LP Formulation of Asymmetric Zero-Sum Stochastic Games

Lichun Li and Jeff Shamma

Abstract— **This paper provides an efficient linear programming (LP) formulation of asymmetric two player zero-sum stochastic games with finite horizon. In these stochastic games, only one player is informed of the state at each stage, and the transition law is only controlled by the informed player. Compared with the LP formulation of extensive stochastic games whose size grows polynomially with respect to the size of the state and the size of the uninformed player's actions, our proposed LP formulation has its size to be linear with respect to the size of the state and the size of the uninformed player, and hence greatly reduces the computational complexity. A travelling inspector problem is used to demonstrate the efficiency of the proposed LP formulation.**

I. INTRODUCTION

Different from the asymmetric repeated games introduced by Aumann and Maschler in 1960's [1]–[3] where the same game is played repeatedly, stochastic games can change the game according to some transition law. If the transition law is Markovian [4], or Markovian controlled by the informed player [5], then stochastic games have values. Recently, [6] generalized the results to the case when players observe signals rather than actions, and showed that the value of the stochastic game existed if one of the players was fully informed and controlled the transition of the state. A more general case when the players have common information but neither one can have full knowledge of the other one's full information is discussed in [7]–[9], which is out of the scope of this paper.

The existence of values in stochastic games means that there is at least one equilibrium such that no player has incentive to deviate from his strategy given that others do not deviate from theirs. But computing the value and the optimal strategy remains an important problem for making game-theoretic solution concepts operational. The most closely related work is in the area of repeated games and the area of symmetric stochastic games. When both players are informed of the game, the value of repeated games can be solved through a polynomial time algorithm [10]. If only one player is informed, while Ponssard and Sorin provided a linear programming algorithm to solve a single shot game [11], Gilpin and Sandholm showed that computing the value of the repeated game with infinite horizon is non-convex [12], [13]. References [14]–[17] analyzed finite-stage asymmetric repeated games in extensive form, and give a LP formulation of the extensive games whose size is linear over the extensive tree size. Based on the idea of the extensive forms of finite-stage asymmetric repeated games, [18], [19] developed a decomposition algorithm to obtain recursive solutions to network interdiction games which had special nested imperfect information structure. In the area of symmetric stochastic games, [20] provided a linear program based algorithm to compute the value and the optimal strategies of symmetric stochastic games when the transition is controlled by only one player.

This paper focuses on computing the values and the optimal strategies for finite horizon, two player, zero sum stochastic games with one player fully informed and controlling the transition. At the beginning of every stage, the state of nature (game) is chosen according to a transition law which relies on the previous state and the previous action of player 1. The chosen state is, then, observed by player 1 only. Player 1 and 2 simultaneously choose their actions which are observed by both players before proceeding to the next stage. For a finite horizon case, stochastic games can be written in an extensive form , and hence be solved through linear programming. The size of the linear program, however, grows exponentially over the length of the horizon with the base to be the product of the size of the states, the size of player 1's actions and the size of player 2's actions [16], [17].

This paper proposes an efficient linear program (LP) to compute the values and the optimal strategies of asymmetric stochastic games where the transition law is controlled by the informed player, and the size of the proposed linear program grows only linearly over the size of the states and the size of the uninformed player's actions but exponentially over the length of the horizon with the base to be the size of the informed player's actions. Compared with the linear program derived from the extensive form, our proposed linear program dramatically reduces the computational complexity by cutting the size of the linear program to be linear over the size of the states, and the size of the uninformed player's actions. Therefore, the proposed linear programming formulation of the asymmetric stochastic games is especially useful when the actions of the informed player are simple, such as on or off, play or quit, in or out, etc., while the state of the nature and the opponent's actions are complicated.

II. MATHEMATICAL MODEL

The mathematical model of stochastic games in this paper is mainly adopted from [4] and [2].

Lichun Li is with the department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308, USA. lichun.li@ece.gatech.edu

J.S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, shamma@gatech.edu, and with King Abdullah University of Science and Technology (KAUST), jeff.shamma@kaust.edu.sa.

The authors acknowledge the financial support of the Army Research Office (W911NF-09-1-0553).

Let $\mathbb{R}^n$ indicate the n-dimensional real space, and $\mathscr{S}$ be a finite set. $|\mathscr{S}|$ denotes its cardinality, and $\Delta(\mathscr{S})$ indicates the set of probability distributions over $\mathscr{S}$. $(p^s)_{s\in\mathscr{S}}$ is a $|\mathscr{S}|$-dimensional row vector whose element is p^s. $\mathbf{1}$ and $\mathbf{0}$ are appropriately dimensional column vectors with all their elements to be 1 and 0, respectively.

A *two-player zero-sum stochastic game with the informed player as a single controller* is described by

- three finite non-empty sets: a set S of states of nature, a set I of actions of player 1 (maximizer), and a set J of actions of player 2 (minimizer). Let $s_t \in S$, $i_t \in I$ and $j_t \in J$ denote the state of nature, the action of player 1 and the action of player 2 at some stage $t \geq 1$.
- an initial distribution $p \in \Delta(S)$.
- a payoff function $g : S \times I \times J \to \mathbb{R}$. $G^s \in \mathbb{R}^{|I|\times|J|}$ denotes the payoff matrix of state s.
- a transition rule $q : S \times I \to \Delta(S)$. $Q^i \in \mathbb{R}^{|S|\times|S|}$ denotes the transition matrix given the action of player 1 to be i. $Q^i_{s,s'}$ is the conditional probability that the next state is s' given the current action is i and the current state is s.

