On Nonterminating Stochastic Games
Author(s): A. J. Hoffman and R. M. Karp
Source: *Management Science*, Vol. 12, No. 5, Series A, Sciences (Jan., 1966), pp. 359-370
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/2627820

MANAGEMENT SCIENCE
Vol. 12, No. 5, January, 1966
Printed in U.S.A.

ON NONTERMINATING STOCHASTIC GAMES*

A. J. HOFFMAN† AND R. M. KARP

IBM Watson Research Center

A stochastic game is played in a sequence of steps; at each step the play is said to be in some state i, chosen from a finite collection of states. If the play is in state i, the first player chooses move k and the second player chooses move l, then the first player receives a reward a_i^{kl}, and, with probability p_{ij}^{kl}, the next state is j.

The concept of stochastic games was introduced by Shapley with the proviso that, with probability 1, play terminates. The authors consider the case when play never terminates, and show properties of such games and offer a convergent algorithm for their solution. In the special case when one of the players is a dummy, the nonterminating stochastic game reduces to a Markovian decision process, and the present work can be regarded as the extension to a game theoretic context of known results on Markovian decision processes.

1. Introduction

In 1953, L. S. Shapley [13] introduced the concept of a stochastic game. Such a game is played in a sequence of steps; at each step the play is said to be in some state i, chosen from a finite collection of states. If the play is in state i, the first player chooses move k and the second player chooses move l, then the first player receives a reward a_i^{kl}, and, with probability p_{ij}^{kl}, the next state is j.

Shapley postulated that, for each i, k, and l,

$$\sum_j p_{ij}^{kl} < 1$$

and that play terminates with probability

$$1 - \sum_j p_{ij}^{kl}.$$

He stated an appropriate definition for the value of such a stochastic game, derived a set of functional equations that the value must satisfy, and prescribed an algorithm for finding the value and optimal strategies. Extensions of these ideas were also given by Everett [5].

The purpose of the present paper is to treat the situation where, for each i, k, and l,

$$\sum_j p_{ij}^{kl} = 1;$$

i.e., play never terminates. This situation has also been studied by Gillette [8], whose results somewhat overlap ours. We shall, for these nonterminating stochastic games, make definitions and derive results analogous to Shapley's.

A stochastic game in which one of the two players is a dummy having no

* Received August 1964, revised July 1965.

† This research was supported in part by the Office of Naval Research under Contract No. Nonr 3775(00), NR 047040.

choice of strategies is precisely a Markovian decision process. Such processes are discussed extensively in a book by Howard [10]. Our work builds on Howard's, and also on the linear programming formulation of Markovian decision processes which has been pointed out by several authors ([2], [3], [4], [6], [7], [12], [16]). In addition, we depend very heavily on results of Aumann [1] and Derman [2] which enable us to restrict the class of strategies considered.

We shall also need to know sufficient conditions for the optimal value of a linear program to be a continuous function of its data, and such conditions will be given in the Appendix.

The present work may be viewed as generalizing to a game-theoretic setting the problem of maximizing the expected reward per period in a Markovian decision process. The analogous generalization of the problem of maximizing total discounted reward has already been supplied by Shapley [13].

2. Summary of Results from Markovian Decision Processes

A nonterminating Markovian decision process is a nonterminating stochastic game (NSG) in which the second player is a "dummy", having only one move available in each of the n states. The play begins in a specified initial state X_0. At a general step, when the play is in state i, and the (first) player chooses move k from a finite set of choices (indexed $1, 2, \cdots, M_i$) he receives an amount a_i^k, and the probability of a transition to state j is p_{ij}^k, where

$$\sum_{j=1}^{n} p_{ij}^k = 1.$$

A pure strategy for such a process is defined by a set of functions

$$F_{kp}(X_0, \Delta_0, X_1, \Delta_1, \cdots, X_{p-1}, \Delta_{p-1}, X_p), \qquad F_{kp} \varepsilon \{0, 1\}, \qquad \sum_k F_{kp} = 1,$$

where X_q is the state occurring at step q, and Δ_q is the move made at step q. $F_{kp} = 1$ if, given the history of the process, the player will choose move k at step p; otherwise, $F_{kp} = 0$. The payoff $Q(S)$ associated with a pure strategy S is taken to be

$$\limsup_{r \to \infty} (1/r) \sum_{p=0}^{r} E(r_p),$$

where $E(r_p)$ is the expected reward at the p^{th} step. The payoff is therefore the average reward per step.

