stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 122/80 APRIL
(DEPARTMENT OF OPERATIONS RESEARCH)

O.J. VRIEZE

LINEAR PROGRAMMING AND UNDISCOUNTED STOCHASTIC GAMES IN WHICH ONE PLAYER CONTROLS TRANSITIONS

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

1980 Mathematics subject classification: 90D15, 93E25

Linear programming and undiscounted stochastic games in which one player controls transitions *)

by

O.J. Vrieze

ABSTRACT

This paper considers non-cooperative two-person zero-sum undiscounted stochastic games with finite state and action spaces. It is assumed that one player governs the transition rules. We give a linear programming algorithm and show, that an optimal solution to this program corresponds to the value of the game and to optimal stationary strategies for both players. Moreover, this linear programming formulation results in an existence proof of the value and of optimal stationary strategies for both players.

KEY WORDS & PHRASES: *undiscounted stochastic games, linear programming*

---

*) This report will be submitted for publication elsewhere.

1. Introduction and preliminaries

This paper considers non-cooperative two-person zero-sum stochastic games with finite state and action spaces, where the transition probabilities are governed by one player.

More formally, we consider games $\Gamma = \langle S, \{A_n(k), n \in \{1,2\}\}, k \in S, r, P \rangle$, where $S = \{1,\ldots,N\}$ is called the state space: $A_n(k) = \{1,\ldots,m_n(k)\}$ is the set of pure actions for player n in state k; $r : T \to \mathbb{R}$ is a real-valued function with $T = \{(k,i,j);\ k \in S,\ i \in A_1(h),\ j \in A_2(k)\}$ and r is called the payoff function; $P = \{p(\ell \mid k,i,j);\ \ell \in S,\ (k,i,j) \in T\}$ prescribes the transition rules, i.e. $p(\ell \mid k,i,j) \geq 0$ and $\sum_{\ell=1}^{N} p(\ell \mid k,i,j) = 1$, so $p(\ell \mid k,i,j)$ denotes the chance that the system will move to state ℓ, if in state k player 1 chooses pure action $i \in A_1(h)$ and player 2 pure action $j \in A_2(k)$.

Throughout this paper we will assume that $p(\ell \mid k,i,j)$ does not depend on i, i.e. player 2 governs the transition rules. In consequence we will suppress the subscript i in $p(\ell \mid k,i,j)$, e.g. $P = \{p(\ell \mid k,j);\ (\ell,k) \in S \times S,\ j \in A_2(k)\}$.

The course of the play proceeds as usually in stochastic games. We will examine the undiscounted version of such stochastic games and the limit expected average payoff criterium will be used.

The notions of behavioral strategy, stationary strategy, limit expected average payoff, value und (ε-) optimal strategies will be adopted in the usual way.

The proof of the following lemma can be found in STERN [5], BEWLEY and KOHLBERG [1] and PARTHASARATHY & RAGHAVAN [4].

Lemma 1.1. *A two-person zero-sum undiscounted stochastic game with finite state and action space, where the transition probabilities depend on one player only, has a value and both player posses optimal stationary strategies.*

A stationary strategy for player n will be denoted as π_n, where $\pi_n = (\pi_n(1),\ldots,\pi_n(N))$ and $\pi_n(k) = (\pi_n(k,1),\ldots,\pi_n(k,m_n(k)))$ with $\pi_n(k,i) \geq 0$ and $\sum_{i=1}^{m_n(k)} \pi_n(k,i) = 1$.
The set of behavioral (history dependent) strategies for player n will be denoted as Π_n^H.
Π_n will denote the set of stationary strategies for player n.
If π_n is such that for each $k \in S$ there exist a $i_k \in A_n(k)$ with $\pi_n(k,i_k) = 1$, then π_n is called a pure stationary strategy and such a strategy will be notated as π_n^p. Let Π_n^p be the finite set of pure stationary strategies for player n.

For a stationary strategy π_2 of player 2, we denote by $P(\pi_2)$ the $N \times N$-matrix of transition probabilities, where the (k,ℓ)-th element equals $\sum_{j=1}^{m_2(k)} p(\ell \mid k,j).\pi_2(k,j)$.
$Q(\pi_2)$ will denote the Cesaro-limit of $P(\pi_2)$, i.e. $Q(\pi_2) = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{P^k(\pi_2)}{n+1}$, where $P^0 = I$, the unit matrix and $P^k(\pi_2) = P(P^{k-1}(\pi_2))$.

