

Using Lemke's Algorithm to Solve Two-Person Single-Controller Stochastic Games

T. E. S. Raghavan^{1,2} and Zamir Syed³

Abstract

Given a two-person, nonzero-sum stochastic game where the second player controls the transitions, we formulate a linear complementarity problem $LCP(q, M)$ whose solution gives a Nash equilibrium pair of stationary strategies under the limiting average payoff criterion. The matrix M will be constructed so that Lemke's algorithm will process it. We will also do the same for a special class of N -person stochastic games called polystochastic games.

Keywords: Stochastic games, Linear Complementarity Problem, Lemke's algorithm

1. Introduction

Discounted stochastic games were first introduced by Shapley (see [1]). In a stochastic game Γ , we have a finite set of states $S = \{1, 2, \dots, s\}$, and for each state $t \in S$ there are two finite sets $A(t) = \{1, 2, \dots, a_t\}$ and $B(t) = \{1, 2, \dots, b_t\}$ called the action sets for players I and II respectively. For each triple (t, a, b) with $a \in A(t)$ and $b \in B(t)$ there is a pair of immediate costs $(r^1(t, a, b), r^2(t, a, b))$ for players (I,II) as well as a probability distribution $p[t, a, b]$ on the set S . Given an initial state $t_0 \in S$, the game is played as follows. The players simultaneously choose actions $a^0 \in A(t_0)$ and $b^0 \in B(t_0)$ resulting in the costs $r^1(t_0, a^0, b^0)$ and $r^2(t_0, a^0, b^0)$ to be paid by players I and II respectively. The system moves to a new state t_1 according to $p[t_0, a^0, b^0]$, and the players again choose actions $a^1 \in A(t_1)$ and $b^1 \in B(t_1)$. Accordingly

⁰Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago; e-mail: ter@uic.edu

¹Partially Funded by NSF Grant DMS 930-1052, and DMS970-4951

²Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago; e-mail: ter@uic.edu

the costs $r^1(t_1, a^1, b^1)$ and $r^2(t_1, a^1, b^1)$ are again paid by their respective players, and the game moves to a new state t_2 according to $p[t_1, a^1, b^1]$. The game continues indefinitely generating two streams of costs $r^1(t_i, a^i, b^i)$ and $r^2(t_i, a^i, b^i)$ as $i = 1, 2, \dots$. A general strategy for a player would be a function from the set of all possible histories into the set of probability distributions over the player's action space. A general strategy can therefore be very complicated but nevertheless, given a pair of strategies (π, ρ) for both players, we can evaluate the expected β -discounted values (for $0 \leq \beta < 1$):

$$\begin{aligned} (1) \quad \phi_\beta^1(\pi, \rho)(t_0) &= \sum_{n=0}^{\infty} \beta^n r_n^1(t_0, \pi, \rho) \\ (2) \quad \phi_\beta^2(\pi, \rho)(t_0) &= \sum_{n=0}^{\infty} \beta^n r_n^2(t_0, \pi, \rho) \end{aligned}$$

where t_0 is the starting state, and $r_n^1(t_0, \pi, \rho)$ ($r_n^2(t_0, \pi, \rho)$) is the expected cost to player I(II) at the n th stage when the players are using π and ρ . Thus we treat $\phi_\beta^1(\pi, \rho)$ and $\phi_\beta^2(\pi, \rho)$ as payoff vectors indexed by the starting state. We say that a pair of strategies (π^*, ρ^*) comprise a β -discounted equilibrium point (in the sense of Nash) if for any pair of strategies (π, ρ) the following conditions hold simultaneously:

$$\begin{aligned} (3) \quad \phi_\beta^1(\pi^*, \rho^*)(t) &\leq \phi_\beta^1(\pi, \rho^*)(t) \\ (4) \quad \phi_\beta^2(\pi^*, \rho^*)(t) &\leq \phi_\beta^2(\pi^*, \rho)(t) \end{aligned}$$

for $t = 1, \dots, s$. Another, generally more difficult to "handle", payoff criterion is the *limiting average cost*. Under the same notation as above we explicitly write:

$$\begin{aligned} (5) \quad \phi^1(\pi, \rho)(t_0) &= \limsup_{T \rightarrow \infty} \frac{1}{T+1} \sum_{x=0}^T r_x^1(t_0, \pi, \rho) \\ (6) \quad \phi^2(\pi, \rho)(t_0) &= \limsup_{T \rightarrow \infty} \frac{1}{T+1} \sum_{x=0}^T r_x^2(t_0, \pi, \rho) \end{aligned}$$

to represent the expected limiting average cost to the respective players when they are using the pair of strategies (π, ρ) . By dropping the β 's we get the analogues of (3) and (4) under this second cost criterion:

$$\begin{aligned} (7) \quad \phi^1(\pi^*, \rho^*)(t) &\leq \phi^1(\pi, \rho^*)(t) \\ (8) \quad \phi^2(\pi^*, \rho^*)(t) &\leq \phi^2(\pi^*, \rho)(t) \end{aligned}$$

It should be noted that adding a fixed constant to all the payoffs does not affect the equilibrium points of the game. Hence, without loss of generality, we can assume that all the immediate costs are strictly positive.

A strategy π (ρ) for player I (II) is called *stationary* if it only depends on the current state. Such strategies can generally be mixed. Nevertheless,

they are much simpler than behavioral strategies, thus giving hope for finite algorithms to solve games which possess *stationary equilibrium points* (equilibrium pairs with both strategies being stationary). Here the notation to be used will be that for any $a \in A(t)$, $\pi(t, a)$ is the probability that player I chooses action a in state t under the strategy π . Similarly for any $b \in B(t)$, $\rho(t, b)$ is the probability that player II chooses action b in state t under the strategy ρ .

Given a pair of stationary strategies (π, ρ) , we write $r(\pi, \rho)$ to be the $s \times 1$ column vector whose k th coordinate is given by $\sum_{i=1}^{a_k} \sum_{j=1}^{b_k} r(k, i, j)\pi(k, i)\rho(k, j)$. We also write $P(\pi, \rho)$ to be the $s \times s$ matrix whose (k, l) th entry is given by $\sum_{i=1}^{a_k} \sum_{j=1}^{b_k} p[k, i, j]_l \pi(k, i)\rho(k, j)$. Note that $P(\pi, \rho)$ is simply the transition matrix of the Markov chain induced by π and ρ . Accordingly, $P^*(\pi, \rho)$ will denote the stationary matrix $\lim_{T \rightarrow \infty} \frac{1}{T+1} \sum_{t=0}^T [P(\pi, \rho)]^t$.

