2024 Tenth Indian Control Conference (ICC)
9 - 11 December, 2024
IISER Bhopal, India

Mathematical Programming Formulations for Computing Nash Equilibrium of Stochastic Games

Ayush Verma[1], Vikas Vikram Singh[2], Prashant Palkar[3]

***Abstract*— In this paper, we propose a trilinear optimization formulation and four mixed integer bilinear formulations to compute the Nash equilibria of a 2-player discounted stochastic game. The trilinear optimization formulation is derived from an existing nonlinear programming problem whose global minimum gives a Nash equilibrium of the discounted stochastic game. The mixed integer bilinear formulations use the optimality condition of a best response Markov decision problem of a player for a fixed strategy of the other player. We compare the performance of our formulations with the existing nonlinear programming formulation. Numerical experiments show that our trilinear formulation outperforms the nonlinear programming formulation as well as the mixed integer bilinear formulations.**

I. Introduction

Strategic interactions among rational agents can be modeled as a non-cooperative game where a Nash equilibrium signifies a stable solution concept against unilateral deviations. In 1950, Nash [13] showed that there exists a mixed strategy Nash equilibrium of a finite strategic game. Since then, the Nash equilibrium of non-cooperative games with general payoffs and continuous strategy sets has been extensively studied in the literature. Under certain conditions on the payoff functions and strategy sets, there exists a Nash equilibrium of the game [1], [2]. The computation of Nash equilibria of these games is closely related to the techniques from optimization theory. For example, a saddle point equilibrium of a zero-sum game can be computed by solving a primal-dual pair of linear programs. A Nash equilibrium of a two-player bimatrix game can be computed using the Lemke-Howson algorithm [10]. Magasarian and Stone [11], around the same time as Lemke and Howson [10], proposed an equivalent quadratic program whose global maximum gives a Nash equilibrium of the game. Sandholm et al. [19] proposed equivalent mixed integer linear programming based methods to compute the Nash equilibria of a 2-player bimatrix game. These methods outperform the Lemke-Howson algorithm on certain data sets. Fischer and Gupte [6] generalized the mixed integer formulations given in [19] as well as the results of Magasarian and Stone [11] for the n-player case by proposing an equivalent multilinear programming problem whose global maximum gives a Nash equilibrium of the game. Using the fact that the objective function value of the multilinear programming problem is nonpositive, they proposed a multilinear feasibility problem that outperforms all other mixed integer formulations and multilinear programming problems.

The aforementioned research focuses only on static games. Dynamic games where players interact over a period of time and the game moves from one state to another according to a Markov chain controlled by all the players were first introduced by L.S. Shapley in his seminal paper [20]. These games are called stochastic games and are extensively studied in the literature. When all the players use the expected discounted payoff criterion, there exists a stationary Nash equilibrium [5], [21]. Similar to the case of static games, different approaches from optimization theory are used to compute a Nash equilibrium of the stochastic game. For example, the zero-sum single controller stochastic game can be solved using a linear program [14] whereas the Nash equilibria of a nonzero-sum game of the same class and switching controller games can be computed by solving linear complementarity problems [8], [17] which are Lemke's processible [9]. For algorithmic aspects of stochastic games, including special classes, we refer the readers to the survey [16]. For a general stochastic game, a Nash equilibrium can be computed by finding a global minimum of a certain nonlinear programming problem [3], [4]. This result can be viewed as a generalization of the result for bimatrix game [11] to the case of stochastic game.

To the best of our knowledge, mixed integer formulations to compute Nash equilibria of stochastic games have not been considered before in the literature. In this paper, we consider a two-player discounted stochastic game. We first propose a set of nonlinear inequality constraints whose feasible solutions are Nash equilibria of the game. Then, we propose four different mixed integer programs where the first one is posed as a feasibility problem, and the rest are with an objective function. We show that a feasible solution of the first mixed integer program and a global minimum of other formulations correspond to a Nash equilibrium of the original game. We perform numerical experiments using two state-of-art global optimization solvers, the Octeract Engine [22], and Baron [18] (both of which can also handle integer variables), on randomly generated instances of various sizes using all our formulations and the formulation proposed in [4]. We observe that as the number of states and actions increases, the solvers are unable to solve the optimization problem proposed in [4]. We also compare the performance of all the formulations and observe that the nonlinear feasibility

[1] Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, 110016, New Delhi, India ayushv148@gmail.com
[2] Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, 110016, New Delhi, India vikassingh@maths.iitd.ac.in
[3] Department of Mechanical Engineering, Indian Institute of Technology Delhi, Hauz Khas, 110016, New Delhi, India ppalkar@mech.iitd.ac.in

979-8-3315-1721-2/24/$31.00 ©2024 IEEE

problem formulation outperforms all other formulations.