For a pair of stationary strategies (π_1,π_2) the limit expected average payoff will be notated as $V(\pi_1,\pi_2) = (V_1(\pi_1,\pi_2),\ldots,V_N(\pi_1,\pi_2))$, where $V_k(\pi_1,\pi_2)$ corresponds to the game starting in state k.
It is known, that $V(\pi_1,\pi_2) = Q(\pi_2).r(\pi_1,\pi_2)$, where $r(\pi_1,\pi_2)$ is a N-vector with as k-th component:
$\sum_{i=1}^{m_1(k)} \sum_{j=1}^{m_2(k)} \pi_1(k,i).\pi_2(k,j).r(k,i,j)$.

The value of the game will be denoted as $V(\Gamma) = (V_1(\Gamma),\dots,V_N(\Gamma))$. The following lemma, which looks obviously, but needs a precise argument, can be found in VRIEZE [6].

Lemma 1.2. *For a stationary strategy* π_1 (π_2) *of player* 1 (2) *it holds:*

$$\min_{\pi_2\in\Pi_2^H} V(\pi_1,\pi_2) = \min_{\pi_2^P\in\Pi_2^P} V(\pi_1,\pi_2)\ .\ \left(\max_{\pi_1\in\Pi_1^H} V(\pi_1,\pi_2) = \max_{\pi_1^P\in\Pi_1^P} V(\pi_1^P,\pi_2)\right).$$

In [2], RAGHAVAN and FILAR gave an algorithm for solving undiscounted stochastic games, where one player controls the transition rules; solving means finding the value and optimal stationary strategies for the both players. Their algorithm is finite in the sense that it needs a finite number of basic computations.
The procedure of FILAR and RAGHAVAN consists of four steps:
In the first step for each $(\pi_1^P,\pi_2^P)\in\Pi_1^P\times\Pi_2^P$ the payoff $V(\pi_1^P,\pi_2^P)$ is computed. In the second step for each state $k\in S$ a matrix game is constructed from the numbers $V_k(\pi_1^P,\pi_2^P)$. It turns out that the value of this matrix game equals $V_k(\Gamma)$ and that an optimal stationary strategy for player 1 can be constructed by means of optimal actions for these matrix games. In the third step the game is reduced with respect to player 2, i.e. for each state pure actions of player 2 are deleted as long as this does not influence the value of the game.
In the fourth step from the remaining game an optimal stationary strategy for player 2 is constructed by means of solving a LP-problem.

In section 2 we will present an alternative algorithm for solving undiscounted stochastic games in which one player controls the transition law. Our algorithm solves the whole problem in one blow, by means of a LP-formulation of the problem. This LP-problem is of size $(2N+\sum_{k=1}^{N} m_1(k))$ by $(N+2\sum_{k=1}^{N} m_2(k))$.

2. The algorithm

We will first state the algorithm, next show that this linear programming problem has a solution and then prove that this solution corresponds to a solution of the stochastic game.

Consider the following linear programming problem in the variables $(g_1,\dots,g_N),(v_1,\dots,v_N)$, $x_i(k)$, $i \in A_1(k)$, $k \in S$:

$$\max \sum_{k=1}^{N} g_k, \text{ subject to:}$$

$$\text{(i)}\quad g_k - \sum_{\ell=1}^{N} p(\ell \mid k,j) g_\ell \le 0, \qquad k \in S,\ j \in A_2(k)$$

$$\text{(ii)}\quad g_k + v_k - \sum_{i=1}^{m_1(k)} x_i(k).r(k,i,j) - \sum_{\ell=1}^{N} p(\ell \mid k,j).v_\ell \le 0, \quad k \in S,\ j \in A_2(k)$$

$$\text{(iii)}\quad \sum_{i=1}^{m_1(k)} x_i(k) = 1, \qquad k \in S$$

$$x_i(k) \ge 0, \qquad k \in S,\ i \in A_1(k).$$

The dual linear programming in the variables $(w_1,\dots,w_N),y_j(k),z_j(k)$, $k \in S$, $j \in A_2(k)$, is:

$$\min \sum_{k=1}^{N} w_k, \text{ subject to:}$$

$$\text{(j)}\quad \sum_{k=1}^{N} \sum_{j=1}^{m_2(k)} (\delta_{k\ell} - p(\ell \mid k,j)) y_j(k) + \sum_{j=1}^{m_2(\ell)} z_j(\ell) = 1, \qquad \ell \in S$$

$$\text{(jj)}\quad \sum_{k=1}^{N} \sum_{j=1}^{m_2(k)} (\delta_{k\ell} - p(\ell \mid k,j)) z_j(k) = 0, \qquad \ell \in S$$

$$\text{(jjj)}\quad -\sum_{j=1}^{m_2(k)} z_j(k).r(k,i,j) + w_k \ge 0, \qquad k \in S,\ i \in A_1(k)$$