2. Single-Controller Stochastic Games

A stochastic game is of the *single-controller* type if the transition probabilities only depend on one of the players. In this paper we will assume the controlling player to be player II. In terms of our notation this means that:

$$(9) \quad p[t, a, b] = p[t, 1, b] \quad \forall t \in S, \forall (a, b) \in (A(t), B(t))$$

Hence we can write $p[t, a, b] = p[t, b]$. The following stochastic game is an example of a single-controller game.

In each state player I is the row player and player II is the column player. The ordered pairs represent the immediate costs to the players, the first coordinate for player I and the second coordinate for player II. In both states the transition probability vector depends only on the column (i.e. player II's choice).

It is known that the class of single-controller stochastic games (both zero-sum and nonzero-sum) possess stationary equilibria. That is, for single-controller games, there always exists a pair of stationary strategies (π^*, ρ^*) satisfying (7) and (8) above. Furthermore, the *orderfield property* holds for single-controller games (that is, the solution to the game lies in the same Archimedean ordered field as the data). The orderfield property is a key evidence to suggest that a finite algorithm to solve the game may exist. Indeed finite step algorithms were given for *discounted* and *irreducible undiscounted* single-controller nonzero sum stochastic games [Nowak and Raghavan (1982)]. An efficient algorithm for solving the discounted case with a single LCP was recently given by Mohan, Neogy, and Parthasarathy (1997).

3. The Linear Complementarity Problem

The linear complementarity problem can be stated as follows. Given a vector $q \in R^n$ and a matrix $M \in R^{n \times n}$, find a vector z such that:

$$(10) \quad w = q + Mz$$

$$(11) \quad z, w \geq 0$$

$$(12) \quad z^T w = 0$$

The system (10)-(12) is usually denoted $LCP(q, M)$. It can be shown that the LCP is a generalization of the well known LP (linear program). In a historic work of C. E. Lemke[1964], a simplex-like pivoting algorithm is given to process LCP's. Unfortunately, the algorithm does not always find a solution to a given LCP. There are, however, certain classes of matrices M for which Lemke's algorithm will process $LCP(q, M)$.

It is well known that an ordinary, zero-sum matrix game can be solved by a single LP. The same was done for zero-sum, single-controller, stochastic games for both discounted and limiting average costs. For nonzero-sum, bi-matrix games, linear programming is not enough. However, it has been shown

that they can be solved by a single LCP. Furthermore, Lemke's algorithm will always generate a solution for the class of LCP's arising from bimatrix games. This gives hope for the nonzero-sum, single-controller stochastic games. Using the discounted cost criteria, it has been shown that such games can be solved by a single LCP and that Lemke's algorithm will process the LCP (Mohan, Neogy, and Parthasarathy [1997]). In this paper we will show that under the limiting average cost criterion, these games can be solved by a single, Lemke-processible, LCP as well.

4. Formulation

We begin by analyzing the conditions (7) and (8) for the single-controller game.

We first note that if in (7) and (8) we restrict π^* , ρ^* , π , and ρ to only stationary strategies, then (π^*, ρ^*) would still be an equilibrium point as defined without the restriction. This can be seen by noticing that when one player's strategy is fixed to a stationary one, the other player is faced with an MDP. From dynamic programming literature (Blackwell[1962]) it is well-known that $\phi^i(\pi, \rho) = P^*(\pi, \rho)r^i(\pi, \rho)$ for $i = 1, 2$. Since the transitions only depend on player II's actions, $P(\pi, \rho)$ can be written as $P(\rho)$. Now (7) reduces to componentwise inequality

$$(13) \quad P^*(\rho^*)r(\pi^*, \rho^*) \leq P^*(\rho^*)r(\pi, \rho^*) \quad \forall \pi.$$

The standard reduction of (13) would entail "cancelling" the $P^*(\rho^*)$ from both sides of the inequality and simply requiring that

$$(14) \quad r(\pi^*, \rho^*) \leq r(\pi, \rho^*) \quad \forall \pi.$$

Here (14) is a stronger condition than (13), and in many cases it is much more than we actually need. After all, (13) is satisfied as long as $r(\pi^*, \rho^*)_t \leq r(\pi, \rho^*)_t, \forall \pi$ is true for all recurrent states t in the Markov chain induced by ρ^* (because $P^*(\rho^*)$ is zero in the columns corresponding to transient states) whereas (14) requires that it be true for all states.

As for player II's side of the equilibrium condition, there is really no reduction. The requirement in (8), unlike in (7), can change entirely with each strategy ρ . Thus, the task of formulating a mathematical program to solve

(7) and (8) will involve a combination of the traditional MDP (coming from player II's side of the game) and simple linear inequalities (from player I's side of the game). Fortunately limiting average MDP's have a nice linear programming formulation. Therefore, in constructing our LCP, we will embed both the LP arising from (8) and the linear inequalities arising from (14) into a single LCP.

Next we consider the MDP arising from player I fixing his strategy to a stationary strategy π . When player II chooses action b in state t , the immediate cost incurred is given by $\tilde{r}(t, b) = \sum_{i=1}^{a_t} \pi(t, i) r^2(t, i, b)$. The transitions of the MDP are the same as those of the original game since π has no influence on them. Using the LP formulation for limiting average MDP's, player II's best reply to π comes as a solution to the following pair of dual LP's:

Primal

$$\begin{aligned} & \text{Maximize} && \frac{1}{s} \sum_{i=1}^s \phi(i) \\ & \text{s.t. (15)} && \phi(i) - \sum_{j=1}^s p[i, b]_j \phi(j) \leq 0 && \forall i \in S, \forall b \in B(i) \\ & && \text{(16)} && \phi(i) + u(i) - \sum_{j=1}^s p[i, b]_j u(j) \leq \tilde{r}(i, b) && \forall i \in S, \forall b \in B(i) \\ & && \text{(17)} && \phi(i), u(i) \text{ unrestricted} && \forall i \in S \end{aligned}$$

Dual

$$\begin{aligned} & \text{Minimize} && \sum_{i=1}^s \sum_{b \in B(i)} \tilde{r}(i, b) x_{ib} \\ & \text{s.t. (18)} && \sum_{b \in B(j)} x_{jb} + \sum_{b \in B(j)} y_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j y_{ib} = \frac{1}{s} && \forall j \in S \\ & && \text{(19)} && \sum_{b \in B(j)} x_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j x_{ib} = 0 && \forall j \in S \\ & && \text{(20)} && x_{ib} \geq 0, y_{ib} \geq 0 && \forall i \in S, \\ & && && && \forall b \in B(i) \end{aligned}$$

In this setup we have y_{ib} and x_{ib} complementary to (15) and (16) respectively. Conversely we have $\phi(j)$ and $u(j)$ complementary to (18) and (19) respectively. Suppose we have an optimal solution for both programs, say (ϕ^*, u^*, x^*, y^*) . Then player II's optimal strategy ρ^* (against π) would be extracted as follows:

$$\rho^*(i, b) = \begin{cases} x_{ib}^* / \sum_{c \in B(i)} x_{ic}^* & \text{when } \sum_{c \in B(i)} x_{ic}^* > 0 \\ y_{ib}^* / \sum_{c \in B(i)} y_{ic}^* & \text{otherwise} \end{cases}$$

One can verify (using (18)) that ρ^* is well defined. Also we have $\phi^*(i) = \phi^2(\pi, \rho^*)$. A key property of this pair of LP's is that those states $i \in S$ for which $\sum_{c \in B(i)} x_{ic}^* = 0$ are *transient* in the Markov chain induced by ρ^* . Next we make a few minor adjustments to (15)-(20) so that they can be put into LCP form.