The *play* of the N-stage zero-sum game is as follows:

- At stage 1, s_1 is chosen according to p, and told to player 1 only. Player 1 and 2 independently choose an action in their own set of actions, $i_1 \in I$ and $j_1 \in J$, respectively. The stage payoff for player 1 is $g(s_1, i_1, j_1)$, and (i_1, j_1) is publicly announced. The play proceeds to stage 2.
- At stage $t \geq 2$, s_t is chosen according to $q(s_{t-1}, i_{t-1})$, and told to player 1 only. Player 1 and 2 independently choose an action in their own set of actions. If $i_t \in I$ and $j_t \in J$ are selected, the payoff for player 1 at stage t is $g(s_t, i_t, j_t)$. (i_t, j_t) is publicly announced, and the play proceeds to the next stage.

Players are assumed to have perfect recall, and the whole description of the game is public knowledge.

A *behavior*[1] *strategy of player* 1 is an element $\sigma = (\sigma_t)_{t=1}^N$, where for each t, $\sigma_t : (S \times I \times J)^{t-1} \times S \to \Delta(I)$. Since player 2 does not observe the state, a *behavior strategy of player* 2 is an element $\tau = (\tau_t)_{t=1}^N$, where for each t, $\tau_t : (I \times J)^{t-1} \to \Delta(J)$. Denote by Σ_N and $\mathscr{T}_N$ the set of N-stage strategies of player 1 and 2, respectively.

Every distribution p, and a pair of strategy (σ, τ) induce a probability $P_{p,\sigma,\tau}$ over the set of plays $(S \times I \times J)^N$. We denote by $\mathbf{E}_{p,\sigma,\tau}$ the corresponding expectation operator. The *average payoff over N stages* is defined as

$$\gamma_N(p, \sigma, \tau) = \mathbf{E}_{p,\sigma,\tau}\left(\frac{1}{N}\sum_{t=1}^{N} g(s_t, i_t, j_t)\right) \tag{1}$$

The *N-stage game* $\Gamma_N(p)$ is defined as the zero-sum game with strategy spaces Σ_N and $\mathscr{T}_N$, and payoff function $\gamma_N(p, \sigma, \tau)$. We say the game $\Gamma_N(p)$ *has a value* $v_N(p)$ iff $\underline{v}_N(p) = \bar{v}_N(p) \doteq v_N(p)$, where $\underline{v}_N(p) = \sup_{\sigma\in\Sigma}\inf_{\tau\in\mathscr{T}} \gamma_N(p, \sigma, \tau)$ is the maxmin value of the game $\Gamma_N(p)$, and $\bar{v}_N(p) = \inf_{\tau\in\mathscr{T}}\sup_{\sigma\in\Sigma} \gamma_N(p, \sigma, \tau)$ is the minmax value of the game $\Gamma_N(p)$.

According to Kuhn's theorem [21], the game $\Gamma_N(p)$ can be seen as the mixed extension of a finite game, so $\Gamma_N(p)$ *has a value and both players have optimal strategies*.

The value $v_N(p)$ of the game $\Gamma_N(p)$ can be computed by a *recursive formula*. Let $p \in \Delta(S)$ represent player 2's belief on the state of nature at current stage. Player 1's action is chosen according to some $x = (x^s)_{s\in S} \in \Delta(I)^{|S|}$, i.e. that if the state is s, player 1 plays according to $x^s = (x^s(i))_{i\in I}$. Player 2's action is selected according to some $y = (y(j))_{j\in J} \in \Delta(J)$. The expected payoff of player 1 at current stage is

$$G(p, x, y) = \sum_{s\in S} p^s x^{sT} G^s y.$$

The probability that player 1 plays $i \in I$ is

$$\bar{x}_{p,x}(i) = \sum_{s\in S} p^s x^s(i). \tag{2}$$

Given that player 1 played $i \in I$ according to x at current stage, we denote by $p^+(p, x, i) \in \Delta(S)$ the conditional probability over the state at the next stage. We have

$$p^+(p, x, i) = \left(\frac{p^s x^s(i)}{\bar{x}_{p,x}(i)}\right)_{s\in S} Q^i \tag{3}$$

Let

$$T_{p,x}(v_{n-1}) = \sum_{i\in I} \bar{x}_{p,x}(i) v_{n-1}(p^+(p, x, i)).$$

Proposition 5.1 and Remark 5.2 of [4] provided the following lemma.