(jjjj) $y_j(k),\ z_j(k) \geq 0,$ $\qquad k \in S,\ j \in A_1(k)$

($\delta_{k\ell} = 1$ if $k = \ell$ and $\delta_{k\ell} = 0$ if $k \neq \ell$).

Lemma 2.1. *Both linear programming problems are feasible and have bounded solutions.*

Proof. Consider the primal problem. Note that $g_\ell = \min_{(k,i,j)\in T} r(k,i,j)$, $\forall \ell \in S$, $v_\ell = 0$, $\forall \ell \in S$ and $x_i(k) = 1$, $i = 1$ and $x_i(k) = 0$, $i > 1$, obeys the conditions (i) to (iiii), so the primal problem is feasible.

Now let $(g,v,x_i(k))$ be a feasable solution. Let the stationary strategy π_1 for player 1 be such that $\pi_1(k,i) = x_i(k)$ and let $\pi_2^p \in \Pi_2^p$ be arbitrary.

Then from (i) and (ii) we get (in vector notation):

$$g \leq P(\pi_2^p).g \tag{2.1}$$

$$g + v \leq r(\pi_1,\pi_2^p) + P(\pi_2^p).v \tag{2.2}$$

From (2.1) we get $g \leq Q(\pi_2^p).g$ and using this result, after multiplying (2.2) from the left by $Q(\pi_2^p)$ yields:

$$g \leq Q(\pi_2^p).r(\pi_1,\pi_2^p) = V(\pi_1,\pi_2^p) \tag{2.3}$$

Now (2.3) shows that g is bounded from above (e.g. by $\max_{(k,i,j)} r(k,i,j)$).

From the duality theorem, it follows that also the dual problem is feasible and has a bounded solution. □

As we already did in the proof of lemma 2.1, we can associate with a set $\{x_i(k);\ k \in S,\ i \in A_1(k)\}$ a stationary strategy $\pi_1(x)$ by defining $\pi_1(x)(k,i) = x_i(k)$.

Lemma 2.2. *Let* $(g,v,x_i(k))$ *be a feasible solution to the primal problem, then*

$$\min_{\pi_2 \in \Pi_2^H} V(\pi_1(x),\pi_2) \geq g.$$

Proof. From (2.3) we get $\min_{\pi_2^p \in \Pi_2^p} V(\pi_1(x),\pi_2^p) \geq g$ and now lemma 1.2 yields the assertion. □

For a feasable solution $(w,y_j(k),z_j(k))$ to the dual program we will define a number of quantities:

$$u_k = \sum_{j=1}^{m_2(k)} z_j(k), \qquad k \in S \text{ and } u = (u_1,\dots,u_N) \tag{2.4}$$

$$S_0 = \{k;\ k \in S \text{ and } u_k = 0\} \tag{2.5}$$

$$\tilde{z}_j(k) = z_j(k)/u_k, \quad k \in S \setminus S_0 \text{ and } j \in A_2(k) \tag{2.6}$$

$$t_k = \sum_{j=1}^{m_2(k)} (y_j(k) + z_j(k)), \ k \in S \text{ and } t = (t_1,\dots,t_N) \tag{2.7}$$

$$\tilde{y}_j(k) = (y_j(k) + z_j(k))/t_k, \quad k \in S,\ j \in A_2(k) \tag{2.8}$$

(from condition (j): $t_k > 0$ if $u_k = 0$)

$\tilde{\pi}_2$ and π_2^*, both stationary strategies for player 2, by

$$\tilde{\pi}_2(k,j) = \tilde{y}_j(k), \quad k \in S,\ j \in A_2(k) \text{ and} \tag{2.9}$$

$$\pi_2^*(k,j) = \tilde{z}_j(k) \text{ for } k \in S \setminus S_0 \text{ and } \pi_2^*(k,j) = \tilde{y}_j(k) \text{ for } k \in S_0. \tag{2.10}$$