A mixed strategy is a probability distribution over the set of pure strategies. A completely precise definition of the set of mixed strategies is complex, requiring the introduction of an appropriate measurable structure over the space of pure strategies. Fortunately, it is possible, by virtue of Aumann's generalization [1] of a theorem of Kuhn [11], to restrict attention to a special class of mixed strategies known as behavior strategies. Behavior strategies in multimove games are characterized by the property that the player makes an independent randomization at each move, rather than selecting an over-all pure strategy by randomization. In the case of a Markovian decision process, a behavior strategy is defined by a set of functions

$$F_{kp}(X_0, \Delta_0, X_1, \Delta_1, \cdots, X_{p-1}, \Delta_{p-1}, X_p), \qquad F_{kp} \geqq 0, \qquad \sum_k F_{kp} = 1,$$

where X_q and Δ_q are defined as before. In this case, F_{kp} gives the probability that the player will choose move k at step p, given the history of the process.

Aumann's theorem applies to games of perfect recall. Intuitively, a game is of perfect recall with respect to a player if, at every move, he remembers what he knew and what he did at previous moves. Markovian decision processes satisfy the definition of a (one-person) game of perfect recall. It is a consequence of Aumann's theorem that, in a game of perfect recall with respect to a player, that player can restrict himself to behavior strategies.

Within the class of behavior strategies for Markovian decision processes, the stationary strategies may be distinguished. A stationary strategy is a behavior strategy in which F_{kp} depends only on k and X_p; i.e., there is a set of numbers w_i^k, $1 \leqq i \leqq n$, $1 \leqq k \leqq M_i$, such that, for $1 \leqq k \leqq M_{X_p}$ and $0 \leqq p < \infty$, $F_{kp}(X_0, \Delta_0, X_1, \Delta_1, \cdots, X_{p-1}, \Delta_{p-1}, X_p) = w_{X_p}^k$. Let C denote the set of all behavior strategies. The following fundamental result has been given by Derman [2].

Theorem 1. For any Markovian decision process there exists a stationary strategy T such that $Q(T) = \max_{S \varepsilon C} Q(S)$.

When a stationary strategy is used, the process may be described as an n-state Markov chain with initial state X_0 in which each transition probability p_{ij} is given by

$$\sum_{k=1}^{M_i} w_i^k p_{ij}^k .$$

It will be convenient to restrict attention to the case in which every stationary strategy yields an irreducible Markov chain. For this reason, we henceforth assume the following *irreducibility property*: For any given choice of integers $k_1, k_2, \cdots, k_n$ such that $1 \leqq k_i \leqq M_i$, $i = 1, \cdots, n$, the stochastic matrix having $p_{ij}^{k_i}$ as its $i - j$ element is irreducible; i.e., for any i and j, the probability that state j will be reached from the initial state i is 1. On this assumqtion. the payoff $Q(S)$ associated with a stationary strategy S is given by

$$\sum_{i=1}^{n} \Pi_i \sum_{k=1}^{M_i} w_i^k a_i^k,$$

where

$$(\Pi_1, \Pi_2, \cdots, \Pi_n)$$

is the unique positive vector of stationary probabilities. Thus, the payoff is independent of the initial state. The determination of an optimal strategy can be formulated as a linear programming problem as follows·

Problem I.

maximize $$\sum_{i=1}^{n} \sum_{k=1}^{M_i} X_i^k a_i^k$$

subject to

(a) $$X_i^k \geqq 0, \qquad 1 \leqq i \leqq n, \qquad 1 \leqq k \leqq M_i$$

(b) $$\sum_{i=1}^{n} \sum_{k=1}^{M_i} X_i^k = 1$$

(c) $$\sum_{k=1}^{M_i} X_j^k - \sum_{i=1}^{n} \sum_{k=1}^{M_i} X_i^k p_{ij}^k = 0, \qquad j = 1, \cdots, n.$$

From the irreducibility property we have, for any feasible solution to the linear program,

$$\sum_{k=1}^{M_i} X_i^k > 0, \qquad \text{for all } i.$$

It follows that there is a one-to-one correspondence between feasible solutions to Problem I and stationary strategies:

$$\Pi_i = \sum_{k=1}^{M_i} X_i^k; \qquad \Pi_i w_i^k = X_i^k.$$

For a given feasible solution to the linear program, the value of the objective function is equal to the payoff for the associated stationary strategy. Also, it may be observed that the n constraints given in (c) are not linearly independent, since the sum of their left-hand sides is identically zero. Accordingly, the linear program has only n linearly independent equality constraints, and its basic feasible solutions, at most n nonzero variables. Thus, in a basic feasible solution, exactly one of the X_i^k is nonzero for each i; and, in the associated stationary strategy, one of the w_i^k is equal to 1, for each i, the others being zero. Since a linear programming problem has one or more basic feasible solutions among its optimal solutions, it follows that there is an optimal stationary strategy that is also a pure strategy. This confirms the intuitive belief that randomization is not required of the player in a Markovian decision process.