Lemma II.1. *For each $n \geq 1$ and $p \in \Delta(S)$,*

$$\begin{aligned} v_n(p) &= \max_{x\in\Delta(I)^{|S|}} \min_{y\in\Delta(J)} \left(\frac{1}{n}G(p, x, y) + \frac{n-1}{n} T_{p,x}(v_{n-1})\right) \\ &= \min_{y\in\Delta(J)} \max_{x\in\Delta(I)^{|S|}} \left(\frac{1}{n}G(p, x, y) + \frac{n-1}{n} T_{p,x}(v_{n-1})\right) \end{aligned}$$

Player 1 *has an optimal strategy which is Markovian, i.e. the action played at the any stage t only depends on the current state and player* 1*'s past actions* $i_1, \ldots, i_{t-1}$.

III. LP FORMULATION OF ASYMMETRIC STOCHASTIC GAMES

In this section, we begin our study about LP formulation of asymmetric stochastic games from one-stage and two-stage games, and then extend our results to N-stage games followed by a proof.

For the convenience of the rest of this section, let us introduce two lemmas whose proofs are given in an appendix.

Lemma III.1. *For any finite positive integer n and any constant $\alpha \geq 0$, the value $v_n(p)$ of an n-stage game $\Gamma_n(p)$ satisfies*

$$\alpha v_n(p) = v_n(\alpha p). \tag{4}$$

Lemma III.1 indicates that

$$\bar{x}_{p,x}(i) v_{n-1}(p^+(p, x, i)) = v_{n-1}\left((p^s x^s(i))_{s\in S} Q^i\right), \tag{5}$$

[1]Since the game is of perfect recall, there is no loss of generality in limiting ourselves to behavior strategies [21], [22].

and hence we have

$$T_{p,x}(v_{n-1}) = \sum_{i\in I} v_{n-1}((p^s x^s(i))_{s\in S} Q^i). \tag{6}$$

Lemma III.2. *Let $u(p,x) = \min_{y\in\Delta(J)} G(p,x,y)$. The value of $u(p,x)$ is the same value of the following linear program*

$$\begin{aligned} &\max_{\ell\in\mathbb{R}} \ \ell \\ &s.t. \sum_{s\in S} p^s G^{sT} x^s \geq \ell \mathbf{1}. \end{aligned} \tag{7}$$

A. One-Stage Games

Consider a one-stage stochastic game $\Gamma_1(p)$ whose value $v_1(p)$ satisfies

$$v_1(p) = \max_{x_1\in\Delta(I)^{|S|}} \min_{y_1\in\Delta(J)} \sum_{s\in S} p^s x_1^{sT} G^s y_1.$$

It is shown in [11] that $v_1(p)$ is the value of the following LP

$$\begin{aligned} &\max_{x,\ell} && \ell && \\ &s.t. && \sum_{s\in S} p^s G^{sT} x^s \geq \ell \mathbf{1} && \\ & && \mathbf{1}^T x^s = 1, && \forall s\in S \\ & && x^s \geq \mathbf{0}, && \forall s\in S \end{aligned} \tag{8}$$

For ease of the comparison with the LP of two-stage games, we let $z^s = p^s x^s$, and rewrite the above linear program as

$$\begin{aligned} &\max_{z,\ell} && \ell && \\ &s.t. && \sum_{s\in S} G^{sT} z^s \geq \ell \mathbf{1} && \\ & && \mathbf{1}^T z^s = p^s, && \forall s\in S \\ & && z^s \geq \mathbf{0}, && \forall s\in S \end{aligned} \tag{9}$$

B. Two-Stage Games

Now, let's consider a game with one more stage. A two-stage game $\Gamma_2(p)$ has a value $v_2(p)$, according to Lemma II.1, satisfying

$$\begin{aligned} v_2(p) &= \max_{x_1\in\Delta(I)^{|S|}} \min_{y_1\in\Delta(J)} \left(\frac{1}{2} G(p,x_1,y_1) + \frac{1}{2} T_{p,x_1}(v_1)\right) \\ &= \frac{1}{2} \max_{x_1\in\Delta(I)^{|S|}} \left(\sum_{i_2\in I} v_1((p^s x_1^s(i_1))_{s\in S} Q^{i_1}) \right. \\ &\quad \left. + \min_{y_1\in\Delta(J)} G(p,x_1,y_1) \right). \end{aligned} \tag{10}$$

The second equality holds because $T_{p,x_1}(v_1)$ is independent of y_1, and can be written as in equation (6). Now, let us analyze $v_2(p)$ term by term, and give a conclusion at the end of this subsection.

The first term $v_1((p^s x_2^s(i_2))_{s\in S} Q^{i_2})$, according to the analysis of one-stage games, has the value of the program (9). Let $h_t = \{i_1, i_2, \ldots, i_{t-1}\}$ be the history action of player 1 at stage t, and H_t be a set including all possible history actions of player 1 at stage t. We notice that the history action $h_2 = \{i_1\}$ of player 1 influences the value of v_1 and the behavior strategy at stage 2. To emphasize the dependence on player 1's history action, we add h_2 as an extra subscript to the variables x and ℓ. Therefore, for each $h_2 \in H_2$, the value of $v_1((p^s x_1^s(i_2))_{s\in S} Q^{i_2})$ can be computed by solving the following program

$$\begin{aligned} &\max_{z_{2|h_2},\ell_{2|h_2}} && \ell_{2|h_2} && \\ &s.t. && \sum_{s\in S} G^{sT} z^s_{2|h_2} \geq \ell_{2|h_2} \mathbf{1} && \\ & && \mathbf{1}^T z^s_{2|h_2} = \left((p^s x_1^s(i_1))_{s\in S} Q^{i_1}\right)^s, && \forall s\in S \\ & && z^s_{2|h_2} \geq \mathbf{0}, && \forall s\in S \end{aligned} \tag{11}$$