Remark 2.3. Note, that from (j) and (jj) we also have

$$\sum_{k=1}^{N} \sum_{j=1}^{m_2(k)} (\delta_{k\ell} - p(\ell \mid k,j))(y_j(k)+z_j(k)) + \sum_{j=1}^{m_2(\ell)} z_j(\ell) = 1, \ \ell \in S, \tag{2.11}$$

which is equivalent to:

$$t_\ell - \sum_{k=1}^{N} p(\ell \mid k,\tilde{\pi}_2).t_k + u_\ell = 1, \qquad \ell \in S, \tag{2.12}$$

Let for a stationary strategy π_2, after suitable renumbering of the states, $P(\pi_2)$ be as:

$$P(\pi_2) = \begin{pmatrix} P_{11}(\pi_2) & & & \\ & \ddots & & 0 \\ 0 & & & P_{\tau\tau}(\pi_2) & \\ P_{\tau+1\ 1}(\pi_2) & \cdots & & P_{\tau+1\ \tau}(\pi_2) & P_{\tau+1\ \tau+1}(\pi_2) \end{pmatrix}$$

$P_{nn}(\pi_2)$ corresponds to the n^{th} ergodic class of $P(\pi_2)$, whose set of states will be notated as $E_n(\pi_2)$, $n \in \{1,\dots,\tau\}$.

$P_{\tau+1\ \tau+1}(\pi_2)$ corresponds to the transient states of $P(\pi_2)$ and this set of transient states will be notated as $T(\pi_2)$.

Remark 2.4. $\sum_{\ell \in E_n(\tilde{\pi}_2)} (t_\ell - \sum_{k \in E_n(\tilde{\pi}_2)} p(\ell \mid k,\tilde{\pi}_2) t_k) = 0$, $n \in \{1,\dots,\tilde{\tau}\}$. (2.13)

This can be seen at once as $\sum_{\ell \in E_n(\tilde{\pi}_2)} p(\ell \mid k,\tilde{\pi}_2) = 1$ for $k \in E_n(\tilde{\pi}_2)$.

Lemma 2.5. (a) $u = u\, P(\pi_2^*)$.

(b) *The transient states for* $P(\pi_2^*)$ *are exactly the states* S_0.

Proof. (a) This follows immediately after inserting definitions (2.4) and (2.10) in (jj).

(b) Note first that summing up condition (j) over $\ell \in N$ yields $\sum_{k=1}^{N} u_k = N$.
From the theory of Markov chains it follows that if $u = u\, P(\pi_2^*)$, with $\sum_{k=1}^{N} u_k = N$, then u can be written as:

$$u = \lambda_1(q_1:0:\dots:0) + \lambda_2(0:q_2:0:\dots:0) + \dots + \lambda_{\tau^*}(0:\dots:0:q_{\tau^*}:0)$$

with $\lambda_n \geq 0$, $\sum_{n=1}^{\tau^*} \lambda_n = N$ and q_n equals the invariant distribution of $P(\pi_2^*)$, $n \in \{1,\dots,\tau^*\}$.

It follows, that, if for $k \in S$, $u_k > 0$, then for some $n \in \{1,\dots,\tau^*\}$, $k \in E_n(\pi_2^*)$ and furthermore $u_\ell > 0$ for all $\ell \in E_n(\pi_2^*)$.

So, if we want to show, that S_0 are exactly the transient states of $P(\pi_2^*)$, it is enough to show, that there does not exist an ergodic class entirely within S_0. Therefore, suppose for some $n \in \{1,\ldots,\tau^*\}$ $E_n(\pi_2^*) \subset S_0$. Summing up (2.12) over $\ell \in E_n(\pi_2^*)$ then, yields for the left hand side (remember remark 2.4):

$$- \sum_{\ell \in E_n(\pi_2^*)} \sum_{k \notin E_n(\pi_2^*)} \sum_{j=1}^{m_2(k)} p(\ell \mid k,j).y_j(k),$$

which leads to a contradiction, as the right hand side is strictly positive. This shows that the assumption $E_n(\pi_2^*) \subset S_0$ was wrong by which the lemma is proved. □

Corollary 2.6. u *can be written as*

$$u = \lambda_1(q_1:0:\ldots:0) + \lambda_2(0:q_2:0:\ldots:0) + \ldots + \lambda_{\tau^*}(0:\ldots:0:q_{\tau^*}:0),$$

with $\lambda_n > 0$, $n \ \{1,\ldots,\tau^*\}$, $\sum_{n=1}^{\tau^*} \lambda_n = N$ *and* q_n *equals the invariant distribution of* $P_{nn}(\pi_2^*)$.