The first point is that we can replace (19) with

$$(21) \quad \sum_{b \in B(j)} x_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j x_{ib} \geq 0 \quad \forall j \in S$$

Clearly (19) implies (21). To see the other direction we just sum (21) over j to get

$$\begin{aligned} 0 &\leq \sum_{j \in S} [\sum_{b \in B(j)} x_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j x_{ib}] \\ &= \sum_{j \in S} \sum_{b \in B(j)} x_{jb} - \sum_{j \in S} \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j x_{ib} \\ &= \sum_{j \in S} \sum_{b \in B(j)} x_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} \sum_{j \in S} p[i, b]_j x_{ib} \\ &= \sum_{j \in S} \sum_{b \in B(j)} x_{jb} - \sum_{j \in S} \sum_{b \in B(j)} x_{jb} \\ &= 0 \end{aligned}$$

Hence (21) implies (19). After replacing (19) with (21) and remembering that equality will hold for any optimal solution, we can tighten another restriction. Namely, the complementary slackness variables to (19), the $u(i)$'s, can be restricted to being nonnegative. To see this, note that changing from u to \tilde{u} where $\tilde{u}(i) = u(i) + \theta, \forall i \in S, \forall \theta \in R$ has no effect on the objective function of the primal LP, does not change (16) in any way. Further the complementary slackness with (21) is untouched. Hence we can change the condition on u in (17) to $u(i) \geq 0 \forall i \in S$.

The next changes will have no effect on the optimal solutions of the the LP's. Remember that we are assuming $r(t, a, b) > 0, \forall (t, a, b)$. Let $m = \min_{(t, a, b)} r(t, a, b)$. Consider the pair (ϕ, \tilde{u}) where $\phi(i) = m, \forall i \in S$ and $\tilde{u} = 0, \forall i \in S$. On substituting (ϕ, \tilde{u}) into the primal LP, it is immediate that it is a feasible solution. It is well known that an optimal ϕ for this LP will dominate all feasible ϕ in every coordinate. Hence, any optimal solution ϕ^*, u^* will satisfy $\phi^*(i) \geq \phi(i) = m > 0$. Therefore we can reduce the feasible set of the primal even further by requiring $\phi(i) \geq 0, \forall i \in S$. Finally, since we know that the optimal ϕ^* has all positive coordinates, by complementary slackness we must have equality in (18) for any optimal pair (x^*, y^*) . Hence we can change (18) to

$$(22) \quad \sum_{b \in B(j)} x_{jb} + \sum_{b \in B(j)} y_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j y_{ib} \geq \frac{1}{s} \quad \forall j \in S$$

In other words, having strict inequality in the j th inequality of (22) would only hurt ϕ by bringing $\phi(j)$ to 0. One additional adjustment (the importance of which will be seen later) will be to change (16) to

$$(16') \quad \phi(i) + u(i) - \sum_{j=1}^s p[i, b]_j u(j) \leq \tilde{r}(i, b) + \sum_{i=1}^s \sum_{b \in B(i)} x_{ib}$$

Here we have simply added the fixed quantity $\sum_{i=1}^s \sum_{b \in B(i)} x_{ib}$ to $\tilde{r}(i, b)$ for $\forall i \in S, \forall b \in B(i)$. Since $x_{ib} \geq 0$, we have that at a complementary solution $\phi(i) > 0, \forall i \in S$. Using complementary slackness in the dual LP and summing (22) over $j \in S$ we get $\sum_{i=1}^s \sum_{b \in B(i)} x_{ib} = 1$. Hence the only effect of changing (16) to (16') is that the optimal ϕ value will have all coordinates increased by 1. From dynamic programming we know that adding a fixed constant to all the immediate costs has no effect on the optimal strategy.

Putting in all the changes gives the following set of inequalities:

$$\begin{aligned}
\phi(i) - \sum_{j=1}^s p[i, b]_j \phi(j) &\leq 0 && \forall i \in S, \forall b \in B(i) \\
\phi(i) + u(i) - \sum_{j=1}^s p[i, b]_j u(j) &\leq \tilde{r}(i, b) + \sum_{i=1}^s \sum_{b \in B(i)} x_{ib} && \forall i \in S, \forall b \in B(i) \\
\sum_{b \in B(j)} x_{jb} + \sum_{b \in B(j)} y_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j y_{ib} &\geq \frac{1}{s} && \forall j \in S \\
\sum_{b \in B(j)} x_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j x_{ib} &\geq 0 && \forall j \in S \\
\phi(i), u(i), x_{ib}, y_{ib} &\geq 0 && \forall i \in S, \forall b \in B(i)
\end{aligned}$$

Any solution to these inequalities which maintains the complementary conditions from the LP's will provide an optimal stationary strategy for player II when player I fixes his strategy to an arbitrary stationary π .

Next we will construct a set of complementary inequalities to take care of (13). Recall that (13) will be valid if the vector inequality (14) is true on those coordinates corresponding to recurrent states of the Markov chain induced by ρ^* (again assume that π^*, ρ^* satisfy (7) and (8)). Consider the following set of inequalities:

$$\begin{aligned}
(23) \quad \sum_{b \in B(i)} r^1(i, a, b) \rho^*(i, b) &\geq v(i) && \forall i \in S, \forall a \in A(i) \\
(24) \quad \sum_{a \in A(i)} \tilde{z}_{ia} &\geq 1 && \forall i \in S \\
(25) \quad \tilde{z}_{ia}, v(i) &\geq 0 && \forall i \in S, \forall a \in A(i)
\end{aligned}$$

along with the complementary conditions

$$\begin{aligned}
(26) \quad \tilde{z}_{ia} [\sum_{b \in B(i)} r^1(i, a, b) \rho^*(i, b) - v(i)] &= 0 && \forall i \in S, \forall a \in A(i) \\
(27) \quad v(i) [\sum_{a \in A(i)} \tilde{z}_{ia} - 1] &= 0 && \forall i \in S
\end{aligned}$$

