$

for all $m \geq 1$ and $(z'_1, \ldots, z'_m) \in Z^m$. Given strategies σ and τ for players I and II, the probability $P_{\sigma[p],\tau[p]} = P_{x_n,\sigma[p],\tau[p]}$ is easily seen to be a $P_{\sigma,\tau}$ conditional distribution for $(z_{n+1}, z_{n+2}, \ldots)$ given $(z_1, \ldots, z_n)$. Thus, if g: $H \to R$ is bounded and Borel measurable,

$$E_{\sigma,\tau} g = \int \{E_{\sigma[p_n(h)],\tau[p_n(h)]}(gp_n(h))\} dP_{\sigma,\tau}(h) \tag{2.1}$$

where, for $p = p_n(h) = (z_1, \ldots, z_n)$, gp is the p-section of g defined on H by

$$(gp)(h') = (gp)(z'_1, z'_2, \ldots) = g(z_1, \ldots, z_n, z'_1, z'_2, \ldots).$$

In the special case when $g(h) = u^*(h) = \lim_n \sup u(x_n)$, the function u^*p is just u^* and (2.1) simplifies to

$$E_{\sigma,\tau} u^* = \int (E_{\sigma[p_n(h)],\tau[p_n(h)]} u^*) dP_{\sigma,\tau}(h).$$

A **stopping time** t is a mapping from H to $\{0, 1, \ldots\} \cup \{\infty\}$ such that, for $n = 0, 1, \ldots$, if $t(h) = n$ and h' agrees with h in the first n coordinates, then $t(h') = n$. (Notice that, if $t(h) = 0$ for some h, then t is identically zero.) A **stop rule** t is a stopping time which is everywhere finite.

If t is a stopping time, $h = (z_1, z_2, \ldots) = ((a_1, b_1, x_1), (a_2, b_2, x_2), \ldots)$, and $t(h) < \infty$, we define the functions z_t, x_t, p_t to have values $z_{t(h)}, x_{t(h)}, p_{t(h)} = (z_1, \ldots, z_{t(h)})$ at h. If t is a stop rule, $P_{\sigma[p_t],\tau[p_t]} = P_{x_t,\sigma[p_t],\tau[p_t]}$ is a $P_{\sigma,\tau}$ conditional distribution for $(z_{t+1}, z_{t+2}, \ldots)$ given $(z_1, \ldots, z_t)$ and (2.1) generalizes to

$$E_{\sigma,\tau} g = \int \{E_{\sigma[p_t],\tau[p_t]}(gp_t)\} dP_{\sigma,\tau}. \tag{2.2}$$

If t is a stop rule and $p = (z_1, \ldots, z_n)$ is a partial history, define $t[p]$ on H by

$$t[p](z'_1, z'_2, \ldots) = t(z_1, \ldots, z_n, z'_1, z'_2, \ldots) - n.$$

Notice that, if $t(z_1, \ldots, z_n, \ldots) \geq n$, then $t[p]$ is itself a stop rule in which case $t[p]$ is called a **conditional stop rule given** p. When $p = (z)$, we write z for p and $t[z]$ for $t[p]$.

There is a natural way to associate with every stop rule t an ordinal number $j(t)$ called the **index** of t by setting $j(0) = 0$ and requiring, for $t > 0$, that

$$j(t) = \sup\{j(t[z]) + 1 \colon z \in Z\}.$$

This definition of the index is equivalent to that of Dellacherie and Meyer [5], as was pointed out by Maitra, Pestien, and Ramakrishnan [11, Proposition 4.1]. Furthermore, $j(t)$ is familiar to students of Dubins and Savage as being the structure of the finitary function z_t (cf. [7, sections 2.7 and 2.9]) except for the uninteresting case when Z is a singleton. One of our arguments will use transfinite induction on $j(t)$ and it is important to notice that, for $t > 0$ and all z, $j(t[z])$ is strictly less than $j(t)$.