The dual of Problem I is the following linear program.

Problem II.

Minimize g

Subject to

$$g + v_i \geqq a_i^k + \sum_{j=1}^{n} p_{ij}^k v_j, \qquad 1 \leqq i \leqq n, \qquad 1 \leqq k \leqq M_i.$$

The dual variable g corresponds to the primal constraint (b), and v_j corresponds to the j^{th} constraint in the set (c). It should be noted that, if $(g, v_1, v_2, \cdots, v_n)$ is a feasible solution to this linear program, so also is $(g, v_1 + c, v_2 + c, \cdots, v_n + c)$, where c is an arbitrary constant. This degree of freedom is due to the dependence existing among the constraints (c) in Problem I.

Lemma 1. Any optimal solution $(g, v_1, v_2, \cdots, v_n)$ of Problem II satisfies:

(1) $$g + v_i = \max_{1 \leqq k \leqq M_i} a_i^k + \sum_{j=1}^{n} p_{ij}^k v_j, \qquad 1 \leqq i \leqq n.$$

Proof. By the principle of complementary slackness [14, p. 94], an optimal program for Problem II must satisfy

$$g + v_i = a_i^k + \sum_{j=1}^{n} p_{ij}^k v_j$$

if X_i^k is nonzero in any optimal program for Problem I. But we have remarked earlier that for any value of i, any feasible solution to Problem I assigns a positive value to at least one variable X_i^k.

Theorem 2. If $(g, v_1, v_2, \cdots, v_n)$ and $(g', v'_1, v'_2, \cdots, v'_n)$ both satisfy (1), then $g = g'$ and there is a constant c such that $v_i - v'_i = c$, $1 \leqq i \leqq n$.

Proof. For each i, choose k_i such that

$$g + v_i = a_i^{k_i} + \sum_{j=1}^{n} p_{ij}^{k_i} v_j ; \qquad \text{in addition,}$$

$$g' + v'_i \geqq a_i^{k_i} + \sum_{j=1}^{n} p_{ij}^{k_i} v'_j , \qquad \text{so that}$$

$$(g' - g) + (v'_i - v_i) \geqq \sum_{j=1}^{n} p_{ij}^{k_i} (v'_j - v_j), \qquad 1 \leqq i \leqq n.$$

If $g' - g$ were positive, each component of the vector

$$v' - v = (v'_1 - v_1, v'_2 - v_2, \cdots, v'_n - v_n)$$

would exceed a convex combination of components. Since this is impossible, $g' - g \leqq 0$; interchanging the roles of the solutions (g', v') and (g, v), we may establish similarly that $g - g' \leqq 0$. Therefore, $g = g'$, and, setting $g' - g$ equal to zero we have

$$(2) \qquad v' - v \geqq P(v' - v),$$

where P is an $n \times n$ matrix having $p_{ij}^{k_i}$ as its $i - j$ element. By the irreducibility assumption P is an irreducible Markov matrix, so that the only solutions to (2) are those in which all components of $v' - v$ are equal. This completes the proof.

Thus, Problem II may be solved by finding the essentially unique solution to (1). And further, a behavior strategy is optimal if and only if $w_i^k > 0$ implies that

$$g + v_i = a_i^k + \sum_{j=1}^{n} p_{ij}^k v_j ,$$

where $(g, v_1, \cdots, v_n)$ is a solution to (2).

Howard [10] has given a finite iterative procedure for solving these equations, and Wolfe and Dantzig [16] have shown that the decomposition principle of linear programming may be used to effect a solution.