Suppose (23)-(27) are satisfied by some (v, z) . If $v(i) > 0$ for all $i \in S$ then by (27) we would have $\sum_{a \in A(i)} \tilde{z}_{ia} = 1$ for $\forall i \in S$. Define the stationary strategy $\tilde{\pi}$ by $\tilde{\pi}(i, a) = \tilde{z}_{ia}$. Fixing i and summing (26) over $a \in A(i)$ gives $r^1(\tilde{\pi}, \rho^*)_i = v(i)$. Substituting this into (23) yields $r^1(\tilde{\pi}, \rho^*)_i \geq r^1(\pi, \rho^*)_i$ for all stationary π . Hence, (14) is satisfied by $\pi^* = \tilde{\pi}$. Unfortunately this implication was a result of $v(i) > 0$ for all $i \in S$. Such a condition is

not necessarily true although with a small adjustment this problem can be alleviated. We replace (23) with

$$(28) \quad \sum_{i=1}^s \sum_{a \in A(i)} \tilde{z}_{ia} + \sum_{b \in B(i)} r^1(i, a, b) \rho^*(i, b) \geq v(i) \quad \forall i \in S, \forall a \in A(i)$$

and accordingly replace (26) with

$$(29) \quad \tilde{z}_{ia} [\sum_{i=1}^s \sum_{a \in A(i)} \tilde{z}_{ia} + \sum_{b \in B(i)} r^1(i, a, b) \rho^*(i, b) - v(i)] = 0 \quad \forall i \in S, \\ \forall a \in A(i)$$

From (24) we know that for each $i \in S$ there is some $a[i] \in A(i)$ with $\tilde{z}_{ia[i]} > 0$. Therefore, using $\tilde{z}_{ia[i]}$'s complementary condition in (29), we have that $v(i) > 0$. Now $v(i) > 0$ implies that $\sum_{a \in A(i)} \tilde{z}_{ia} = 1$ (using (27)). Thus $\sum_{i=1}^s \sum_{a \in A(i)} \tilde{z}_{ia} = s$ so that we have only added a constant to all the inequalities of (26) in order to get (28). It is easy to check that this adjustment does not affect the properties of $\tilde{\pi}$.

We are now ready to define an LCP whose solutions correspond to equilibrium points. We begin by writing the complementary inequalities needed.

$$\begin{aligned} w_{ia}^1 &= \sum_{i=1}^s \sum_{a \in A(i)} \pi_{ia} + \sum_{b \in B(i)} r^1(i, a, b) x_{ib} - v(i) \\ w_{ib}^2 &= \sum_{j=1}^s p[i, b]_j \phi(j) - \phi(i) \\ w_{ib}^3 &= \sum_{i=1}^s \sum_{b \in B(i)} x_{ib} + \sum_{a \in A_i} \pi_{ia} r^2(t, i, b) + \sum_{j=1}^s p[i, b]_j u(j) - u(i) - \phi(i) \\ w_j^4 &= \sum_{b \in B(j)} x_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j x_{ib} \\ w_j^5 &= -\frac{1}{s} + \sum_{b \in B(j)} x_{jb} + \sum_{b \in B(j)} y_{jb} - \sum_{i=1}^s \sum_{b \in B(i)} p[i, b]_j y_{ib} \\ w_i^6 &= -1 + \sum_{a \in A(i)} \pi_{ia} \\ w_{ia}^1 &\geq 0, \pi_{ia} \geq 0, \pi_{ia} w_{ia}^1 = 0, \forall i \in S, \forall a \in A(i) \\ w_{ib}^2 &\geq 0, y_{ib} \geq 0, y_{ib} w_{ib}^2 = 0, \forall i \in S, \forall b \in B(i) \\ w_{ib}^3 &\geq 0, x_{ib} \geq 0, x_{ib} w_{ib}^3 = 0, \forall i \in S, \forall b \in B(i) \\ w_j^4 &\geq 0, u(j) \geq 0, u(j) w_j^4 = 0, \forall j \in S \\ w_j^5 &\geq 0, \phi(j) \geq 0, \phi(j) w_j^5 = 0, \forall j \in S \\ w_i^6 &\geq 0, v(i) \geq 0, v(i) w_i^6 = 0, \forall i \in S \end{aligned}$$

This set of complementary inequalities will be referred to as \mathcal{K} . Next we will define z, M , and q (recall that $w = q + Mz$). However, we advise the reader to first examine the example that follows as it will give a clearer picture as to how (30)-(42) can be written as an LCP. We write

$$z = \left(\pi_{11}, \dots, \pi_{sa_s}, y_{11}, \dots, y_{sb_s}, x_{11}, \dots, x_{sb_s}, \right. \\ \left. \phi(1), \dots, \phi(s), u(1), \dots, u(s), v(1), \dots, v(s) \right)^T$$

In order to label certain subsets of the index set of the matrices involved, we will use the pattern of $z = (\pi, y, x, \phi, u, v)$. Let n be the number of coordinates in z . The matrix $n \times n$ matrix M will be constructed by various partitions of its rows and columns. To refer to an entry of M we will use the labels of z . Thus we can speak of the entry in "row x_{ib} " and "column $\phi(j)$ " of M . The first partition is to write

$$(30) \quad M = \begin{bmatrix} \mathcal{R} & \mathcal{A} \\ \mathcal{B} & 0 \end{bmatrix}$$

where \mathcal{R} is square and contains the indices (π, y, x) . The 0 simply represents the zero matrix and by default it contains the indices (ϕ, u, v) . We next write

$$(31) \quad \mathcal{R} = \begin{bmatrix} \mathcal{C}_1 & 0 & \mathcal{D} \\ 0 & 0 & 0 \\ \mathcal{E} & 0 & \mathcal{C}_2 \end{bmatrix}$$

where the three way split is partitioned as $\pi|y|x$ for the rows of \mathcal{R} as well as its columns. We define all entries of \mathcal{C}_1 and \mathcal{C}_2 to be equal to 1. The (π_{ia}, x_{ib}) th entry of \mathcal{D} is $r^1(i, a, b)$ for $\forall i \in S, \forall a \in A(i), \forall b \in B(i)$. The other entries of \mathcal{D} are 0. The (x_{ib}, π_{ia}) th entry of \mathcal{E} is $r^2(i, a, b)$ for $\forall i \in S, \forall a \in A(i), \forall b \in B(i)$. The other entries of \mathcal{E} are 0. This completes the definition of \mathcal{R} . We now partition the rows and columns of \mathcal{A} by $\pi|y|x$ and $\phi|u|v$ respectively:

$$(32) \quad \mathcal{A} = \begin{bmatrix} 0 & 0 & \mathcal{F} \\ \mathcal{P}_1 & 0 & 0 \\ \mathcal{G} & \mathcal{P}_2 & 0 \end{bmatrix}$$

The $(\pi_{ia}, v(i))$ th entry of \mathcal{F} is -1 for $\forall i \in S, \forall a \in A(i)$. The rest of the entries of \mathcal{F} are 0. If $i \neq j$ then the $(y_{ib}, \phi(j))$ th entry of \mathcal{P}_1 is given by $p[i, b]_j$. If $i = j$ then the $(y_{ib}, \phi(j))$ th entry is $p[i, b]_j - 1$. \mathcal{P}_1 and \mathcal{P}_2 are actually identical. Formally we have that the $(x_{ib}, u(j))$ th entry of the former is the same as the $(y_{ib}, \phi(j))$ th entry of the latter. The $(x_{ib}, \phi(i))$ th entry of \mathcal{G} is -1 for $\forall i \in S, \forall b \in B(i)$. The other entries of \mathcal{G} are all 0. To complete the definition of M we define $\mathcal{B} = -\mathcal{A}^T$.