Corollary 2.7. Let $\pi_1^p \in \Pi_1^p$, then

$$\sum_{k=1}^{N} \sum_{j=1}^{m_2(k)} z_j(k).r(k,\pi_1^p,j) = \sum_{n=1}^{\tau^*} \lambda_n V(\pi_1^p,\pi_2)(n).$$

(Here $V(\pi_1^p,\pi_2^*)(n)$ equals the expected average payoff for the pair (π_1^p,π_2^*) with as starting state a state belonging to $E_n(\pi_2^*)$.)

Corollary 2.7 can be checked by inserting the expression for u of corollary 2.6 in the left hand side and remembering

$$\sum_{k \in E_n(\pi_2^*)} q_n(k).r(k,\pi_1^p,\pi_2^*) = V(\pi_1^p,\pi_2^*)(n).$$

From now on $(q,v,x_i(k))$ and $(w,y_j(k),z_j(k))$ will correspond to a dual pair of optimal solutions.

Lemma 2.8. (a) $w_k = 0$ *for* $k \in S_0$.

(b) $\sum_{k \in E_n(\pi_2^*)} w_k = \lambda_n \max_{\pi_1^p} V(\pi_1^p, \pi_2^*)(n)$.

Proof. (a) Follows at once from $u_k = 0$, $k \in S_0$ and (jjj).

(b) Follows at once from corollary 2.7 and lemma 1.2. □

Lemma 2.9. $P(\pi_2^*).g = g$ *and* $P(\tilde{\pi}_2).g = g$.

Proof. From $P(\pi_2^*).g \geq g$, it follows that the equality sign holds for components, belonging to the recurrent states of $P(\pi_2^*)$, i.e. to $S \setminus S_0$. This yields: if $z_j(k) > 0$, then $\sum_{\ell=1}^{N} p(\ell \mid k, j)\ g_\ell = g_k$. (2.14)

From the complementary slackness property we get:

$$\text{if } y_j(k) > 0, \text{ then } \sum_{\ell=1}^{N} p(\ell \mid k, j)\ g_\ell = g_k \qquad (2.15)$$

(2.14) and (2.15) together with the definitions of π_2^* and $\tilde{\pi}_2$ gives the lemma. □

Corollary 2.10. (a) For each $n \in \{1,\ldots,\tau^*\}$, g_k is constant for $h \in E_n(\pi_2^*)$.

(b) For each $n \in \{1,\ldots,\tilde{\tau}\}$, g_k is constant for $k \in E_n(\tilde{\pi}_2)$.

In the following $g(n)$, for $n \in \{1,\ldots,\tau^*\}$ or $n \in \{1,\ldots,\tilde{\tau}\}$, will denote the value of g on $E_n(\pi_2^*)$ or $E_n(\tilde{\pi}_2)$.

Lemma 2.11. *For* $(n_1, n_2) \in \{1,\ldots,\tilde{\tau}\} \times \{1,\ldots,\tau^*\}$ *we have*
either $E_{n_1}(\tilde{\pi}_2) \cap E_{n_2}(\pi_2^*) = E_{n_2}(\pi_2^*)$
or $E_{n_1}(\tilde{\pi}_2) \cap E_{n_2}(\pi_2^*) = \emptyset$.

Proof. Let $k \in E_{n_1}(\tilde{\pi}_2) \cap E_{n_2}(\pi_2^*)$ and let $\ell \in E_{n_2}(\pi_2^*)$, then ℓ and k communicate for $P(\pi_2^*)$. But then, as $\tilde{y}_j(k) > 0$ if $\tilde{z}_j(k) > 0$ it follows that ℓ and k also communicate for $P(\tilde{\pi}_2)$, so $\ell \in E_{n_1}(\tilde{\pi}_2)$. □

Let $D_n = \{n;\ E_n(\tilde{\pi}_2) \supset E_{\tilde{n}}(\pi_2^*)\}$, $\quad n \in \{1,\ldots,\tilde{\tau}\}$

$T_n = E_n(\tilde{\pi}_2) \cap T(\pi_2^*)$, $\quad n \in \{1,\ldots,\tilde{\tau}\}$

$T = \{\tilde{n};\ T(\tilde{\pi}_2) \supset E_{\tilde{n}}(\pi_2^*)\}$

$TT = T(\tilde{\pi}_2) \cap T(\pi_2^*)$.