The structure of the paper is as follows. We present a stochastic game model in Section II. Section III comprises of the new reformulations of the stochastic game. The description of numerical experiments, conclusions and some future directions of research are elaborated in Section IV.

II. Model

A stochastic game is used to model repeated interaction among several players in which the underlying state of the environment changes probabilistically, and it depends on the actions of the players. In this paper, we focus on a 2-player stochastic game that has a finite number of states, and each player has a finite number of actions available at each state. Let S denote the finite set of states and $A^1(s)$, $A^2(s)$ denote the finite action sets of player 1 and player 2 at state $s \in S$, respectively. At any time t if the state is s and player 1 chooses action a^1 and player 2 chooses action a^2, they receive rewards $r^1(s, a^1, a^2)$ and $r^2(s, a^1, a^2)$, respectively. The game moves to a state s' with probability $p(s'|s, a^1, a^2)$ at time $t + 1$ where again the same thing gets repeated, and the process continues infinitely. The actions taken by both the players at time t may depend on a sequence of states and actions up to time $t - 1$ and state at time t, which defines a history $h_t = (s_0, a_0, s_1, \ldots, s_{t-1}, a_{t-1}, s_t)$. The choice of such action is given by a history-dependent decision rule, and the sequence of such decision rules forms a history-dependent strategy. We refer to Chapter 1 of [15] for different types of decision rules and strategies. It is well known that for a finite state-action stochastic game under discounted cost criterion, there exists a Nash equilibrium among stationary strategies [5], [21]. A stationary strategy f of player 1 is defined by a vector $f = \big(f(s)\big)_{s\in S}$, where $f(s) \in \wp(A(s))$ for each $s \in S$; $\wp(A(s))$ denotes the set of probability distributions on finite set $A(s)$. As per stationary strategy f, whenever the state of the game is s, player 1 chooses action a^1 with probability $f(s, a^1)$. A stationary strategy g of player 2 is defined similarly. We denote the set of stationary strategies of player 1 and player 2 by F_S and G_S, respectively. These sets are defined as

$$F_S = \big\{f \mid f(s, a^1) \geq 0, \sum_{a^1\in A^1(s)} f(s, a^1) = 1 \ \forall\ s \in S, \\ a^1 \in A^1(s)\big\},$$

$$G_S = \big\{g \mid g(s, a^2) \geq 0, \sum_{a^2\in A^2(s)} g(s, a^2) = 1, \ \forall\ s \in S, \\ a^2 \in A^2(s)\big\}.$$

For an initial state s, and a given stationary strategy pair $(f, g) \in F_S \times G_S$ and a discount factor $\beta \in [0, 1)$, the expected discounted value of player i, $i = 1, 2$, is defined by

$$v^i_\beta(s, f, g) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}^{f,g}_s(r^i(X_t, A^1_t, A^2_t)), \qquad (1)$$

where X_t, A^1_t, and A^2_t denote the state at time t and the actions taken by players 1 and 2 at time t, respectively. Let $P(f, g)$ be a transition probability matrix induced by strategy pair (f, g) and its components are defined by $p(s'|s, f, g) = \sum_{a^1\in A^1(s)} \sum_{a^2\in A^2(s)} f(s, a^1)p(s'|s, a^1, a^2)g(s, a^2)$, for all $s, s' \in S$. Define the expected reward vector $r^i(f, g) = (r^i(s, f, g))_{s\in S}$, $i = 1, 2$, where for all $s \in S$, $r^i(s, f, g) = \sum_{a^1\in A^1(s)} \sum_{a^2\in A^2(s)} f(s, a^1)r^i(s, a^1, a^2)g(s, a^2)$. It follows from [3], the value vector of player i at strategy profile (f, g) defined by (1) can be written as