To complete the construction of the LCP we need to define the vector q . Like z , q is also an $n \times 1$ vector. We set all the coordinates of q to 0 with the exception of the indices (u, v) . Those coordinates of q in u will have value $-\frac{1}{s}$. The coordinates in v will have value -1 .

Under this construction of q , M , and z , it is easy to check that conditions (10), (11), and (12) are equivalent to \mathcal{K} . Before proceeding we write the q , M , and z for the stochastic game given above. We have

$$q = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -\frac{1}{2} \\ -\frac{1}{2} \\ 0 \\ 0 \\ -1 \\ -1 \end{bmatrix}, z = \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{21} \\ \pi_{22} \\ y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{23} \\ x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \\ x_{23} \\ \phi(1) \\ \phi(2) \\ u(1) \\ u(2) \\ v(1) \\ v(2) \end{bmatrix}$$

5. Main Results

Given a two-player, nonzero-sum, single-controller, stochastic game Γ , the following algorithm is suggested:

Algorithm 1

1. Input the data $s, A_1, \dots, A_s, B_1, \dots, B_s, r^1(i, a, b), r^2(i, a, b), p[i, b]$
2. Construct q and M as specified above.
3. Use Lemke's algorithm to process LCP(q, M)
4. From the solution z obtained in step 3, set $\pi^*(i, a) = \pi_{ia}, \forall i \in S, \forall a \in A(i)$ and set $\rho^*(i, b)$ according to the following rule:

$$\rho^*(i, b) = \begin{cases} x_{ib} / \sum_{c \in B(i)} x_{ic} & \text{if } \sum_{c \in B(i)} x_{ic} > 0 \\ y_{ib} / \sum_{c \in B(i)} y_{ic} & \text{otherwise} \end{cases}$$

Theorem 1: The pair (π^*, ρ^*) obtained in step 4 of the Algorithm 1 will constitute an equilibrium point of Γ .

Verifying Theorem 1 requires proving that step 3 does indeed provide a complementary solution of LCP(q, M) and further that the pair (π^*, ρ^*) satisfies conditions (7) and (8). We will prove these in the next two lemmas:

Lemma 1: Lemke's algorithm will provide a solution to LCP(q, M).

Proof: We will show that LCP(q, M) is feasible and that the matrix M is copositive plus, that is:

$$\begin{aligned} (33) \quad & z^T M z \geq 0, \forall z \geq 0 \\ (34) \quad & z^T M z = 0 \Rightarrow (M + M^T)z = 0, \forall z \geq 0 \end{aligned}$$

To show that (33) is satisfied by M , we simply use the partition in (30). Writing $z = [z_1, z_2]$ in accordance with the partition in (30) we get $z^T M z = z_1^T \mathcal{R} z_1 + z_1^T (\mathcal{A} + \mathcal{B}^T) z_1 = z_1^T \mathcal{R} z_1$ as $\mathcal{A} = -\mathcal{B}^T$. Since \mathcal{R} has nonnegative entries, $z_1^T \mathcal{R} z_1 \geq 0$. To show (34), assume that $z^T M z = 0$ is true. Since $z^T M z = z_1^T \mathcal{R} z_1$ we have that $z_1^T \mathcal{R} z_1 = 0$. Now using the partition of (31) and splitting $z_1 = [z_{11}, z_{12}, z_{13}]$ accordingly, we have $0 = z_1^T \mathcal{R} z_1 = z_{11}^T \mathcal{C}_1 z_{11} + z_{13}^T (\mathcal{E} + \mathcal{D}) z_{11} + z_{13}^T \mathcal{C}_2 z_{13}$. Since $\mathcal{E} + \mathcal{D} \geq 0$, and since \mathcal{C}_1 and \mathcal{C}_2 have all entries positive, we can conclude that $z_{11} = 0$ and $z_{13} = 0$. From this is easy to see now that $(M + M^T)z = 0$.

The feasibility of $\text{LCP}(q, M)$ entails showing that (10) and (11) together have at least one point (z, w) . Of course we need only specify z as $w = q + Mz$ is readily computable. We will now give a value of z by giving the values of its coordinates (π, y, x, ϕ, u, v) . Let ρ be an arbitrary stationary strategy for player II and let $P(\rho)$ be the transition matrix for the induced Markov chain. As before, let $P^*(\rho)$ be the stationary matrix of $P(\rho)$ and let $D(\rho) = [I - P(\rho) + P(\rho)^*]^{-1} - P(\rho)^*$ be the *deviation matrix* of $P(\rho)$. Let m be the number of ergodic classes in $P(\rho)$ and let $E_i, 1 \leq i \leq m$ be the classes themselves. Let F denote the set of transient states. Thus $F \cup \bigcup_{i=1}^m E_i = S$ and all the sets are disjoint. For the vector z , let $\pi_{ia} = 1$ for $\forall i \in S, a \in A(i)$ and set

$$x_{jb} = \left(\frac{1}{s} \sum_{i \in S} P^*(\pi)_{ij} \right) \rho_{jb}, \forall j \in S, \forall b \in B(j)$$

$$y_{jb} = \left(\frac{1}{s} \sum_{i \in S} D(\pi)_{ij} + \sum_{i \in S} \gamma_i P^*(\pi)_{ij} \right) \rho_{jb}, \forall j \in S, \forall b \in B(j)$$

where

$$\gamma_i = \begin{cases} 0 & i \in F \\ \max_{l \in E_j} \left(-\frac{1}{s} \sum_{k \in E_j} D(\pi)_{kl} / \sum_{k \in E_j} P(\pi)_{kl}^* \right) & i \in E_j, 1 \leq j \leq m \end{cases}$$

Set all other coordinates of z equal to zero. This choice of x and y was taken from Kallenberg where it is shown that such a solution is feasible in the inequalities of \mathcal{K} involving x_{ib} and y_{ib} . That the rest of the inequalities are satisfied is simple to check. Thus we have $z \geq 0$ and $q + Mz \geq 0$. Feasibility of $\text{LCP}(q, M)$ and M being copositive plus imply that a complementary solution exists and that Lemke's algorithm will find one. This complete the proof. \diamond

Remark: Given a stationary strategy ρ for player II, a feasible solution for $\text{LCP}(q, M)$ was constructed. The choice of (x, y) in this solution actually preserves optimality. That is, an optimal ρ would preserve complementarity in $\text{LCP}(q, M)$. This shows that given an equilibrium pair (π^*, ρ^*) , a corresponding solution z^* to $\text{LCP}(q, M)$ can be computed. Furthermore, applying step 4 of the algorithm to this value of z^* will give back the pair (π^*, ρ^*) .