Consider the special case of (2.1) where $n = 1$ and $g = u(x_t)$ for a stop rule $t > 0$. Notice that $(x_t z_1)(z_2, \ldots) = x_t(z_1, z_2, \ldots) = x_{t[z_1]}(z_2, \ldots)$ if we make the convention that $x_{t[z_1]}(z_2, \ldots) = x_1$ when $t[z_1] = 0$. Thus (2.1) gives

$$E_{\sigma,\tau} u(x_t) = \int \{E_{\sigma[z_1],\tau[z_1]} u(x_{t[z_1]})\} dP_{\sigma_0,\tau_0}(z_1). \tag{2.3}$$

We conclude these preliminaries by stating a result which is needed in order to approximate the game with payoff u^* by leavable games.

LEMMA 2.1: *[15, Theorem 3.2] If u is a bounded, real-valued function on X and P is a probability measure on the Borel subsets of H, then*

$$\int u^* dP = \inf_s \sup_{t \geq s} \int u(x_t) dP$$

where s and t vary over the set of stop rules.

3. Leavable games

Let u be a bounded, real-valued function defined on X. Then u and an initial position x determine a **leavable game** $\mathcal{L}(u)(x)$ in which player I chooses a strategy σ and a stop rule t, player II chooses a strategy τ, and II pays I the quantity $E_{\sigma,\tau} u(x_t)$. Here we allow $t = 0$ and require $x_0 = x$.

THEOREM 3.1: *The leavable game $\mathcal{L}(u)(x)$ has a value $U(x) = \sup U_n(x)$, where the functions U_n are as defined by (1.3) and (1.4).*

For the proof, we will also consider, for $n = 0, 1, \ldots$, an n-day leavable game $\mathcal{L}_n(u)(x)$ with the same rules except that player I must choose a stop rule $t \leq n$.

LEMMA 3.2: *The n-day leavable game $\mathcal{L}_n(u)(x)$ has value $U_n(x)$, and both players have optimal strategies.*

Proof: If $n = 0$, the only stop rule allowed to player I is $t = 0$. So the value of $\mathcal{L}_0(u)(x)$ is clearly $U_0(x) = u(x)$.

Assume the result for n and let $\overline{U}_{n+1}(x)$, $\underline{U}_{n+1}(x)$ be the upper and lower values for $\mathcal{L}_{n+1}(u)(x)$.

To see that $\underline{U}_{n+1}(x) \geq U_{n+1}(x)$, consider two cases. First suppose $u(x) = U_{n+1}(x)$. Then player I takes $t = 0$ and any σ to get $E_{\sigma,\tau}u(x_t) = u(x) = U_{n+1}(x)$ for all τ. Next suppose $SU_n(x) = U_{n+1}(x)$. In this case player I chooses σ_0 to be optimal in the auxiliary game $\mathcal{A}(U_n)(x)$ defined in the introduction and, for each $z_1 = (a_1, b_1, x_1)$, chooses a conditional strategy $\sigma[z_1]$ and stop rule $t[z_1]$ to be optimal in $\mathcal{L}_n(u)(x_1)$. Then, by (2.3), for any τ,

$$
\begin{aligned}
(3.1) \qquad E_{\sigma,\tau}u(x_t) &= \int \{E_{\sigma[z_1],\tau[z_1]}u(x_{t[z_1]})\}dP_{\sigma_0,\tau_0}(z_1) \\
&\geq \int U_n(x_1)dP_{\sigma_0,\tau_0}(z_1) \\
&\geq SU_n(x) \\
&= U_{n+1}(x).
\end{aligned}
$$

To see that $\overline{U}_{n+1}(x) \leq U_{n+1}(x)$, let τ_0 be optimal for player II in the auxiliary game $\mathcal{A}(U_n)(x)$ and, for each $z_1 = (a_1, b_1, x_1)$, let $\tau[z_1]$ be optimal for II in $\mathcal{L}_n(u)(x)$. Given any σ and $t \leq n+1$ for player I, repeat the calculation in (3.1) above. The inequalities reverse to give the desired result. ∎

The next lemma gives two useful properties of the operator S.

LEMMA 3.3: *Let $\varphi_1 \leq \varphi_2 \leq \cdots$ be uniformly bounded, real-valued functions on X. Then*

(a) $S\varphi_1 \leq S\varphi_2$ *and*

(b) $\lim_n S\varphi_n = S(\lim_n \varphi_n)$.