Meanwhile, the second term $\min_{y_1\in\Delta(J)} G(p,x_1,y_1)$, according to Lemma III.2, has the value of another linear program

$$\begin{aligned} &\max_{\ell_1\in\mathbb{R}} \ \ell_1 \\ &s.t. \sum_{s\in S} p^s G^{sT} x_1^s \geq \ell_1 \mathbf{1}. \end{aligned} \tag{12}$$

To conclude, the value $v_2(p)$ of a two-stage game is the same as the value of the following linear program according to equation (10), (11), and (12)

$$\begin{aligned} &\frac{1}{2} \max_{x_1} \left(\sum_{h_2\in H_2} \max_{z_{2|h_2},\ell_{2|h_2}} \ell_{2|h_2} + \max_{\ell_1} \ell_1 \right) \\ &s.t. \quad \sum_{s\in S} p^s G^{sT} x_1^s \geq \ell_1 \mathbf{1}. \\ & \qquad \mathbf{1}^T x_1^s = 1, && \forall s\in S \\ & \qquad x_1^s \geq \mathbf{0}, && \forall s\in S \\ & \qquad \sum_{s\in S} G^{sT} z^s_{2|h_2} \geq \ell_{2|h_2} \mathbf{1} && \forall h_2\in H_2 \\ & \qquad \mathbf{1}^T z^s_{2|h_2} = \left((p^s x_1^s(i_1))_{s\in S} Q^{i_1}\right)^s, && \forall s\in S, h_2\in H_2 \\ & \qquad z^s_{2|h_2} \geq \mathbf{0}, && \forall s\in S, h_2\in H_2 \end{aligned}$$

Let $z_1^s = p^s x_1^s$. We rewrite the LP above as

$$\begin{aligned} &\frac{1}{2} \max_{z_1,\ell_1,z_{2|h_2},\ell_{2|h_2},\forall h_2\in H_2} \left(\ell_1 + \sum_{h_2\in H_2} \ell_{2|h_2} \right) \\ &s.t. \quad \sum_{s\in S} G^{sT} z_1^s \geq \ell_1 \mathbf{1}. \\ & \qquad \mathbf{1}^T z_1^s = p^s, && \forall s\in S \\ & \qquad z_1^s \geq \mathbf{0}, && \forall s\in S \\ & \qquad \sum_{s\in S} G^{sT} z^s_{2|h_2} \geq \ell_{2|h_2} \mathbf{1} && \forall h_2\in H_2 \\ & \qquad \mathbf{1}^T z^s_{2|h_2} = \left((p^s x_1^s(i_1))_{s\in S} Q^{i_1}\right)^s, && \forall s\in S, h_2\in H_2 \\ & \qquad z^s_{2|h_2} \geq \mathbf{0}, && \forall s\in S, h_2\in H_2 \end{aligned}$$

To make the form more compact, for each stage t and each possible history action h_t of player 1, we define two variables $z_{t|h_t}$ and $\ell_{t|h_t}$. Hence, z_1 and ℓ_1 in the above LP are

replaced by $z_{1|h_1}$ and $\ell_{1|h_1}$, and we have the following form.

$$\frac{1}{2}\max_{\substack{z_{t|h_t},\ell_{t|h_t},\\ \forall t=1,2,\forall h_t\in H_t}} \sum_{t=1}^{2}\sum_{h_t\in H_t}\ell_{t|h_t} \tag{13}$$

$$\begin{aligned} s.t. \quad & \forall t=1,2,\forall h_t\in H_t \\ & \sum_{s\in S} G^{sT} z^s_{t|h_t} \geq \ell_{t|h_t}\mathbf{1} \\ & \mathbf{1}^T z^s_{t|h_t} = \left((z^s_{t-1}(i_{t-1}))_{s\in S}Q^{i_{t-1}}\right)^s, \quad \forall s\in S \\ & z^s_{t|h_t} \geq \mathbf{0}, \quad \forall s\in S, \end{aligned}$$

where $z^s_0(i_0)=p^s$ and Q^{i_0} is an identity matrix.

C. N-Stage Games

Compared the LP formulation (9) of one-stage games and the LP formulation (13) of two-stage games, we find that they are in *a uniform form* which is for each stage t and each possible history action h_t, the corresponding strategy variable $z_{t|h_t}$ and the corresponding value variable $\ell_{t|h_t}$ satisfy the same constraints.

This uniform LP formulation is also true for N-stage games, and we have the following main theorem. The proof is based on mathematical induction, and is given in the appendix.