Lemma 2: The pair (π^*, ρ^*) obtained in step 4 of the algorithm satisfies conditions (7) and (8).

In the proof we assume now that z is a solution to the LCP in the algorithm.

Proof: We first note that in a complementary solution to $\text{LCP}(q, M)$, we must have $v(i) > 0, \forall i \in S$ so that $\sum_{a \in A(i)} \pi_{ia} = 1, \forall i \in S$. So ρ^* is indeed a stationary strategy for player I. When player I fixes this strategy, the conditions in \mathcal{K} are sufficient to show that ρ^* is optimal in the resulting MDP. Thus, (8) is satisfied.

Next we show that (7) holds. Since $\sum_{i \in S} \sum_{b \in B(i)} x_{ib} = 1$, there must exist a subset $T \subseteq S$ for which $\sum_{b \in B(t)} x_{tb} > 0$ for all $t \in T$ (recall that these are precisely the recurrent states of the Markov chain induced by ρ^*). Now using the complementarity conditions with π_{ia} we get for each $t \in T$

$$(35) \quad \sum_{a \in A(t)} \sum_{b \in B(t)} \pi^*(t, a) r^1(t, a, b) x_{tb} = v(t)$$

Since for each $t \in T$ we have $\rho^*(t, b) = x_{tb}$ we can write

$$(36) \quad \sum_{a \in A(t)} \sum_{b \in B(t)} \pi^*(t, a) r^1(t, a, b) \rho_{tb}^* = \bar{v}(t), \forall t \in T$$

where $\bar{v}(t) = v(t) / \sum_{b \in B(t)} x_{tb}$. But this together with the inequalities in \mathcal{K} imply that $r(\pi^*, \rho^*)_t \leq r(\pi, \rho^*)_t \forall \pi, \forall t \in T$. If $t \notin T$ then $\sum_{b \in B(t)} x_{tb} > 0$, and we cannot conclude that necessarily $r(\pi^*, \rho^*)_t \leq r(\pi, \rho^*)_t \forall \pi$ will hold. But this won't matter because the states outside T are transient in the Markov chain induced by ρ^* . Thus, the matrix $P^*(\rho)$ vanishes in columns outside t . From this we can conclude that (7) does indeed hold. This completes the proof. \diamond

Any solution of $\text{LCP}(q, M)$ is a solution to the game Γ . Also, the proof of Theorem I in no way uses the existence of stationary equilibria for single-controller games and is entirely constructive thus giving a new proof for the existence of stationary equilibria in this class of stochastic games.

Given a solution $z = (\pi, y, x, \phi, u, v)$ to $\text{LCP}(q, M)$, the optimal pair of strategies (π^*, ρ^*) are obtained from the algorithm after some trivial arithmetic operations. We also have $\phi^2(\pi^*, \rho^*) = \phi - \bar{e}$ where \bar{e} is the $s \times 1$ column vector of with all coordinates 1. The only work needed is to compute $\phi^1(\pi^*, \rho^*) = P^*(\rho^*)r^1(\pi^*, \rho^*)$. Here $r^1(\pi^*, \rho^*) = v - s\bar{e}$. Thus it is clear that Γ possesses the orderfield property.

6. Solving Polystochastic Games

In this section we will mimic the work of the previous sections in order to give a constructive proof of the existence of stationary equilibria in a spe-

cial class of N -person stochastic games, namely the class of single-controller polystochastic games. This class was introduced in Mohan, Neogy, and Parthasarathy (1997). We will try to make the notation as similar to the previous sections as possible.

Here we have N players $\mathcal{N} = \{\tau_1, \tau_2, \dots, \tau_N\}$ each of which possesses an action set $A^i(t)$ for player τ_i in each state t in the state space $S = \{1, \dots, s\}$. The game is played just as before. Given a starting state t_0 , all N players simultaneously choose actions $(a_1^0, a_2^0, \dots, a_N^0)$, $a_i^0 \in A^i(t_0)$, whence come N immediate costs $r^i(t_0, a_1^0, a_2^0, \dots, a_N^0)$, $i = 1, \dots, N$ where τ_i pays r^i . The game moves to a new state t_1 according to a probability distribution $p[t_0, a_1^0, \dots, a_N^0]$ on S and the process continues indefinitely. Given a collection of strategies (π_1, \dots, π_N) for the players, we again write:

$$(37) \quad \phi^i(\pi_1, \dots, \pi_N)(t_0) = \limsup_{T \rightarrow \infty} \frac{1}{T+1} \sum_{x=0}^T r_x^i(t_0, \pi_1, \dots, \pi_N), i = 1, \dots, N$$

where $r_x^i(\pi_1, \dots, \pi_N)$ is the expected cost to τ_i on the x th day when the players use (π_1, \dots, π_N) and the starting state is t_0 . We say that an N -tuple of strategies $(\pi_1^*, \dots, \pi_N^*)$ constitute an equilibrium point if for any N -tuple of strategies (π_1, \dots, π_N) we have

$$(38) \quad \phi^i(\pi_1^*, \dots, \pi_N^*) \leq \phi^i(\pi_1^*, \dots, \pi_N^* | \pi_i), i = 1, \dots, N$$

where $(\pi_1^*, \dots, \pi_N^* | \pi_i)$ means to replace π_i^* with π_i and to keep all other strategies fixed. We again point out that if we restrict $(\pi_1^*, \dots, \pi_N^*)$ and (π_1, \dots, π_N) to stationary strategies then (38) will still yield an equilibrium point in the sense of the definition without the restriction.

