Linear Complementarity and Discounted Switching Controller Stochastic Games

T. A. Schultz

Communicated by P. L. Yu

Abstract. The class of discounted switching controller stochastic games can be solved in one step by a linear complementarity program (LCP). Following the proof of this technical result is a discussion of a special formulation and initialization of a standard LCP pivoting algorithm which has, in numerical experiments, always terminated in a complementary solution. That the LCP algorithm as formulated always finds a complementary solution has not yet been proven, but these theoretical and experimental results have the potential to provide an alternative proof of the ordered field property for these games. Numerical experimentation with the reformulated LCP is reviewed.

Key Words. Ordered field property, complementary pivots, linear complementarity, discounted switching controller, stochastic games.

1. Introduction

Stochastic games and their solution algorithms have received considerable theoretical and applied attention; see Raghavan and Filar (Ref. 1) for a survey. A recurring theme has been development of one-step solution methods (i.e., solving a single mathematical program solves the game). Effective one-step methods have typically required structural assumptions about the game (e.g., a single controller or additive rewards and state independent transitions). Considerable attention, including development of finite solution algorithms, has been focused on switching controller stochastic games (e.g., see Refs. 2–5), where it is assumed that only one decision maker's choice of action determines the transition probabilities for the game next state, but the controlling decision maker varies from state to state.

1Associate Professor, School of Business Administration, Augusta College, Augusta, Georgia.
This paper reduces the solution of discounted switching controller stochastic games to the solution of a single linear complementarity program (LCP). An appropriate formulation and initialization of a standard pivoting scheme [Lemke (Ref. 6), Cottle and Dantzig (Ref. 7)] for the LCP have always led to a complementary solution in numerical experimentation, but a proof that a complementary solution will always be found has not yet been developed. Together, the proof of the LCP approach and the numerical evidence provide a heuristic solution method for these games and have the potential for an alternative proof that discounted switching controller games possess the ordered field property.

Following definitions and background in the remainder of Section 1, Section 2 contains the LCP result, Section 3 describes the LCP formulation and an example, and Section 4 reviews results of experimentation.

A two person, zero-sum stochastic game is a competitive, dynamic decision process involving two decision makers, DM1 and DM2. Parameters which define the process are: a finite state space \( S = \{1, \ldots, N\} \); for each state \( s \in S \), finite action sets \( A(s) = \{1, \ldots, m_s\} \) for DM1 and \( B(s) = \{1, \ldots, n_s\} \) for DM2; rewards \( R_{a,b}(s) \) for \( s \in S, a \in A(s), b \in B(s) \); and transitions \( Q_{a,b}(s,t) \) for \( s, t \in S, a \in A(s), b \in B(s) \). As notation, let \( R(s) \) with entries \( R_{a,b}(s) \) denote the \( m_s \times n_s \) matrix of rewards in state \( s \in S \), and let \( Q(s,t) \) with entries \( Q_{a,b}(s,t) \) denote the \( m_t \times n_t \) matrix of transitions from state \( s \in S \) to state \( t \in S \).

At each stage in an infinite horizon, DM1 and DM2 independently choose respective actions \( a \in A(s) \) and \( b \in B(s) \) based on the game parameters and the history of actions taken and states visited (including the current state \( s \in S \)). As a consequence of the action choices, DM2 pays DM1 the amount \( R_{a,b}(s) \), and the decision makers begin the next stage in state \( t \in S \) with probability \( Q_{a,b}(s,t) \) where the game continues.

When the decision makers discount future rewards, it is sufficient for them to choose actions based upon stationary strategies which ignore the game history, except for the current state. For DM1, a stationary strategy \( f \in F_S = \left\{ f_a(s) \mid s \in S, a \in A(s), f_a(s) \geq 0, \sum_{a \in A(s)} f_a(s) = 1 \right\} \) indicates that action \( a \in A(s) \) should be chosen with probability \( f_a(s) \) when the process is in state \( s \in S \). For DM2, stationary strategies \( g \in G_S \) are defined similarly. As notation, for \( f \in F_S \) or \( g \in G_S \), let \( f(s) \) or \( g(s) \) denote the \( m_s \)-dimensional or \( n_s \)-dimensional column vector with elements \( f_a(s) \) or \( g_b(s) \).

With a fixed discount factor \( \alpha \in [0, 1) \), each pair of strategies \( f \) and \( g \) determines an expected discounted game reward \( \Phi_{fg}(\alpha ; s) \) to DM1 for each starting state \( s \in S \) which can be represented as an infinite discounted sum of