Proof: (a) is obvious. For (b), set $\varphi = \lim \varphi_n$. Fix x and choose $\mu \in P(A)$ so that, for all $b \in B$,

$$\int \int \varphi(x_1) q(dx_1 | x, a, b) \mu(da) \geq S\varphi(x).$$

Let $\epsilon > 0$. Then, for n sufficiently large and all b,

$$\int \int \varphi_n(x_1) q(dx_1 | x, a, b) \mu(da) \geq S\varphi(x) - \epsilon.$$

Hence, for n sufficiently large, $(S\varphi_n)(x) \geq (S\varphi)(x) - \epsilon$. ∎

LEMMA 3.4: $u = U_0 \leq U_1 \leq \cdots$.

Proof: Use Lemma 3.3(a) or Lemma 3.2. ∎

Let $\overline{U}(x)$ and $\underline{U}(x)$ be the upper and lower values of $\mathcal{L}(u)(x)$.

LEMMA 3.5: $\underline{U}(x) \geq U(x)$.

Proof: Let $\epsilon > 0$. Choose n so that $U_n(x) > U(x) - \epsilon$. Let σ, t be optimal in $\mathcal{L}_n(u)(x)$. Then, for every τ, $E_{\sigma,\tau} u(x_t) \geq U_n(X) > U(x) - \epsilon$. ∎

The next result in an extension to leavable games of a fundamental result of Dubins and Savage [7, Corollary 2.14.1].

LEMMA 3.6: *U is the least, bounded, real-valued function φ on X such that $(a)\varphi \geq u$ and $(b) S\varphi \leq \varphi$.*

Proof: Suppose φ satisfies (a) and (b). So $\varphi \geq U_0 = u$. Assume $\varphi \geq U_n$. Then $\varphi \geq S\varphi \geq SU_n$ and $\varphi \geq u \vee SU_n = U_{n+1}$. Hence, $\varphi \geq U_n$ for all n and $\varphi \geq U$.

Obviously, $U \geq u$, and, by Lemma 3.3(b), $SU = S(\lim U_n) = \lim SU_n \leq \lim U_{n+1} = U$. ∎

For each $x \in X$, let $\nu(x)$ be a probability on B which is optimal for player II in the auxiliary game $\mathcal{A}(U)(x)$. Then define τ^x to be the strategy for II such that

$$\tau_0 = \nu(x) \quad \text{and} \quad \tau_n^x(z_1, \ldots, z_n) = \nu(x_n) \qquad \text{for all } n \text{ and } z_1, \ldots, z_n.$$

(Here $z_n = (a_n, b_n, x_n)$.)

LEMMA 3.7: *$\overline{U}(x) \leq U(x)$ and τ^x is an optimal strategy for player II in $\mathcal{L}(u)(x)$.*

Proof: Let σ be a strategy for player II and let t be a stop rule. We will show that

$$E_{\sigma,\tau^x} u(x_t) \leq U(x)$$

by induction on $j(t)$. The inequality is obvious when $j(t) = 0$, i.e. when $t = 0$. Let t be a stop rule with index $j(t) = \alpha > 0$ and assume the inequality holds for all σ, x and stop rules of index less than α. Then, by (2.3)

$$\begin{aligned} E_{\sigma,\tau^x} u(x_t) &= \int \{E_{\sigma[z_1],\tau^{x_1}} u(x_{t[z_1]})\} dP_{\sigma_0,\nu(x)}(z_1) \\ &\leq \int U(x_1) dP_{\sigma_0,\nu(x)}(z_1) \\ &\leq SU(x) \\ &\leq U(x). \qquad \blacksquare \end{aligned}$$

In view of Lemmas 3.5 and 3.7, the proof of Theorem 3.1 is complete.

The next result is a form of the optimality equation of dynamic programming.

LEMMA 3.8: *$U = u \vee SU$.*

Proof: That $U \geq u \vee SU$ is immediate from Lemma 3.6. For the opposite inequality, fix x and suppose $u(x) < U(x)$. Then, for n sufficiently large, $u(x) < U_n(x)$ and so

$$U(x) = \lim_n U_{n+1}(x) = \lim_n SU_n(x) = SU(x)$$

by Lemma 3.3(b). $\blacksquare$

Consider now a slight modification $\mathcal{L}^*(u)(x)$ of the leavable game in which player I chooses a strategy σ and a stop rule $t \geq 1$, player II chooses a strategy τ and, as before, II pays I the quantity $E_{\sigma,\tau} u(x_t)$. The only difference is that player I is not allowed to take $t = 0$.