3. Optimal Strategies for NSG's

A pure strategy for a player in a NSG with initial state X_0 is a set of functions

$$F_{kp}(X_0, \Delta_0, \Sigma_0, \cdots, X_{p-1}, \Delta_{p-1}, \Sigma_{p-1}, X_p), \qquad F_{kp} \in \{0, 1\}, \qquad \sum_k F_{kp} = 1,$$

where X_q is the state occurring at step q, and Δ_q and Σ_q are the moves of players 1 and 2, respectively, at step q; $F_{kp} = 1$ if the player will choose move k at step p, knowing the entire history of the play, including the moves of both players. If the players choose pure strategies S and S', respectively, the resulting payoff to the first player is taken to be

$$Q(S, S') = \limsup_{r\to\infty} (1/r) \sum_{p=0}^{r} E(r_p)$$

where $E(r_p)$ is the expected value of the reward received by the first player at the p^{th} step given that the strategy pair S, S' is used. The payoff is therefore the average reward per step, just as in the special case of Markovian decision processes.

The definition of a behavior strategy for a NSG differs from the definition of a pure strategy in only one respect: the condition '$F_{kp} \varepsilon \{0, 1\}$' is replaced by '$F_{kp} \geqq 0$'. Since a NSG is a game of perfect recall for each player, it follows from Aumann's theorem that each player may restrict himself to behavior strategies. It will be shown later that, in fact, each player may restrict himself to stationary strategies.

Let us suppose that the second player follows a stationary strategy S', always choosing move l with probability u_i^l when the play is in state i. If the first player knows his opponent's strategy, he can maximize his payoff by following an optimal strategy for the Markovian decision process in which

$$(3) \qquad a_i^k = \sum_{l=1}^{N_i} a_i^{kl} u_i^l, \qquad 1 \leqq i \leqq n, \qquad \text{and}$$

$$(4) \qquad p_{ij}^k = \sum_{l=1}^{N_i} p_{ij}^{kl} u_i^l, \qquad 1 \leqq i \leqq n, \qquad 1 \leqq j \leqq n.$$

Here N_i denotes the number of moves available to the second player in state i. Let this maximum payoff be called $g(S')$. Our first undertaking in the present section will be to show that $g(S')$ assumes a minimum value, and to derive a functional equation characterizing this minimum. To ensure that the techniques of the previous section will be applicable, we shall restrict consideration to NSG's with the following property.

Irreducibility Property for NSG's. for any given choice of integers $k_1, k_2, \cdots, k_n$ and $l_1, l_2, \cdots, l_n$ such that $1 \leqq k_i \leqq M_i$ and $1 \leqq l_i \leqq N_i$, for $i = 1, 2, \cdots, n$, the stochastic matrix having $p_{ij}^{k_i l_i}$ as its $i - j$ element is irreducible.

Any stationary strategy S' for the second player may be specified as a point $(u_1^1, \cdots, u_1^{N_1}, u_2^1, \cdots, u_2^{N_2}, \cdots, u_n^1, \cdots, u_n^{N_n})$ in a Euclidean space of $\sum_{i=1}^n N_i$ dimensions; the set Σ of all points corresponding to stationary strategies is clearly compact. Moreover, $g(S')$ is the optimal value of the objective function for a linear program (Problem II) whose data, as expressed in (3) and (4), depend continuously on the qualities u_i^l; thus $g(S')$ is a continuous function of these quantities (see the Appendix), and assumes its minimum at at least one point of Σ. Let such a point be denoted

$$(t_1^1, t_1^2, \cdots, t_1^{N_1}, t_2^1, \cdots, t_2^{N_2}, \cdots, t_n^1, t_n^2, \cdots, t_n^{N_n}),$$

and call the associated stationary strategy T.

It will be proven that T is an optimum strategy for the second player in the NSG. Consider Problem II, with

$$(5) \qquad a_i^k = \sum_{l=1}^{N_i} a_i^{kl} t_i^l, \qquad 1 \leqq i \leqq n, \qquad 1 \leqq k \leqq M_i, \qquad \text{and}$$

$$(6) \qquad p_{ij}^k = \sum_{l=1}^{N_i} p_{ij}^{kl} t_i^l, \qquad 1 \leqq i \leqq n, \qquad 1 \leqq j \leqq n,$$

and let an optimal solution by given by $g^*, v_1^*, \cdots, v_n^*$. We shall derive a set of equations characterizing this solution.