Theorem III.3. *For each finite positive integer N, the value $v_N(p)$ of N-stage asymmetric stochastic game $\Gamma_N(p)$ has the value of the linear program*

$$\frac{1}{N}\max_{\substack{z_{t|h_t},\ell_{t|h_t},\\ \forall t=1,\ldots,N,\forall h_t\in H_t}} \sum_{t=1}^{N}\sum_{h_t\in H_t}\ell_{t|h_t} \tag{14}$$

$$\begin{aligned} s.t. \quad & \forall t=1,\ldots,N \ \forall h_t\in H_t \\ & \sum_{s\in S} G^{sT} z^s_{t|h_t} \geq \ell_{t|h_t}\mathbf{1} \\ & \mathbf{1}^T z^s_{t|h_t} = \left((z^s_{t-1}(i_{t-1}))_{s\in S}Q^{i_{t-1}}\right)^s, \quad \forall s\in S \\ & z^s_{t|h_t} \geq \mathbf{0}, \quad \forall s\in S, \end{aligned}$$

where $z^s_0(i_0)=p^s$ and Q^{i_0} is an identity matrix. The optimal behavior strategy ${x^s}^*_{t|h_t}$ at stage t given the history action to be h_t and the current state to be s is

$${x^s_{t|h_t}}^* = \begin{cases} \frac{{z^s}^*_{t|h_t}}{\mathbf{1}^T {z^s}^*_{t|h_t}}, & \text{if } {z^s}^*_{t|h_t}\neq\mathbf{0}; \\ 0, & \text{otherwise} \end{cases}$$

where ${z^s}^*_{t|h_t}$ is the optimal solution of the LP formulation (14).

Remark III.4. *${x^s_{t|h_t}}^*$ equals to* 0 *means that it is impossible for the current state to be s given the initial distribution p and the history action of player* 1 *to be h_t, or it is impossible for player* 1*'s history action to be h_t.*

The size of this linear program is *linear* with respect to the size $|S|$ of the state set and the size $|J|$ of player 2's action set, *polynomial* with respect to the size $|I|$ of player 1's action set, and *exponential* with respect to the length N of the horizon. It is easy to see that for an N-stage game, there are at most $|I|^{t-1}$ different history actions of player 1 at stage t. For each history action at each stage, we need an independent pair of variables $z\in\mathbb{R}^{|I|\times|S|}$ and $\ell\in\mathbb{R}$, and hence in total there are $(1+|I|+\cdots+|I|^{N-1})(1+|I||S|)=\mathbf{O}(|\mathbf{S}||\mathbf{I}^{\mathbf{N+1}}|)$ scalar variables. Meanwhile, for each history action at each stage, there are three sets of constraints. Constraint set 1 has $|J|$ inequalities, constraint set 2 has $|S|$ equalities, and constraint set 3 has $|I||S|$ inequalities. In total, there are $(1+|I|+\cdots+|I|^{N-1})(|J|+|S|+|I||S|)=\mathbf{O}(|\mathbf{J}||\mathbf{I}|^{\mathbf{N}}+|\mathbf{S}||\mathbf{I}|^{\mathbf{N+1}})$ constraints. Therefore, we say the size of the LP formulation (14) grows only linearly with respect to $|J|$ and $|S|$, polynomially with respect to $|I|$, and exponentially with respect to N.

The LP formulation given in Theorem III.3 has the advantage of cutting the computational complexity from a polynomial function to a linear function with respect to $|S|$ and $|J|$ over the LP formulation of the extensive form of game $\Gamma_N(p)$. If we model the stochastic game $\Gamma_N(p)$ in an extensive form and express this extensive form as a tree, then there will be $|S|^N|I|^N|J|^N$ leaves, and the corresponding LP formulation will have the linear size in the size of the game tree [16], [17]. Compared with the LP formulation of the extensive form of the game $\Gamma_N(p)$, the LP formulation (14) greatly decreases the computational complexity.

IV. EXAMPLE: A TRAVELLING INSPECTOR PROBLEM

This section will apply the LP formulation of asymmetric stochastic games proposed in Theorem III.3 to the travelling inspector problem which is developed from [20], [23], and discuss the computational complexity of the LP formulation proposed in Theorem III.3, the optimal policy and the value functions of this problem.

An inspector wants to prevent illegal dumping of toxic wastes in the region to which he is assigned, and needs the optimal travelling strategy. In his region, there are two big cities: A and B. In city *A*, there is plant 2 with toxic waste capacity to be 2 tons. In city *B*, there is plant 3 with toxic waste capacity to be 3 tons. Between city *A* and *B*, there is plant 1 which is relatively small, and has toxic waste capacity 1 ton. The three plants coordinate their dumping strategy to dump as much toxic wastes as possible without being fined excessively.

The actions for each plant on each day are to dump (1) or not to dump (0), and there are 8 coordinated actions for the three plants. On each day, the inspector can only visit one city. In city *A*, he can inspect either plant 1 or plant 2. In city *B*, he can inspect either plant 1 or plant 3. If the plant he visits is dumping toxic wastes, then the plant will be fined 2 thousand dollars of its toxic waste capacity. Meanwhile, the cost to treat the uninspected dumped wastes is 1 thousand dollars per ton. We can model the payoff matrices for the inspector in city *A* and *B* as in Table I.