THEOREM 3.2: *The game $\mathcal{L}^*(u)(x)$ has a value equal to $SU(x)$ and τ^x, as defined before Lemma 3.7, is optimal for player II.*

Proof: Fix $x \in X$ and let y be an element outside X. Consider a new problem with state space $X' = X \cup \{y\}$, the same action sets as before, the same utility u and law of motion q on X and extended to y as follows:

$$\begin{aligned} u(y) &= \inf\{u(x') \colon x' \in X\} - 1, \\ q(\bullet|y,a,b) &= q(\bullet|x,a,b). \end{aligned}$$

In other words, the utility at y is such that player I has every incentive to leave y and the law of motion from y is such that it takes the system to the same states with the same distribution as the law of motion from x.

It now follows that the leavable game $\mathcal{L}(u)(y)$ is equivalent to $\mathcal{L}^*(u)(x)$ as player I has no incentive to use $t = 0$ when the initial state is y. Thus $U(y)$ is the value of $\mathcal{L}^*(u)(x)$ and, by Lemma 3.8,

$$U(y) = u(y) \vee SU(y) = SU(x).$$

The proof of Lemma 3.7 shows that, for all σ and $t \geq 1$,

$$E_{\sigma,\tau^x} u(x_t) \leq SU(x).$$

So τ^x is optimal. ∎

As in the introduction, we denote the value of the game $\mathcal{L}^*(u)(x)$ by $Tu(x)$.

4. Nonleavable games

For each $x \in X$, let $\mathcal{N}(u)(x)$ be the game described in the introduction in which, starting from x, player I chooses a strategy σ, player II chooses a strategy τ, and II pays I the quantity $E_{\sigma,\tau} u^*$.

THEOREM 4.1: *The game $\mathcal{N}(u)(x)$ has a value $V(x)$ which is equal to $Q(x)$, where Q is defined by (1.8).*

Let $\overline{V}(x)$ and $\underline{V}(x)$ be the upper and lower values, respectively, of $\mathcal{N}(u)(x)$.

PROPOSITION 4.2: *If φ is a bounded, real-valued function on X such that $T(u \wedge \varphi) \geq \varphi$, then $\underline{V} \geq \varphi$. In particular, $\underline{V} \geq Q$.*

Proof: As mentioned in the introduction, $T(u \wedge Q) = Q$. So it suffices to prove the first assertion. The proof is similar to that of Theorem 5.1 in [6], but, for the sake of completeness, we will give the details.

Fix $x_0 \in X$ and $\epsilon > 0$. We will construct a strategy σ for player I such that, for every strategy τ for I,

$$E_{\sigma,\tau} u^* \geq \varphi(x_0) - \epsilon. \tag{4.1}$$

The construction involves the composition of a sequence of increasingly better strategies for I in the game $\mathcal{L}^*(u \wedge \varphi)$. So, for each $x \in X$ and $\delta > 0$, let

$\sigma(x,\delta), t(x,\delta)$ be δ-optimal for I in $\mathcal{L}^*(u \wedge \varphi)(x)$. Then, for every τ,

$$E_{\sigma(x,\delta),\tau}(u \wedge \varphi)(x_{t(x,\delta)}) \geq T(u \wedge \varphi)(x) - \delta \geq \varphi(x) - \delta. \tag{4.2}$$

Now choose positive numbers $\delta_0, \delta_1, \ldots$ such that $\sum \delta_n < \epsilon$ and, for each x and n, set $\sigma^n(x) = \sigma(x,\delta_n), t_n(x) = t(x,\delta_n)$. We take the strategy σ to be the **sequential composition of the** (σ_n, t_n) **starting from** x_0. Intuitively, σ follows $\sigma^0(x_0)$ up to time $t_0(x_0)$, then switches to $\sigma^1(x_{t_0(x_0)})$ and so on. To be precise, first define stop rules $s_0 < s_1 < \cdots$ by setting, for each $h = (z_1, z_2, \ldots) \in H$,

$$s_0(h) = t_0(x_0)(h),$$
$$s_{n+1}(h) = s_n(h) + t_{n+1}(x_{s_n})(z_{s_n+1}, z_{s_n+2}, \ldots).$$