Definition. $\Gamma_i(v_1, \cdots, v_n)$ is the $M_i \times N_i$ matrix game having

$$a_i^{kl} + \sum_{j=1}^n p_{ij}^{kl} v_j$$

as its $k - l$ payoff element. The value of this game is denoted as

$$\text{Val } \Gamma_i(v_1, \cdots, v_n).$$

Theorem 3. For

$$1 \leqq i \leqq n, \tag{7}$$
$$g^* + v_i^* = \mathrm{Val}\,\Gamma_i(v_i^*, \cdots, v_n^*).$$

Proof. By Lemma 1,

$$g^* + v_i^* = \max_{1 \leqq k \leqq M_i} \sum_{l=1}^{N_i} (a_i^{kl} + \sum_{j=1}^{n} p_{ij}^{kl} v_j^*) t_i^l.$$

Thus we have to show that, for each i, $(t_i^1, t_i^2, \cdots, t_i^{N_i})$ is an optimum mixed strategy for $\Gamma_i(v_1^*, \cdots, v_n^*)$. Suppose, to the contrary, that there is a better strategy $w_h^1, \cdots, w_h^{N_h}$ for the game $\Gamma_h(v_1^*, \cdots, v_n^*)$; then

$$g^* + v_h^* > \max_{1 \leqq k \leqq M_h} \sum_{l=1}^{N_h} (a_h^{kl} + \sum_{j=1}^{n} p_{hj}^{kl} v_j^*) w_h^l. \tag{8}$$

Now consider the behavior strategy S' which is the same as T except that $(w_h^1, \cdots, w_h^{N_h})$ replaces $(t_h^1, \cdots, t_h^{N_h})$. Then $g(S')$ is the optimal value of the objective function for Problem II, with a_i^k and p_{ij}^k as given in (5) and (6), except for the case $i = h$; in that case,

$$a_h^k = \sum_{l=1}^{N_h} a_h^{kl} w_h^l, \quad \text{and} \quad p_{hj}^k = \sum_{l=1}^{N_h} p_{hj}^{kl} w_h^l,$$
$$\text{for} \quad 1 \leqq k \leqq M_h, \quad \text{and} \quad 1 \leqq j \leqq n.$$

Now $(g^*, v_1^*, \cdots, v_n^*)$ is a feasible solution to this linear program, but not an optimal one, since the strict inequality (8) violates the condition for optimality given in Lemma 1. Thus $g(S') < g^* = g(T)$, contradicting the definition of T. This contradiction completes the proof.

Theorem 4. If $g + v_i = \mathrm{Val}\,\Gamma_i(v_1, \cdots, v_n)$ and $g^* + v_i^* = \mathrm{Val}\,\Gamma_i(v_i^*, \cdots, v_n^*)$, $1 \leqq i \leqq n$, then $g = g^*$ and there is a constant c such that, for all i, $v_i - v_i^* = c$.

Proof. For each i, let $Y_i = (y_i^1, \cdots, y_i^{N_i})$ be an optimal mixed strategy for the second player in $\Gamma_i(v_1, \cdots, v_n)$, and let $Z_i = (z_i^1, \cdots, z_i^{M_i})$ be an optimal mixed strategy for the first player in $\Gamma_i(v_1^*, \cdots, v_n^*)$. When the strategy pair (Z_i, Y_i) is played in $\Gamma_i(v_1, \cdots, v_n)$, the first player will receive not more than the value of the game, since the second player is using his optimal strategy. Therefore

$$\mathrm{Val}\,\Gamma_i(v_1, \cdots, v_n) = g + v_i \geqq \sum_{k=1}^{M_i} \sum_{l=1}^{N_i} (a_i^{kl} + \sum_{j=1}^{n} p_{ij}^{kl} v_j) z_i^k y_i^l.$$

If the same strategy pair is used in $\Gamma_i(v_1^*, \cdots, v_n^*)$, the first player will receive at least the value of the game, so that

$$g^* + v_i^* \leqq \sum_{k=1}^{M_i} \sum_{l=1}^{N_i} (a_i^{kl} + \sum_{j=1}^{n} p_{ij}^{kl} v_j^*) z_i^k y_i^l.$$

Subtracting the second inequality from the first one obtains

$$(g - g^*) + (v_i - v_i^*) \geqq \sum_{j=1}^{n} (\sum_{k=l}^{M_i} \sum_{l=1}^{N_i} p_{ij}^{kl} z_i^k y_i^l)(v_j - v_j^*).$$

Given these inequalities, one for each i, the remainder of the proof is precisely analogous to the proof of Theorem 2.