For a finite set B, we mean by $|B|$ the number of elements of B.

Lemma 2.12. *For* $n \in \{1,\ldots,\tilde{\tau}\}$ *we have*:

$$\sum_{\ell \in E_n(\tilde{\pi})} u_\ell = \sum_{\tilde{n} \in D_n} \lambda_{\tilde{n}} = \sum_{\tilde{n} \in D_n} |E_{\tilde{n}}(\pi_2^*)| + |T_n| + \sum_{\ell \in E_n(\tilde{\pi})} \sum_{k \in T(\tilde{\pi}_2)} p(\ell \mid k, \tilde{\pi}_2) . t_k .$$

Proof. Summing up (2.12) over $\ell \in E_n(\tilde{\pi}_2)$ and using remark 2.4 yields the assertion. □

Remark 2.13. (a) (2.12) for $\ell \in TT$ gives

$$t_\ell = 1 + \sum_{k \in T(\tilde{\pi}_2)} p(\ell \mid k, \tilde{\pi}_2) t_k . \tag{2.16}$$

(b) (2.12) summing up over $E_n(\pi_2^*)$ for $n \in T$ yields:

$$\lambda_n = |E_n(\pi_2^*)| - \sum_{\ell \in E_n(\pi_2^*)} t_\ell + \sum_{\ell \in E_n(\pi_2^*)} \sum_{k \in T(\tilde{\pi}_2)} p(\ell \mid k, \tilde{\pi}_2) . t_k . \tag{2.17}$$

Lemma 2.14. $\max_{\pi_1^p} V(\pi_1^p, \pi_2^*)(n) = g(n)$, $n \in \{1,\ldots,\tau^*\}$.

Proof. From duality theorem and lemma 2.8 (a) and (b) we get:

$$\sum_{\ell \in S} g_\ell = \sum_{\ell \in S} w_\ell = \sum_{n=1}^{\tau^*} \lambda_n \max_{\pi_1^p} V(\pi_1^p, \pi_2^*)(n) =$$

$$= \sum_{\tilde{n}=1}^{\tilde{\tau}} \max_{\pi_1^p} V(\pi_1^p, \pi_2^*)(\tilde{n}) . \sum_{n \in D_{\tilde{n}}} \lambda_n + \sum_{n \in T} \lambda_n \max_{\pi_1^p} V(\pi_1^p, \pi_2^*)(n) \tag{2.18}$$

Now it should be noted from lemma 1.2 and lemma 2.2 that for $n \in \{1,\ldots,\tau^*\}$:

$$\max_{\pi_1^p} V(\pi_1^p, \pi_2^*)(\tilde{n}) \geq V(\pi_1(x), \pi_2^*)(\tilde{n}) \geq g(\tilde{n}) \tag{2.19}$$

Substituting inequality (2.19) in (2.18) yields:

$$\sum_{k=1}^{N} g_k \geq \sum_{\tilde{n}=1}^{\tilde{\tau}} g(\tilde{n}) \sum_{n\in D_{\tilde{n}}} \lambda_n + \sum_{n\in T} g(n).\lambda_n. \qquad (2.20)$$

If we insert in the right hand side of (2.20) the expressions for $\sum_{n\in D_{\tilde{n}}} \lambda_n$, $n \in \{1,\ldots,\tilde{\tau}\}$ and λ_n, $n \in T$ of lemma 2.12 and of (2.17), then, after suitable rearrangings of terms and using the expression (2.16), it follows, that the right hand side of (2.20) exactly equals $\sum_{k=1}^{N} g_k$. But, as $\lambda_n > 0$ for each $n \in \{1,\ldots,\tau^*\}$ this means that in (2.19) the equality sign holds, which proves the lemma. □

Theorem 2.15. *The game has a value and both players posses optimal stationary strategies;* g *is the value of the game,* $\pi_1(x)$ *is an optimal stationary strategy for player* 1 *and* π_1^* *is optimal for player* 2.