Now define

$$\sigma_0 = \sigma^0(x_0)_0,$$

$$\sigma_n(z_1, \ldots, z_n) = \begin{cases} \sigma^0(x_0)_n(z_1, \ldots, z_n) & \text{if } n < s_0(h) \\ \sigma^{k+1}(x_{s_k})_{n-s_k}(z_{s_k+1}, \ldots, z_n) & \text{if } s_k(h) \leq n < s_{k+1}(h). \end{cases}$$

We shall now verify (4.1). Fix a strategy τ for II and let $P = P_{\sigma,\tau}$. The expectations and conditional expectations below are all with respect to P.

Set $Y_n = (u \wedge \varphi)(x_{s_n})$, $n \geq 0$. By assumption,

$$E(Y_0) \geq \varphi(x_0) - \delta_0$$

and, for $n \geq 1$,

$$E(Y_n | p_{s_{n-1}}) \geq \varphi(x_{s_{n-1}}) - \delta_n.$$

So, for $n \geq 1$,

$$E(Y_n) \geq E(\varphi(x_{s_{n-1}})) - \delta_n \geq E(Y_{n-1}) - \delta_n.$$

By iterating this inequality, we get

$$E(Y_n) \geq E(Y_0) - (\delta_1 + \delta_2 + \cdots + \delta_n) \geq \varphi(x_0) - (\delta_0 + \delta_1 + \cdots + \delta_n) \geq \varphi(x_0) - \epsilon, \qquad n \geq 0.$$

Hence

$$\limsup_n E(Y_n) \geq \varphi(x_0) - \epsilon.$$

But

$$\begin{aligned} E(u^*) &= E(\limsup_n u(x_n)) \\ &\geq E(\limsup_n u(x_{s_n})) \\ &\geq E(\limsup_n Y_n) \\ &\geq \limsup_n E(Y_n) \\ &\geq \varphi(x_0) - \epsilon. \end{aligned}$$

This completes the proof of (4.1) and of the proposition. ∎

LEMMA 4.3: $\overline{V} \leq Q$.

Proof: It suffices to show $\overline{V} \leq Q_\xi$ for each countable ordinal ξ and we will do so by induction on ξ.

To see that $\overline{V} \leq Q_0$, fix x and let τ be optimal for II in $\mathcal{L}^*(u)(x)$. Then for any σ for I, it follows from Lemma 2.1 that

$$E_{\sigma,\tau} u^* \leq \sup_{t \geq 1} E_{\sigma,\tau} u(x_t) \leq Tu(x) = Q_0(x).$$

Now let ξ be a positive ordinal and assume that $\overline{V} \leq \inf_{\eta<\xi} Q_\eta$. Set $R_\xi = \inf_{\eta<\xi} Q_\eta$. To show $\overline{V} \leq Q_\xi$, fix x and $\epsilon > 0$. We will find a strategy τ for II such that, for all σ for I,

$$E_{\sigma,\tau} u^* \leq Q_\xi(x) + \epsilon, \tag{4.3}$$

which clearly suffices.

To define τ, first choose τ^1 to be an optimal strategy for II in $\mathcal{L}^*(u \wedge R_\xi)(x)$ and, for every $y \in X$, choose $\overline{\tau}(y)$ for II in $\mathcal{N}(u)(y)$ so that, for all σ,

$$E_{\sigma,\overline{\tau}(y)} u^* < \overline{V}(y) + \epsilon/2 \leq R_\xi(y) + \epsilon/2. \tag{4.4}$$

For each $h = (z_1, z_2, \ldots) = ((a_1, b_1, x_1), (a_2, b_2, x_2), \ldots)$, let

$$\lambda(h) = \inf\{k\colon u(x_k) > R_\xi(x_k)\}. \tag{4.5}$$

Then λ is a stopping time with ∞ as a possible value. Now take τ to be that strategy which follows τ^1 prior to time λ and then switches to $\overline{\tau}(x_\lambda)$; that is,

$$\tau_0 = \tau_0^1,$$

$$\tau_n(z_1,\dots,z_n) = \begin{cases} \tau_n^1(z_1,\dots,z_n) & \text{if } n < \lambda(h), \\ \bar{\tau}(x_\lambda)_{n-\lambda}(z_{\lambda+1},\dots,z_n) & \text{if } n \geq \lambda(h). \end{cases}$$