In a single-controller polystochastic game a simplified cost structure is assumed. This simplification (along with the single-controller condition) will allow for (38) to be written as linear inequalities with complementarity conditions. We will assume that τ_N controls the transitions of the game so that $p[t, a_1, \dots, a_N] = p[t, a_N]$. To help distinguish player τ_N in the notation we write $B(t) = A^N(t)$ and refer to τ_N 's strategy as ρ instead of π_N . We also write $\mathcal{M} = \{\tau_1, \dots, \tau_{N-1}\}$. As in polymatrix games, we assume that the costs can be split as

$$(39) \quad r^i(t, a_1, \dots, a_N) = \sum_{j \in \mathcal{N}, j \neq i} R_{ij}(t, a_i, a_j), \forall i \in \mathcal{N}.$$

Here $R_{ij}(t, a_i, a_j)$ can be thought of as a partial cost paid by τ_i as a result of the mutual action (a_i, a_j) of τ_i and τ_j . Hence the cost functions r^i are

completely determined by the partial costs R_{ij} . If $N = 2$ it is easy to check the the two-person game of the previous section can be formulated to fit into this model.

The following set of inequalities will provide an equilibrium point for such a game. The verification is nearly identical to that of the previous sections (with obvious modifications).

$$\begin{aligned}
w_i^1(t, a) &= \sum_{j \in S} \sum_{c \in A(j)} \pi_i(j, c) + \sum_{k \in \mathcal{M}, k \neq i} \sum_{c \in A^k(t)} R_{ik}(a, c) \pi_k(t, c) + \sum_{b \in B} R_{iN}(a, b) x_{tb} - v^i(t) \\
w^2(t, b) &= \sum_{j \in S} p[t, b]_j \phi(j) - \phi(t) \\
w^3(t, b) &= \sum_{j \in S} \sum_{c \in B(j)} x_{jb} + \sum_{k \in \mathcal{M}} \sum_{c \in A^k(t)} R_{ik}(a, c) \pi_k(t, c) + \sum_{j \in S} p[t, b]_j u(j) - u(t) - \phi(t) \\
w^4(t) &= \sum_{b \in B(t)} x_{tb} - \sum_{j \in S} \sum_{b \in B(t)} p[t, b]_j x_{tb} \\
w^5(t) &= -\frac{1}{s} + \sum_{b \in B(t)} x_{tb} + \sum_{b \in B(t)} y_{tb} - \sum_{j \in S} \sum_{b \in B(t)} p[t, b]_j y_{tb} \\
w_i^6(t) &= -1 + \sum_{a \in A^i(t)} \pi^i(t, a) \\
w_i^1(t, a) &\geq 0, \pi_i(t, a) \geq 0, \pi_i(t, a) w_i^1(t, a) = 0, \forall i \in \mathcal{M}, t \in S, \forall a \in A^i(t) \\
w^2(t, b) &\geq 0, y_{tb} \geq 0, y_{tb} w^2(t, b) = 0, \forall t \in S, \forall b \in B(t) \\
w^3(t, b) &\geq 0, x_{tb} \geq 0, x_{tb} w^3(t, b) = 0, \forall t \in S, \forall b \in B(t) \\
w^4(t) &\geq 0, u(t) \geq 0, u(t) w^4(t) = 0, \forall t \in S \\
w^5(t) &\geq 0, \phi(t) \geq 0, \phi(t) w^5(t) = 0, \forall t \in S \\
w_i^6(t) &\geq 0, v^i(t) \geq 0, v^i(t) w_i^6(t) = 0, \forall i \in \mathcal{M}, \forall t \in S
\end{aligned}$$

In this set of inequalities, call it $\bar{\mathcal{K}}$, the notation is slightly different from that of \mathcal{K} . For $i \in \mathcal{M}, t \in S$, and $a \in A^i(t)$, $\pi_i(t, a)$ is the probability that player τ_i chooses action a when the system is in state t . Player τ_N 's strategy ρ is extracted from (x, y) just as before.

Next we construct $\text{LCP}(q, M)$ whose solution corresponds to a solution of $\bar{\mathcal{K}}$. Let a_t^i be the number of elements in the set $A^i(t)$, and b_t the number of elements in $B(t)$. We begin by defining the coordinates of z .

$$\pi_i = \begin{bmatrix} \pi_i(1, 1) \\ \pi_i(1, 2) \\ \cdot \\ \cdot \\ \cdot \\ \pi_i(s, a_s^i) \end{bmatrix}, i \in \mathcal{M}, \text{ and } x = \begin{bmatrix} x_{11} \\ x_{12} \\ \cdot \\ \cdot \\ \cdot \\ x_{sb_s} \end{bmatrix}, y = \begin{bmatrix} y_{11} \\ y_{12} \\ \cdot \\ \cdot \\ \cdot \\ y_{sb_s} \end{bmatrix}.$$

$$\phi = \begin{bmatrix} \phi(1) \\ \phi(2) \\ \vdots \\ \vdots \\ \phi(s) \end{bmatrix}, u = \begin{bmatrix} u(1) \\ u(2) \\ \vdots \\ \vdots \\ u(s) \end{bmatrix}, \text{ and } v^i = \begin{bmatrix} v^i(1) \\ v^i(2) \\ \vdots \\ \vdots \\ v^i(s) \end{bmatrix}, i \in \mathcal{M}.$$

Let

$$v = \begin{bmatrix} v^1 \\ v^2 \\ \vdots \\ \vdots \\ v^{N-1} \end{bmatrix} \text{ and } \pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \vdots \\ \pi_{N-1} \end{bmatrix}.$$

Finally we write z as

$$z = \begin{bmatrix} \pi \\ y \\ x \\ \phi \\ u \\ v \end{bmatrix}.$$

Just as before, we will use the coordinates of z to serve as markers for the partitions (and coordinates) of M and q . We first partition the rows and columns of M by $[\pi y x | \phi u v]$ and write

$$(40) \quad M = \begin{bmatrix} \mathcal{R} & \mathcal{A} \\ \mathcal{B} & 0 \end{bmatrix}$$

The matrix \mathcal{R} contains the immediate cost information. Next we partition \mathcal{R} by $[\pi_1 | \pi_2 | \dots | \pi_{N-1} | y | x]$ and write

$$(41) \quad \mathcal{R} = \begin{bmatrix} \mathcal{E}_1 & \mathcal{R}_{12} & \dots & \mathcal{R}_{1N-1} & 0 & \mathcal{R}_{1N} \\ \mathcal{R}_{21} & \mathcal{E}_2 & \dots & \mathcal{R}_{2N-1} & 0 & \mathcal{R}_{2N} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \mathcal{R}_{N-11} & \mathcal{R}_{N-12} & \dots & \mathcal{E}_{N-1} & 0 & \mathcal{R}_{N-1N} \\ 0 & 0 & \dots & 0 & 0 & 0 \\ \mathcal{R}_{N1} & \mathcal{R}_{N2} & \dots & \mathcal{R}_{NN-1} & 0 & \mathcal{E}_N \end{bmatrix}$$