Thus, if $g^*, v_1^*, \cdots, v_n^*$ is the essentially unique solution of $g + v_i = \mathrm{Val}\,\Gamma_i(v_1, \cdots, v_n)$, $1 \leqq i \leqq n$, the second player can limit the first player's payoff to g^* by playing a stationary strategy such that he employs an optimal mixed strategy for $\Gamma_i(v_1^*, \cdots, v_n^*)$ whenever state i occurs. By a symmetrical argument, the first player can ensure for himself a payoff of g^* by playing an optimal mixed strategy in $\Gamma_i(v_1^*, \cdots, v_n^*)$ whenever in state i. Our conclusion is that the NSG has g^* as its value, and has a solution in stationary strategies.

4. An Algorithm for the Solution of NSG's

We propose the following iteration process for the solution of (7):

(a) Choose a behavior strategy $\{u_i^l(0)\}$ for the second player.

(b) Given $\{u_i^l(t)\}$, compute $g(t), v_1(t), \cdots, v_n(t)$, the unique solution of:

$$g(t) + v_i(t) = \max_{1 \leqq k \leqq M_i} \sum_{l=1}^{N_i} (a_i^{kl} + \sum_{j=1}^{n} p_{ij}^{kl} v_j(t)) u_i^l(t), \qquad 1 \leqq i \leqq n;$$
$$v_n(t) = 0.$$

This may be done using Howard's algorithm for solving Markovian decision processes.

(c) Given $g(t), v_1(t), \cdots, v_n(t)$, compute $\{u_i^l(t+1)\}$ such that, for each i,

$$u_i(t+1) = (u_i^1(t+1), \cdots, u_i^{N_i}(t+1))$$

is an optimal strategy for player 2 in the game

$$\Gamma_i(t) = \Gamma_i(v_1(t), \cdots, v_n(t)).$$

We shall show that $g(t) \to g^*$, and

$$v(t) = (v_1(t), \cdots, v_n(t)) \to v^* = (v_1^*, \cdots, v_n^*)),$$

where

$$g^* + v_i^* = \text{Val}\, \Gamma_i(v_1^*, \cdots, v_n^*), \qquad 1 \leqq i \leqq n.$$

To do this we first note that

(9) $g(t)$ is monotone decreasing,

which is obvious from the definition of our sequence. Secondly,

(10) the vectors $g(t); v(t)$,

which arise in the sequence are all in a compact set.

The justification for (10) is as follows: if we pick a function $k(i)$, and solve the equations

$$g + v_i = \sum_{l=1}^{n_i} (a_i^{k(i)l} + \sum_{j=1}^{n} p_{ij}^{k(i)l} v_j) u_i^l, \qquad i \neq n, \qquad v_n = 0,$$

for given $\{u_i^l\}$, then it follows from the irreducibility property that the solution to these equations exists, is unique, and is a continuous function of the $\{u_i^l\}$. Because the $\{u_i^l\}$ vary in a compact set, so do the solutions to the system of equations. But the vectors $g(t); v(t)$ which can arise in our sequence are contained in the finite union of compact sets given by all possible functions $k(i)$.

Therefore we may choose a convergent subsequence of vectors $g(t), v(t)$, and let us denote the vector to which they converge by $g^+; v^+$. Consider now the corresponding sequence of vectors

$$u(t + 1) = (u_1(t + 1), \cdots, u_n(t + 1)).$$

Since these vectors vary in a compact set, we may extract a convergent subsequence converging to

$$u^+ = (u_1^+, \cdots, u_n^+).$$

Because $u_i(t + 1)$ is an optimal strategy for player 2 in game $\Gamma_i(t)$, it follows from continuity that u_i^+ is an optimal strategy for player 2 in game

$$\Gamma_i(+) = \Gamma_i(v_1^+, \cdots, v_n^+).$$

Since

$$g(t) + v_i(t) \geqq \text{Val } \Gamma_i(t),$$

it follows from continuity that

$$g^+ + v_i^+ \geqq \text{Val } \Gamma_i(+).$$

If, for some i, $g^+ + v_i^+ > \text{Val } \Gamma_i(+)$, then the solution to problem II with $u = u^+$ would yield $g(u^+) < g^+$. But, since $g^+ \leqq g(u(t + 1))$, by (9) and the continuity, it follows that $g^+ \leqq g(u^+)$. This contradiction establishes that

$$g^+ + v_i^+ = \text{Val } \Gamma_i(+)$$

for every i which proves that the convergent subsequence converges to a solution to our functional equations. Because the solution to the functional equations is unique every convergent subsequence has the same limit, and it follows that the sequence produced in the algorithm converges to a solution to the functional equations.