Fix a strategy σ for I and we will verify (4.3). By Lemma 2.1, it suffices to find a stop rule s such that, for all stop rules $t \geq s$,

$$E_{\sigma,\tau}u(x_t) \leq Q_\xi(x) + \epsilon. \tag{4.6}$$

To obtain s, first choose a positive integer m such that

$$P_{\sigma,\tau}[\lambda < \infty] \leq P_{\sigma,\tau}[\lambda \leq m] + \epsilon/(4(\sup|u| + 1)). \tag{4.7}$$

Also, for each partial history $p = (z_1,\dots,z_n)$ with $z_n = (a_n, b_n, x_n)$, use Lemma 2.1 and (4.4) to get a stop rule $\bar{t}(p)$ such that, for all stop rules $t \geq \bar{t}(p)$,

$$E_{\sigma[p],\bar{\tau}(x_n)}u(x_t) \leq R_\xi(x_n) + \epsilon/2. \tag{4.8}$$

Now, for $h = (z_1, z_2, \dots)$, define

$$s(h) = \begin{cases} \lambda(h) + \bar{t}(p_\lambda)(z_{\lambda+1}, z_{\lambda+2}, \dots) & \text{if } \lambda(h) \leq m, \\ m & \text{if } \lambda(h) > m. \end{cases}$$

Let $t \geq s$. To check (4.6), condition on $p_{\lambda\wedge t}$ and calculate:

$$\begin{aligned} E_{\sigma,\tau}u(x_t) &= \int_{\lambda\leq t} E_{\sigma[p_\lambda],\bar{\tau}(x_\lambda)}(u(x_{t[p_\lambda]})dP_{\sigma,\tau} + \int_{\lambda>t} u(x_t)dP_{\sigma,\tau} \\ &\leq \int_{\lambda\leq m} E_{\sigma[p_\lambda],\bar{\tau}(x_\lambda)}(u(x_{t[p_\lambda]}))dP_{\sigma,\tau} + \int_{\lambda>t} u(x_t)dP_{\sigma,\tau} + \epsilon/4 \\ &\leq \int_{\lambda\leq m} R_\xi(x_\lambda)dP_{\sigma,\tau} + \int_{\lambda>t} u(x_t)dP_{\sigma,\tau} + 3\epsilon/4 \\ &\leq \int_{\lambda\leq t} R_\xi(x_\lambda)dP_{\sigma,\tau} + \int_{\lambda>t} u(x_t)dP_{\sigma,\tau} + \epsilon \\ &= \int (u \wedge R_\xi)(x_{\lambda_{\wedge t}})dP_{\sigma,\tau^1} + \epsilon \\ &\leq T(u \wedge R_\xi)(x) + \epsilon \\ &= Q_\xi(x) + \epsilon. \end{aligned}$$

The first line above uses the equality $\tau[p_\lambda] = \bar{\tau}(x_\lambda)$ for $\lambda < \infty$; the second is by (4.7); the third by (4.8) and the fact that $t[p_\lambda] \geq s[p_\lambda] = \bar{t}(p_\lambda)$ if $\lambda \leq m$; the fourth by (4.7) and the fact that $\sup|R_\xi| \leq \sup|u|$; the fifth by (4.5); the last two lines are by choice of τ^1 and definition of Q_ξ, respectively. ∎

Theorem 4.1 is immediate from Proposition 4.2 and Lemma 4.3.

Mertens and Neyman [12] showed not only that the average reward stochastic game with finite state and action spaces has a value, but also that ϵ-optimal strategies exist which are also ϵ-optimal for games of a sufficiently long finite horizon. Theorem 4.1 implies that the average reward game with a countable state space and finite action sets has a value. The stronger result of Mertens and Neyman does not hold for games with a countable state space and finite action spaces. We do not know whether our method of proof when specialized to finite state and action spaces will yield the stronger result.