After inspecting one plant, the inspector may move and visit the other city on the next day. Because plant 2 is far away from city *B*, the inspector prefers to stay in city *A* after inspecting plant 2. For the same reason, the inspector prefers to stay in city *B* after inspecting plant 3. If the inspector visited plant 1, then he can move to either city *A* or city *B*

TABLE I

PAYOFF MATRICES

	000	001	010	011	100	101	110	111
1	0	-3	-2	-5	2	-1	0	-3
2	0	-3	4	1	-1	-4	3	0
3	-100	-100	-100	-100	-100	-100	-100	-100

G^A

	000	001	010	011	100	101	110	111
1	0	-3	-2	-5	2	-1	0	-3
2	-100	-100	-100	-100	-100	-100	-100	-100
3	0	6	-2	4	-1	5	-4	2

G^B

TABLE II

TRANSITION MATRICES

	A	B
A	0.5	0.5
B	0.5	0.5

Q^1

	A	B
A	0.8	0.2
B	0.8	0.2

Q^2

	A	B
A	0.2	0.8
B	0.2	0.8

Q^3

with equal possibility. Therefore, the transition matrices are modeled as in Table II. The objective of the inspector is to maximize his average payoff during a certain period, while the three plants want to minimize his payoff.

If we fix the horizon to be 6 days, there are $(1+3+\cdots+3^6)(8+2+6) = 17488$ equations and $(1+3+\cdots+3^6)(1+6) = 7651$ variables in the linear program formulated in Theorem III.3, which takes a computer with 2.4 GHz CUP about 8 seconds to compute the optimal solution and the game value. If we formulate a linear program for the extensive form of the same game, the number of equations is in the order of $8^6 = 26214$, and the number of variables is in the order of $2^6 3^6 = 46656$. Compared with the LP of extensive game, the LP proposed in Theorem III.3 cuts about $1/3$ of the equations and about $5/6$ of the variables.

Given the initial distribution $p = [0.5\ 0.5]$ and the horizon to be 6, the inspector's optimal strategy at stage 3 is given in Table III. We find that the optimal strategy only depends on the last action of the inspector. This is because the three plants' belief on which city the inspector visits today only depends on which plant the inspector visited yesterday. For example, if plant 2 was visited yesterday, then the inspector must be in city A yesterday no matter which plants the inspector inspected before yesterday, and the plants' belief about where the inspector is today is always [0.8 0.2] according to the transition matrix Q^2.

The value function of the travelling inspector game is given in Figure 1. From the plot, we find that the value functions are piece-wise linear. Reference [11] pointed out that the value functions of asymmetric repeated games with finite horizon were piece-wise linear. We believe that this is still true for the asymmetric stochastic games.

TABLE III

THE OPTIMAL STRATEGY AT STAGE 3

h_3	s_3	$x^*_{3\|h_3}$		
11, 21, 31	A	0.22	0.78	0
	B	0.33	0	0.67
12, 22, 32	A	0.58	0.42	0
	B	0	0	1
13, 23, 33	A	0	1	0
	B	0.58	0	0.42

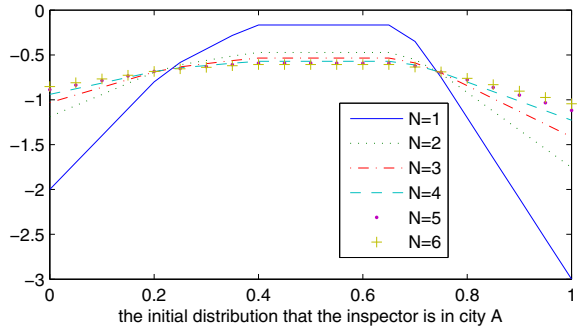

Fig. 1. Value function for $N = 1,2\ldots,6$

V. CONCLUSION

This paper proposed an LP formulation of asymmetric stochastic games. The size of the proposed linear program is only linear with respect to the size of the state and the size of the uninformed player's actions. However, the size of the proposed LP formulation still grows exponentially with respect to the horizon. A tight bound was achieved by implementing a suboptimal policy in a receding horizon manner in [24]. Our next step is to study when the optimal strategy for a finite-horizon game is implemented in a receding horizon manner, whether we can asymptotically achieve the optimal strategy of the game with infinite horizon.

APPENDIX

Proof of Lemma III.1 It is shown by mathematical induction. First, we notice that $G(\alpha p, x_n, y_n) = \alpha G(p, x_n, y_n)$. For $n = 1$, we have

$$\begin{aligned} v_1(\alpha p) &= \max_{x_1 \in \Delta(I)^{|S|}} \min_{y_1 \in \Delta(J)} G(\alpha p, x_1, y_1) \\ &= \alpha \max_{x_1 \in \Delta(I)^{|S|}} \min_{y_1 \in \Delta(J)} G(p, x_1, y_1) = \alpha v_1(p). \end{aligned}$$