Appendix: On the Continuity of $g(S')$

Lemma 2. Let

$$(11)\qquad \begin{aligned} &y'A = c \\ &y \geqq 0 \\ &\max\ (y, b) \end{aligned} \qquad \text{and} \qquad (12)\qquad \begin{aligned} &Ax \geqq b \\ &\min\ (c, x) \end{aligned}$$

be a pair of dual linear programs such that both programs are feasible and have m as the common value of the objective function. Further, let the sets

$$X = \{x \mid Ax \geqq b,\ (c, x) = m\}, \qquad \text{and}$$

$$Y = \{y \mid y'A = c,\ y \geqq 0,\ (b, y) = m\}$$

be bounded. Then, for any $\epsilon > 0$ there exists a $\delta > 0$ such that, if A^*, c^* and b^* have respectively the same dimensions as A, c and b, and if

$$\max \begin{Bmatrix} | a_{ij} - a_{ij}^* |, & \text{all } i, j \\ | c_j - c_j^* |, & \text{all } j \\ | b_i - b_i^* |, & \text{all } i \end{Bmatrix} < \delta \tag{13}$$

then the corresponding programming problems (11)* and (12)* have a common optimum value m^*, and $| m - m^* | < \epsilon$.

What we are asserting here are sufficient conditions for the "obvious" fact that the optimum value of a linear programming problem is a continuous function of its data. It has been shown by Williams [15] that these same conditions are necessary and sufficient for the existence of "marginal values" of the given linear program with respect to perturbations of its data.

Proof. Assume A has p rows and q columns. Let M be the matrix with $2p + 2q + 1$ rows and $p + q$ columns given by:

$$M = \begin{bmatrix} 0 & -A \\ A^T & 0 \\ -A^T & 0 \\ -I & 0 \\ -b' & c' \end{bmatrix}$$

Let r be the vector with $2p + 2q + 1$ components given by:

$$r = \begin{bmatrix} b \\ c \\ -c \\ 0 \\ 0 \end{bmatrix}.$$

Solving the original linear programming problem is equivalent to finding a vector z such that $Mz \leqq r$, where the first p coordinates of z give the vector y solving (11), and the last q coordinates give the vector x solving (12). This, of course, is a consequence of the duality theorem. Further, by our hypothesis the set

$$Z = \{z \mid Mz \leqq r\}$$

is bounded.

For any real number a, we define

$$a_+ = \begin{Bmatrix} a & \text{if} & a \geqq 0 \\ 0 & \text{if} & a < 0 \end{Bmatrix}.$$

For any vector v, whose components are v_i, v_+ is the vector whose components are $(v_i)_+$. Now, for any A^*, c^*, and b^*, having the dimensions of A, b and c respectively, let M^* and r^* be defined appropriately, and let $z^* = \begin{pmatrix} y \\ x \end{pmatrix}$ denote a solution of $M^*z^* \leqq r^*$.

We begin by giving the following result, in which $|v|$ denotes the largest element in absolute value of the vector v: there exists a pair of numbers (L, Δ) such that, if

$$\max\ (\max_{i,j} |a_{ij} - a_{ij}^*|, \max_j |c_j - c_j^*|, \max_i |b_i - b_i^*|) < \Delta,$$

then
(i) $M^*z^* \leqq r^*$ is consistent, and
(ii) any solution z^* of (i) satisfies $|z^*| < L$.

The truth of (i) for some Δ_1 is established by the same argument as Lemma 2 of [15]. To prove (ii), assume the contrary. Then we can find a sequence (A_t, c_t, b_t) approaching (A, b, c), and a sequence (z_t), where $M_t z_t \leqq r_t$, such that $|z_t| \to \infty$. Thus we have $M_t w_t \leqq r_t/|z_t|$, where $w_t = z_t/|z_t|$ has norm 1. Therefore there is a convergent sequence of the w_t converging to a vector w of norm 1, and by continuity we have $Mw \leqq 0$. If z is any vector satisfying $Mz \leqq r$, then $M(z + \alpha w) \leqq r$ for arbitrarily large positive α, contradicting the fact that Z is bounded. This contradiction yields the existence of (L, Δ) satisfying (ii). Let $\Delta = \min(\Delta_1, \Delta_2)$ and (i) and (ii) are established.

We shall also invoke the following theorem [9]: let $|\quad|$ stand for the same vector norm as before. Let M be any matrix. Then there exists a constant K, depending just on M, such that, if r is a vector with $Mz \leqq r$ consistent, then for any vector z^* there is a vector z satisfying $Mz \leqq r$ such that

$$|z - z^*| < K|(Mz^* - r)_+|. \tag{14}$$

In using (14), we will let M and r have the meanings of the previous paragraphs, and z^* satisfy $M^*z^* \leqq r^*$.