This proof that $\mathcal{N}(u)$ has a value is analogous to that given by Blackwell in [3] that his G_δ games have a value. We could also imitate Blackwell's earlier proof in [2] by arguing as in the proof of Lemma 4.3 that $T(u \wedge \overline{V}) \geq \overline{V}$ and then applying Proposition 4.2 to conclude that $\underline{V} \geq \overline{V}$. Such a proof would be sightly shorter, but less constructive as Blackwell pointed out. Our leavable game $\mathcal{L}^*(u)(x)$ is the analog of Blackwell's auxiliary game. We are able to avoid use of Sion's minimax theorem, which Blackwell invokes to show that his auxiliary game has a value, by reducing the game $\mathcal{L}^*(u)(x)$ through backward induction to a one-shot game, at which stage we can apply von Neumann's theorem. As we will show in another paper, the proof given here can be generalized to a Borel measurable setting. We conclude this section with a characterization of V similar to Theorem 7.1 in [6].

THEOREM 4.4: *The value function V for the game $\mathcal{N}(u)$ is the largest, bounded, real-valued function φ on X such that*

$$T(u \wedge \varphi) = \varphi \tag{4.9}$$

Proof: The function V is a solution to (4.9) because $V = Q$. Also, every solution φ of (4.9) is majorized by V as follows from Proposition 4.2. ∎

5. Three examples

To illustrate the use of the operators, we present two simple examples. The first is a very special case of a class of win, lose, or draw games which were suggested to us by David Blackwell.

EXAMPLE 1: *Let $X = \{w, l, d\}$; $u(w) = 1$, $u(l) = -1$, $u(d) = 0$; $A = B = \{0, 1\}$; $q(w|w, a, b) = 1$ and $q(l|l, a, b) = 1$ for all $a \in A, b \in B$; $q(w|d, 1, 1) = q(l|d, 1, 0) = q(l|d, 0, 1) = 1$, $q(w|d, 0, 0) = q(d|d, 0, 0) = 1/2$.*

To simplify notation, we write functions on X in vector form. So, for example, the utility function u becomes $u = (1, -1, 0)$.

The value of the auxiliary game $\mathcal{A}(u)$ is easily found to be $Su = (1, -1, -1/7)$ and, hence, $U_1 = u \vee Su = u$. Consequently, $U_n = u$ for all n, $U = u$, and $Q_0 = Tu = Su = (1, -1, -1/7)$.

Similar calculations show that for $n = 0, 1, \ldots, Q_{n+1} = TQ_n = (1, -1, x_{n+1})$ where $x_0 = -1/7$ and $x_{n+1} = (x_n - 1)/(x_n + 7)$. Furthermore the x_n decrease to a limit $x^* = \sqrt{8} - 3$. It is easily checked that $Q = \inf Q_n = (1, -1, x^*)$ satisfies $T(u \wedge Q) = TQ = Q$. So $(1, -1, x^*)$ is the value. ∎

The game of Example 1 is the same if we take the payoff to be

$$\limsup_n \left(\sum_{i=1}^{n} u(x_i) \right) / n.$$

So its value could also be calculated from that of the discounted games as in [1]. The value for our next example, which corresponds to Example 1 of Orkin [13], cannot be calculated from discounted games.

EXAMPLE 2: *Let* $X = \{w, l, g, d\}$; $u(w) = u(g) = 1$, $u(l) = u(d) = 0$; $A = B = \{0, 1\}$; $q(w|w, a, b) = q(l|l, a, b) = 1$ *for all* $a \in A$, $b \in B$, $q(w|g, 1, 1) = q(w|d, 1, 1) = 1$, $q(g|g, 0, 0) = q(g|d, 0, 0) = 1$, $q(d|g, 0, 1) = q(d|d, 0, 1) = 1$, $g(l|g, 1, 0) = q(l|d, 1, 0) = 1$.

As in Example 1, we use vector notation. So $u = (u(w), u(l), u(g), u(d)) = (1, 0, 1, 0,)$. The value of $\mathcal{A}(u)$ is found to be $Su = (1, 0, 1/2, 1/2)$ and $U_1 = u \vee Su = (1, 0, 1, 1/2)$. One shows inductively, for $n = 0, 1, \ldots$, that $SU_n = (1, 0, x_n, x_n)$ where $x_0 = 1/2$ and $x_{n+1} = (2 - x_n)^{-1}$. Thus $U_{n+1} = u \vee SU_n = (1, 0, 1, x_n)$ and $U = \lim U_n = (1, 0, 1, \lim x_n) = (1, 0, 1, 1)$.