Now, let us assume equation (4) is true for $n-1$. From equation (2), we have $\bar{x}_{\alpha p, x_n}(i) = \alpha \bar{x}_{p,x_n}(i)$. From equation (3), we have $p^+(\alpha p, x_n, i) = p^+(p, x_n, i)$. Therefore, $T_{\alpha p, x_n}(v_{n-1}) = \alpha T_{p,x_n}(v_{n-1})$, and $v_n(\alpha p)$ equals to

$$\begin{aligned} &= \max_{x_n \in \Delta(I)^{|S|}} \min_{y_n \in \Delta(J)} \left(\frac{1}{n} G(\alpha p, x_n, y_n) + \frac{n-1}{n} T_{\alpha p, x_n}(v_{n-1}) \right) \\ &= \max_{x_n \in \Delta(I)^{|S|}} \min_{y_n \in \Delta(J)} \left(\alpha \frac{1}{n} G(p, x_n, y_n) + \alpha \frac{n-1}{n} T_{p, x_n}(v_{n-1}) \right) \\ &= \alpha v_n(p) \end{aligned}$$

Proof of Lemma III.2 Notice that $\sum_{s \in S} p^s G^{sT} x^s_n \doteq c$ is a $|J|$ dimensional vector. So, $u(p, x_n) = \min_{y_n \in \Delta(J)} c^T y_n$. Since y_n is a simplex in $\Delta(J)$, $u(p, x_n)$ actually is the smallest element in c, and can be solved by the linear program (V).

Proof of Theorem III.3 Mathematical induction is used to show Theorem III.3.

From equation (9), we see that Theorem III.3 is true for $N = 1$. Assume that Theorem III.3 is true for $N-1$-stage stochastic games. For N-stage games, Lemma II.1 indicates

$$v_N(p) = \max_{x_1 \in \Delta(I)^{|S|}} \min_{y_1 \in \Delta(J)} \left(\frac{1}{N} G(p, x_1, y_1) + \frac{N-1}{N} T_{p, x_1}(v_{N-1}) \right)$$

$$= \max_{x_1 \in \Delta(I)^{|S|}} \left(\frac{N-1}{N} \sum_{i_1 \in I} v_{N-1}((p^s x_1^s(i_1))_{s\in S} Q^{i_1}) + \min_{y_1 \in \Delta(J)} \frac{1}{N} G(p, x_1, y_1) \right).$$

The 2nd equality holds because $v_{N-1}((p^s x_1^s(i_N))_{s\in S} Q^{i_1})$ is independent of y_1. Let us analyze this equation term by term.

Since Theorem III.3 is true for $N-1$-stage stochastic games, the first term $\frac{N-1}{N} v_{N-1}((p^s x_1^s(i_1))_{s\in S} Q^{i_1})$ can be solved through the linear program defined in (14) with the the initial decision variable $z_0^s(i_0) = \left((p^s x_1^s(i_1))_{s\in S} Q^{i_1}\right)^s$. Since the $N-1$-stage game is a sub-game (from stage 2 to N) of the N-stage game, the tth stage in the $N-1$-stage game is actually the $t+1$th in the N-stage game, the history action h_t of player 1 in the $N-1$-stage game is actually the history action from stage 2 to stage N in the N-stage game, and for player 1's each action i_1 at stage 1 in the N-stage game, there are independent set of variables $z_{t|h_t}$ and $\ell_{t|h_t}$ in the $N-1$-stage game. Let $t = n-1$. The first term has the same value of the following LP.

$$\frac{1}{N} \max_{\substack{z_{n|h_n}, \ell_{n|h_n}, \\ \forall n=2,\ldots,N, \forall h_n \in H_n}} \sum_{n=2}^{N} \sum_{h_n \in H_n} \ell_{n|h_n} \tag{15}$$

$$\begin{aligned} s.t. \quad & \forall n = 2, \ldots, N \ \forall h_n \in H_n \\ & \sum_{s\in S} G^{sT} z^s_{n|h_n} \geq \ell_{n|h_n} \mathbf{1} \\ & \mathbf{1}^T z^s_{n|h_n} = \left((z^s_{n-1}(i_{n-1}))_{s\in S} Q^{i_{n-1}}\right)^s, \quad \forall s \in S \\ & z^s_{t|h_t} \geq \mathbf{0}, \quad \forall s \in S, \end{aligned}$$

where $z_1^s(i_1) = \left((p^s x_1^s(i_1))_{s\in S} Q^{i_1}\right)^s$.

The second term $\min_{y_1 \in \Delta(J)} G(p, x_1, y_1)$, according to Lemma III.2 and $z^s_{1|h_1} = p^s x_1^s$, has the value of

$$\begin{aligned} & \max_{\ell_n \in \mathbb{R}} \ \ell_{1|h_1} \\ & s.t. \sum_{s\in S} G^{sT} z^s_{1|h_1} \geq \ell_{1|h_1} \mathbf{1}. \end{aligned}$$