Observe that, as a vector inequality, we have

$$(Mz^* - r)_+ \leqq ((M - M^*)z^*)_+ + (M^*z^* - r^*)_+ + (r^* - r)_+$$
$$= ((M - M^*)z^*)_+ + (r^* - r)_+ .$$

Therefore,

$$|(Mz^* - r)_+| \leqq |((M - M^*)z^*)_+| + |(r^* - r)_+|. \tag{15}$$

Now, given $\epsilon > 0$, let us choose δ so that $0 < \delta < \Delta$ and

$$\delta(qL + K((p + q)L + 1)\textstyle\sum_j |c_j|) < \epsilon. \tag{16}$$

From (13), (14), and (15) there exists a $z = \begin{pmatrix} y \\ x \end{pmatrix}$ satisfying $Mz \leqq r$ such that

$$|z - z^*| < K(\delta(p + q)L + \delta).$$

Consequently,

$$|x - x^*| \leqq |z - z^*| \leqq K\delta((p + q)L + 1). \tag{17}$$

We also have from (ii) that $|x^*| < L$. Then $|(c^*, x^*) - m| = |(c^*, x^*) - (c, x)| \leqq |(c^* - c, x^*)| + |(c, x^* - x)| < \delta qL + \delta K((p + q)L + 1)\sum |c_j| < \epsilon$. This completes the proof of Lemma 2.

To apply Lemma 2 to the statement in §3 that $g(S')$ is a continuous function of S', it is sufficient to consider problems I and II with the side condition that $v_n = 0$. Problem I obviously has its solution set bounded; and the fact that the optimum solutions for Problem II are bounded is given in §4.

Acknowledgment

The authors wish to thank the referee for pointing out the relevance of the theorems of Kuhn and Aumann.

References

1. AUMANN, R., "Mixed and Behavior Strategies in Infinite Extensive Games," *Advances in Game Theory*, Annals of Mathematics Studies Number 52, Princeton, 1964, pp. 627–650.
2. DERMAN, C. "On Sequential Decisions and Markov Chains," *Management Science*, 9, October 1962, pp. 16–24.
3. D'ÉPENOUX, F., "A Probabilistic Production and Inventory Problem," *Management Science*, *10*, October 1963, pp. 98–108.
4. ——, "Sur un Problème de Production de Stockage dans l'Aleatoire," *Revue Française de Recherche Opérationnelle*, No. *14*, 1960.
5. EVERETT, H., "Recursive Games," *Contributions to the Theory of Games*, Vol. III, 47–78, Princeton, 1957.
6. FREIMER, M., "On Solving a Markovian Decision Problem by Linear Programming," (Unpublished paper, Institute for Defense Analyses, Cambridge, Mass., 12 Dec. 1961).
7. DE GHELLINK, G., "Les Problèmes de Décisions Séquentielles," *Cahiers de Centre d'Étude de Recherche Opérationnelle*, 2, Brussels, 1960.
8. GILLETTE, D., "Stochastic Games with Zero Stop Probabilities," *Contributions to the Theory of Games*, Vol. III, 179–187, Princeton, 1957.
9. HOFFMAN, A. J., "On Approximate Solutions of Systems of Linear Inequalities," *Journal of Research of the National Bureau of Standards*, 49, October 1952, pp. 263–265.
10. HOWARD, R., *Dynamic Programming and Markov Processes*, Technology Press and Wiley, 1960.
11. KUHN, H., "Extensive Games and the Problem of Information," Annals of Mathematics Studies Number 28, Princeton, 1953, pp. 193–216.
12. MANNE, A., "Linear Programming and Sequential Decisions," *Management Science*, 6, April, 1960, pp. 259–267.
13. SHAPLEY, L., "Stochastic Games," *Proc. Nat. Acad. Sci., U.S.A.*, 39, pp. 1095–1100.
14. SIMONNARD, M., *Programmation Linéaire*, Dunod, Paris, 1962.
15. WILLIAMS, A., "Marginal Values in Linear Programming," *Journal of the Society for Industrial and Applied Mathematics*, 11, March, 1963, pp. 82–94.
16. WOLFE, P. AND DANTZIG, G., "Linear Programming in a Markov Chain," *Operations Research*, 10, Sept.-Oct., 1962, pp. 702–710.