For $i, j \in \mathcal{M}, i \neq j$ we define the entry in the $\pi_i(t, a)$ th row and the $\pi_j(t, c)$ th column of \mathcal{R}_{ij} to be $R_{ij}(t, a, c)$ where $t \in S, a \in A^i(t), c \in A^j(t)$, and we define all other entries of \mathcal{R}_{ij} to be zero. The 0's in \mathcal{R} represent zero matrices of appropriate dimensions, and for each $i \in \mathcal{N}$, \mathcal{E}_i has all entries equal to one. For each $i \in \mathcal{M}$ we define the entry in the $\pi_i(t, a)$ th row and the x_{tb} th column of \mathcal{R}_{iN} to be $R_{iN}(t, a, b)$ where $t \in S, a \in A^i(t), b \in B(t)$, and we define all other entries of \mathcal{R}_{iN} to be zero. Similarly for each $j \in \mathcal{M}$ we define the entry in the x_{tb} th row and $\pi_j(t, a)$ th column of \mathcal{R}_{Nj} to be $R_{Nj}(t, b, a)$ where $t \in S, a \in A^i(t), b \in B(t)$, and we define all other entries of \mathcal{R}_{Nj} to be zero. This completes the construction of \mathcal{R} . In order to define \mathcal{A} we partition its rows by $[\pi|y|x]$ and its columns by $[\phi|u|v]$ as

$$(42) \quad \mathcal{A} = \begin{bmatrix} 0 & 0 & \mathcal{F} \\ \mathcal{P}_1 & 0 & 0 \\ \mathcal{G} & \mathcal{P}_2 & 0 \end{bmatrix}$$

For each $i \in \mathcal{M}, t \in S$, and $a \in A(t)$ we define the entry in the $\pi_i(t, a)$ th row and $v^i(t)$ th column of \mathcal{F} to be equal to -1 , and we define all other entries of \mathcal{F} to be zero. For each pair $t, t' \in S$ with $t \neq t'$ define the entry in the y_{tb} row and $\phi(t')$ th column of \mathcal{P}_1 to be $p[t, b]_{t'}$. For $t \in S$ define the entry in the y_{tb} th row and the $\phi(t)$ th column of \mathcal{P}_1 to be $p[t, b]_t - 1$. We define $\mathcal{P}_2 = \mathcal{P}_1$. For each $t \in S$ and $b \in B(t)$ we define the entry in the x_{tb} th row and $\phi(t)$ th column of \mathcal{G} to be -1 , and we define all other entries of \mathcal{G} to be zero. This completes the definition of M . For the vector q we set all coordinates in (π, y, x, ϕ) to zero, in u to $-\frac{1}{s}$, and in v to -1 .

It is easy to check that with this construction of q, M , and z , (10), (11), and (12) together are equivalent to $\bar{\mathcal{K}}$. It can also be verified that M is copositive plus and that $\text{LCP}(q, M)$ is feasible (the verification is virtually identical to that of the previous sections). Given a solution z^* to $\text{LCP}(q, M)$, the equilibrium strategies $(\pi_1^*, \dots, \pi_{N-1}^*)$ can be immediately read off. The strategy for τ_N , namely ρ^* can be extracted from (x^*, y^*) just as before. Thus we have:

Algorithm 2

1. Input the data $s, A^i(t), R_{ij}(t, a, b), p[t, b]$
2. Construct q and M as specified above.
3. Use Lemke's algorithm to process $\text{LCP}(q, M)$

4. From the solution z^* obtained in step 3, $(\pi_1^*, \dots, \pi_{N-1}^*, \rho^*)$ is an equilibrium point with $\pi_i^*(t, a)$, $i \in \mathcal{M}$ being the probability that player τ_i chooses action $a \in A^i(t)$ when in state t , and $\rho^*(t, b)$ being the probability that player N chooses action $b \in B(t)$ when in state t where $\rho^*(t, b)$ is defined by the following rule:

$$\rho^*(t, b) = \begin{cases} x_{tb}^* / \sum_{c \in B(t)} x_{tc}^* & \text{if } \sum_{c \in B(t)} x_{tc}^* > 0 \\ y_{tb}^* / \sum_{c \in B(t)} y_{tc}^* & \text{otherwise} \end{cases}$$

Summing up the results we have:

Theorem 2: Algorithm 2 will solve any single-controller polystochastic game.

Just like Theorem 1, Theorem 2 is entirely constructive in that it does not require the knowledge of the existence of a stationary equilibrium point. Furthermore, it is easy to see that this class of games possesses the orderfield property. On applying our results to the previous example, we arrive at the solution $\pi_{11} = 1, \pi_{12} = 0, \pi_{21} = 0.7, \pi_{22} = 0.3, x_{11} = 0.5, x_{12} = 0, x_{21} = 0.3, x_{22} = 0.2, x_{23} = 0$, and $y_{ib} = 0 \forall (i, b)$. This yields the equilibrium strategies $\pi^*(1, 1) = 1, \pi^*(1, 2) = 0, \pi^*(2, 1) = 0.7, \pi^*(2, 2) = 0.3$ for player I and $\rho^*(1, 1) = 1, \rho^*(1, 2) = 0, \rho^*(2, 1) = 0.6, \rho^*(2, 2) = 0.4, \rho^*(2, 3) = 0$ for player II.

References

Blackwell, D. [1962] : *Discrete Dynamic Programming*. Annals of Mathematical Statistics **33**, 719-726.

Hordijk, A. and L.C.M. Kallenberg [1981] : "Linear programming and Markov games", Game Theory and Mathematical Economics, O. Meschling, D. Pallaschke Editors, North Holland, Amsterdam, :307-319.

Kallenberg, L. C. M. [1983]: *Linear Programming and Finite Markovian Control Problems*. Mathematical Centre Tract 148, Centre for Mathematics and Computer Science, Amsterdam.

Lemke, C. E. [1964] : *Bimatrix equilibrium points and mathematical programming*. Management Sci. **11**, 681-689.

Lemke C.E. and Howson, Jr. J.J. [1964] : Equilibrium points of bimatrix games, J. Soc. Indust. Appl. Math., **12**: 413-423.

S. R. Mohan, S. K. Neogy and T. Parthasarathy [1997] :, "Linear Complementarity and Discounted Polystochastic Games when One Player Controls Transitions", To appear in the Proceedings of the International Conference on Complementarity Problem (SIAM Publication 1997), Editors M. C. Ferris and J. S. Pang.

Nowak, A.S. and T.E.S. Raghavan [1993] : A finite step algorithm via a bimatrix game to a single controller non-zero sum stochastic game, Mathematical Programming, 59: 249-259

Parthasarathy, T. and T.E.S. Raghavan [1981] : An orderfield property for stochastic games when one player controls transition probabilities, J. Optimization Theory and Appl. 33:375-392.

Vrieze, O.J. [1981] : Linear programming and undiscounted stochastic game in which one player controls transitions. OR Spektrum 3: 29-35.