Proof. From lemma 2.14 $\max_{\pi_1^p} V_k(\pi_1^p,\pi_2^*) = g_k$ for $k \in S \setminus T(\pi_2^*)$.
But, as by lemma 2.9 $P(\pi_2^*).g = g$, also for the transient states we have $\max_{\pi_1^p} V_k(\pi_1^p,\pi_2^*) = g_k$, $k \in T(\pi_2^*)$.
So we have

$$\max_{\pi_1^p} V(\pi_1^p,\pi_2^*) \geq g. \qquad (2.21)$$

Using lemma 1.2 (2.21) gives:

$$\max_{\pi_1\in\Pi_1^H} V(\pi_1,\pi_2^*) \geq g \qquad (2.22)$$

(2.22) combining with lemma 2.2 shows the theorem. □

3. Some remarks

Remark 3.1. If in each state player 1 has only 1 action, then the game reduces to a minimizing Markov decision problem. In that case our algorithm reduces to the algorithm proposed by Hordijk and Kallenberg [3].

Parts of their proofs could be projected on our problem; in particular the fact, that S_0 is exactly the set of transient states for $P(\pi_2^*)$ could be proven for both cases in an analogue way.

The problem of proving the optimality of π_2^* is essential different. Namely following their line of argument, would mean showing that π_2^* is "optimal" against all π_1^p such that $\pi_1^p \in \underset{k=1}{\overset{N}{X}} \alpha_k$, where $\alpha_k = \{i_k;\ i_k \in A_1(k) \text{ and } \pi_1(x)(k,i_k) > 0\}$. Clearly this is not enough for showing the optimality of π_2^*.

Remark 3.2. If it is known in advance that for each $\pi_2 \in \Pi_2$ the transition probability matrix $P(\pi_2)$ is such, that the set of all states form an ergodic class, then as well the algorithm as the proofs can be considerably simplified.

Namely the algorithm becomes:

max g (g is a number now) subject to:

(i) $g + v_k - \sum_{i=1}^{m_2(k)} x_i(k) r(k,i,j) - \sum_{\ell=1}^{N} p(\ell \mid k,j) v_\ell \le 0,\quad k \in S,\ j \in A_2(k)$

(ii) $\sum_{i=1}^{m_2(k)} x_i(k) = 1$ and

(iii) $x_i(k) \ge 0$.

The dual of this linear programming problem is:

$\min \sum_{k=1}^{N} w_k$, subject to:

(j) $\sum_{k=1}^{N} \sum_{j=1}^{m_2(k)} z_j(k) = 1$

(jj) $\sum_{k=1}^{N} \sum_{j=1}^{m_2(k)} (\delta_{k\ell} - p(\ell \mid k,j)) z_j(k) = 0, \qquad \ell \in S$

(jjj) $-\sum_{j=1}^{m_2(k)} z_j(k).r(k,i,j) + w_k \geq 0, \qquad i \in A_1(k),\ k \in S$

(jjjj) $z_j(k) \geq 0.$

In this case the stationary strategy π_2 with $\pi_2(k,j) = z_j(k) / \sum_{j=1}^{m_2(k)} z_j(k)$ for each j and k is optimal for player 2, if the $z_j(k)$'s belong to an optimal solution of the dual program.

References

1. *T. Bewley* and *E. Kohlberg*, "On stochastic games with stationary optimal strategies", Math.Oper.Res. 3 (1978), pp.104-125.

2. *J. Filar* and *T.E.S. Raghavan*, "An algorithm for solving an undiscounted stochastic game in which one player controls transitions", Research Memorandum, University of Illinois, Chicago, 1979.

3. *A. Hordijk* and *L.C.M. Kallenberg*, "Linear Programming and Markov decision chains", Management Sci. Vol.25 (1979), pp.352-362.

4. *T. Parthasarathy* and *T.E.S. Raghavan*, "An order field property for stochastic games when one player controls the transitions", Game Theory Conference, Cornell, Ithaca, June 1978.

5. *M.A. Stern*, "On stochastic games with limiting average payoff", Ph.D. dissertation submitted to the university of Illinois, Circle Campus, Chicago 1975.

6. *O.J. Vrieze*, "Characterization of optimal stationary strategies in undiscounted stochastic games", Math. Centre Report BW 102/79, Amsterdam, 1979.