It follows that $Q_0 = Tu = SU = (1, 0, 1, 1)$ also and $Q_1 = T(u \wedge Q_0) = Tu = Q_0$. So Q_0 is a fixed point and $V = Q_0 = (1, 0, 1, 1)$. ∎

Finally consider a modification $I(u)(x)$ of the nonleavable game in which the payoff from player II to player I is

$$u_* = \liminf_n u(x_n).$$

Now $u_* = -(-u)^*$. So, if we reverse the roles of the two players, if follows from Theorem 4.1 that the game $I(u)(x)$ has a value, say $W(x)$. Clearly $W(x) \leq V(x)$

since $u_* \leq u^*$. For average reward games with finite state and action sets as in Mertens and Neyman [12], the values $W(x)$ and $V(x)$ are the same. Here is a simple example to show the values need not be the same for an average reward game with a countable state space.

EXAMPLE 3: *Let X be the set of all finite sequences $x = (n_1, n_2, \ldots, n_k)$ of positive integers; $A = B = \{1\}$; let G be a subset of the positive integers which has inner density zero and outer density one, that is, if r is the indicator function of G,*

$$\liminf_k \left(\sum_{i=1}^{k} r(i) \right) / k = 0,$$

$$\limsup_k \left(\sum_{i=1}^{k} r(i) \right) / k = 1;$$

define

$$u(n_1, \ldots, n_k) = \frac{1}{k} \sum_{i=1}^{k} r(n_i)$$

and

$$q((n_1, n_2, \ldots, n_k, n_k + 1) | (n_1, n_2, \ldots, n_k), 1, 1) = 1.$$

Then, for every x, $W(x) = 0$ and $V(x) = 1$.

References

[1] T. Bewley and E. Kohlberg, *The asymptotic theory of stochastic games,* Math. Oper. Res. **1** (1976), 197–208.

[2] D. Blackwell, *Infinite G_δ games with imperfect information,* Zastos. Mat. **10** (1969), 99–101.

[3] D. Blackwell, *Operator solution of infinite G_δ games of imperfect information,* in *Probability, Statistics and Mathematics,* Papers in Honor of S. Karlin (T. W. Anderson, K. B. Athreya and D. L. Iglehart, eds.), Academic Press, New York, 1989, pp. 83–87.

[4] D. Blackwell and T. S. Ferguson, *The big match,* Ann. Math. Statist. **39** (1968), 159–163.

[5] C. Dellacherie and P. A. Meyer, *Ensembles analytiques et temps d'arrêt,* in *Seminaire de Probabilités* IX, Lecture Notes in Math. **465**, Springer-Verlag, Berlin and New York, 1975, pp. 373–389.

[6] L. Dubins, A. Maitra, R. Purves and W. Sudderth, *Measurable, nonleavable gambling problems,* Isr. J. Math. **67** (1989), 257–271.

[7] L. E. Dubins and L. J. Savage, *Inequalities for Stochastic Processes,* Dover, New York, 1976.

[8] D. Gillette, *Stochastic games with zero-stop probabilities,* in *Contributions to the Theory of Games III,* Ann. Math. Studies, No. 39, Princeton University Press, Princeton, 1957, pp. 179–187.

[9] A. J. Hoffman and R. M. Karp, *On nonterminating stochastic games,* Management Sci. **12** (1966), 359–370.

[10] E. Kohlberg, *Repeated games with absorbing states,* Ann. Statist. **2** (1974), 724–738.

[11] A. Maitra, V. Pestien and S. Ramakrishnan, *Domination by Borel stopping times and some separation properties,* Fund. Math. **35** (1990), 189–201.

[12] J.-F. Mertens and A. Neyman, *Stochastic games,* Int. J. Game Theory **10** (1981), 53–66.

[13] M. Orkin, *Infinite games with imperfect information,* Trans. Am. Math. Soc. **171** (1972), 501–507.

[14] L. Shapley, *Stochastic games,* Proc. Natl. Acad. Sci. U.S.A. **39** (1953), 1095–1100.

[15] W. Sudderth, *On measurable gambling problems,* Ann. Math. Statist. **42** (1971), 260–269.