$$v^i_\beta(f, g) = \big(I - \beta P(f, g)\big)^{-1} r^i(f, g), \ i = 1, 2,$$

where I denotes an $|S| \times |S|$ identity matrix. A strategy pair $(f^*, g^*) \in F_S \times G_S$ is said to be a Nash equilibrium of the discounted stochastic game if for each $s \in S$

$$v^1_\beta(s, f^*, g^*) \geq v^1_\beta(s, f, g^*) \ \forall\ f \in F_S,$$
$$v^2_\beta(s, f^*, g^*) \geq v^2_\beta(s, f^*, g) \ \forall\ g \in G_S.$$

We denote the above discounted stochastic game by G_β. Filar et al. [4] characterized the Nash equilibrium of the stochastic game using a global minimum of a nonlinear programming problem. We use the following matrix notations in order to present the nonlinear programming problem given in [3], [4]. For each $i = 1, 2$, let $v^i = (v^i(s))_{s\in S}$, and for all $s \in S$, define

$$\mathbf{R}^i(s) = \big(r^i(s, a^1, a^2)\big)_{a^1\in A^1(s), a^2\in A^2(s)}, \qquad (2)$$

and

$$\mathbf{T}(s, v^i) = \left(\sum_{s'\in S} p(s'|s, a^1, a^2)v^i(s')\right)_{a^1\in A^1(s), a^2\in A^2(s)}. \qquad (3)$$

Theorem 1. *Consider a vector $z = (v^1, v^2, f, g)$. Then the strategy pair (f, g) of z forms a Nash equilibrium of the game G_β if and only if z is a global minimum, with objective function value zero, of the following nonlinear programming problem* (T1).

$$\begin{aligned}
&\min\ \mathbf{1}^\top_{|S|}[v^1 + v^2 - r^1(f, g) - r^2(f, g) \\
&\qquad - \beta P(f, g)(v^1 + v^2)] \\
&(i)\ \mathbf{R}^1(s)g(s) + \beta\mathbf{T}(s, v^1)g(s) \leq v^1(s)\mathbf{1}_{|A^1(s)|}, \ \forall\, s \in S, \\
&(ii)\ f^\top(s)\mathbf{R}^2(s) + \beta f^\top(s)\mathbf{T}(s, v^2) \leq v^2(s)\mathbf{1}^\top_{|A^2(s)|}, \\
&\qquad \forall\, s \in S, \\
&(iii)\ (f, g) \in \mathbf{F}_S \times \mathbf{G}_S, \qquad \text{(T1)}
\end{aligned}$$

where for a given set C, $\mathbf{1}_{|C|}$ is an $|C| \times 1$ vector of ones.

Proof. See Theorem 3.8.2 of [3] □

III. New optimization reformulations of stochastic games

In this section, we propose new equivalent optimization reformulations of a discounted stochastic game that can be used to compute its Nash equilibria.

A. Trilinear feasibility problem

The objective function value is nonnegative for every feasible solution of (T1). Then, it follows from Theorem 1 that computing a Nash equilibrium is equivalent to the following feasibility problem

$$\mathbf{1}_{|S|}^{\top}(v^1+v^2-r^1(f,g)-r^2(f,g)-\beta P(f,g)(v^1+v^2))\leq 0,$$

$$(i)-(iii) \text{ of (T1).} \tag{T2}$$

Corollary 1. *A strategy pair (f,g) is a Nash equilibrium of the game G_β if and only if there exist vectors v^1 and v^2 such that $z=(v^1,v^2,f,g)$ is a feasible point of* (T2)*.*

Proof. The proof directly follows from Theorem 1 because every feasible solution of (T2) is a global minimum of the optimization problem (T1). □