Therefore, for the N-stage game $\Gamma_N(p)$, the game value $v_N(p)$ is the value of

$$\frac{1}{N} \max_{\substack{z_{t|h_t}, \ell_{t|h_t}, \\ \forall t=1,\ldots,N, \forall h_t \in H_t}} \sum_{t=1}^{N} \sum_{h_t \in H_t} \ell_{t|h_t}$$

$$\begin{aligned} s.t. \quad & \forall t = 1, \ldots, N \ \forall h_t \in H_t \\ & \sum_{s\in S} G^{sT} z^s_{t|h_t} \geq \ell_{t|h_t} \mathbf{1} \\ & \mathbf{1}^T z^s_{t|h_t} = \left((z^s_{t-1}(i_{t-1}))_{s\in S} Q^{i_{t-1}}\right)^s, \quad \forall s \in S \\ & z^s_{t|h_t} \geq \mathbf{0}, \quad \forall s \in S, \end{aligned}$$

where $z_0^s(i_0) = p^s$ and Q^{i_0} is an identity matrix. The behavior strategy $x_1^s = \frac{z^s_{1|h_1}}{\mathbf{1}^T z^s_{1|h_1}}$ if $z^s_{1|h_1} \neq \mathbf{0}$, 0 otherwise. The behavior strategy $x^s_{t|h_t}$ is $\frac{z^s_{t|h_t}}{\mathbf{1}^T z^s_{t|h_t}}$ if $z^s_{t|h_t} \neq \mathbf{0}$, 0 otherwise.

Therefore, Theorem III.3 still holds for N-stage games, and hence complete the proof.

REFERENCES

[1] R. J. Aumann and M. Maschler, *Repeated games with incomplete information*. MIT press, 1995.

[2] S. Sorin, *A first course on zero-sum repeated games*. Springer, 2002, vol. 37.

[3] S. Zamir, "Repeated games of incomplete information: Zero-sum," *Handbook of Game Theory*, vol. 1, pp. 109–154, 1992.

[4] J. Renault, "The value of Markov chain games with lack of information on one side," *Mathematics of Operations Research*, vol. 31, no. 3, pp. 490–512, 2006.

[5] D. Rosenberg, E. Solan, and N. Vieille, "Stochastic games with a single controller and incomplete information," *SIAM journal on control and optimization*, vol. 43, no. 1, pp. 86–110, 2004.

[6] J. Renault, "The value of repeated games with an informed controller," *Mathematics of Operations Research*, vol. 37, no. 1, pp. 154–179, 2012.

[7] J. P. Hespanha and M. Prandini, "Nash equilibria in partial-information games on Markov chains," in *Decision and Control, 2001. Proceedings of the 40th IEEE Conference on*, vol. 3. IEEE, 2001, pp. 2102–2107.

[8] E. Altman, V. Kamble, and A. Silva, "Stochastic games with one step delay sharing information pattern with application to power control," in *Game Theory for Networks, 2009. GameNets' 09. International Conference on*. IEEE, 2009, pp. 124–129.

[9] A. Nayyar, A. Gupta, C. Langbort, and T. Basar, "Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games," 2014.

[10] M. L. Littman and P. Stone, "A polynomial-time nash equilibrium algorithm for repeated games," *Decision Support Systems*, vol. 39, no. 1, pp. 55–66, 2005.

[11] J.-P. Ponssard and S. Sorin, "The LP formulation of finite zero-sum games with incomplete information," *International Journal of Game Theory*, vol. 9, no. 2, pp. 99–105, 1980.

[12] A. Gilpin and T. Sandholm, "Solving two-person zero-sum repeated games of incomplete information," in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 903–910.

[13] T. Sandholm, "The state of solving large incomplete-information games, and application to poker," *AI Magazine*, vol. 31, no. 4, pp. 13–32, 2010.

[14] D. M. Kreps and R. Wilson, "Sequential equilibria," *Econometrica: Journal of the Econometric Society*, pp. 863–894, 1982.

[15] D. Koller and N. Megiddo, "The complexity of two-person zero-sum games in extensive form," *Games and economic behavior*, vol. 4, no. 4, pp. 528–552, 1992.

[16] D. Koller, N. Megiddo, and B. Von Stengel, "Efficient computation of equilibria for extensive two-person games," *Games and Economic Behavior*, vol. 14, no. 2, pp. 247–259, 1996.

[17] B. Von Stengel, "Efficient computation of behavior strategies," *Games and Economic Behavior*, vol. 14, no. 2, pp. 220–246, 1996.

[18] J. Zheng and D. A. Castanon, "Dynamic network interdiction games with imperfect information and deception," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 7758–7763.

[19] ——, "Stochastic dynamic network interdiction games," in *American Control Conference (ACC), 2012*. IEEE, 2012, pp. 1838–1844.

[20] J. Filar and K. Vrieze, *Competitive Markov decision processes*. Springer-Verlag New York, Inc., 1996.

[21] H. W. Kuhn, "Extensive games and the problem of information," *Contributions to the Theory of Games*, vol. 2, no. 28, pp. 193–216, 1953.

[22] R. J. Aumann, "Mixed and behavior strategies in infinite extensive games," DTIC Document, Tech. Rep., 1961.

[23] J. A. Filar, "Player aggregation in the traveling inspector model," *Automatic Control, IEEE Transactions on*, vol. 30, no. 8, pp. 723–729, 1985.

[24] M. Jones and J. S. Shamma, "Policy improvement for repeated zero-sum games with asymmetric information," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 7752–7757.