B. Mixed integer bilinear programs

Before we delve into the mixed integer formulations, we present sufficient conditions of the best response strategy of a player for a fixed strategy of the other player. Later, we use these sufficient conditions to derive equivalent mixed integer bilinear programs. It is well known that for a fixed strategy g of player 2, the problem of finding an optimal strategy of player 1 is a Markov Decision Process (MDP) with rewards and transition probabilities given by

$$r^1(s,a^1,g)=\sum_{a^2\in A^2(s)} r^1(s,a^1,a^2)g(s,a^2),$$
$$p(s'|s,a^1,g)=\sum_{a^2\in A^2(s)} p(s'|s,a^1,a^2)g(s,a^2),$$

for all $s,s'\in S$ and $a^1\in A^1(s)$. We denote such an MDP problem by MDP(g). Similarly, for a fixed strategy f of player 1, player 2 faces an MDP problem with rewards and transition probabilities for all $s,s'\in S$ and $a^2\in A^2(s)$ given by

$$r^2(s,f,a^2)=\sum_{a^1\in A^1(s)} r^2(s,a^1,a^2)f(s,a^1),$$
$$p(s'|s,f,a^2)=\sum_{a^1\in A^1(s)} p(s'|s,a^1,a^2)f(s,a^1).$$

We denote such an MDP problem as MDP(f).

Lemma 1. 1) *A strategy f of player 1 is a best response to a fixed strategy g of player 2 if and only if there exists a vector v^1 such that for all $s\in S$*

$$v^1(s)=\max_{a^1\in A^1(s)}\big\{r^1(s,a^1,g)+\beta\sum_{s'\in S}p(s'|s,a^1,g)v^1(s')\big\}, \tag{4}$$

$$v^1(s)=r^1(s,a^1,g)+\beta\sum_{s'\in S}p(s'|s,a^1,g)v^1(s'), \tag{5}$$
$$\forall\ a^1\in A^1(s)\ \ \textit{such that}\ \ f(s,a^1)>0.$$

2) *A strategy g of player 2 is a best response to a fixed strategy f of player 1 if and only if there exists a vector v^2 such that for all $s\in S$*

$$v^2(s)=\max_{a^2\in A^2(s)}\big\{r^2(s,f,a^2)+\beta\sum_{s'\in S}p(s'|s,f,a^2)v^2(s')\big\}, \tag{6}$$

$$v^2(s)=r^2(s,f,a^2)+\beta\sum_{s'\in S}p(s'|s,f,a^2)v^2(s'), \tag{7}$$
$$\forall\ a^2\in A^2(s)\ \ \textit{such that}\ \ g(s,a^2)>0.$$

Proof. Let f be a best response to g. This implies that f is an optimal policy of the MDP(g) and $v^1_\beta(s,f,g)=\max_{\bar f\in F_S} v^1_\beta(s,\bar f,g)$ for all $s\in S$. Let $v^1=v^1_\beta(f,g)$, then it follows from optimality equation of MDP(g) that v^1 satisfies (4) [3]. The vector v^1 also satisfies (5) because if there exists an a^1 such that $f(s,a^1)>0$ and

$$v^1(s)>r(s,a^1,g)+\beta\sum_{s'\in S}p(s'|s,a^1,g)v^1(s'),$$

we get $v^1(s)>v^1_\beta(s,f,g)$ which gives a contradiction. Conversely, suppose a vector v^1 satisfies (4) and (5). Then, v^1 is a value vector of MDP(g) and $v^1(s)=v^1_\beta(s,f,g)$ for all $s\in S$. This implies that $v^1_\beta(s,f,g)=\max_{\bar f\in F_S} v^1_\beta(s,\bar f,g)$ for all $s\in S$ which proves that f is a best response of g. The proof of the second part follows using arguments similar to those above. □

It is clear from Lemma 1 that if (v^1,v^2,f,g) satisfies (4),(5),(6), and (7), (f,g) is a Nash equilibrium. By reformulating these constraints using integer variables, we propose four equivalent mixer integer bilinear programs.

Mixed integer formulation I

We propose a set of constraints involving integer variables and bilinear terms and show that its feasible solution generates a Nash equilibrium of the game G_β.

$$\begin{aligned}
&(i)\ v^1(s)\mathbf{1}_{|A^1(s)|}\geq \mathbf{R}^1(s)g(s)+\beta\mathbf{T}(s,v^1)g(s),\ \forall\, s\in S,\\
&(ii)\ v^2(s)\mathbf{1}^{\top}_{|A^2(s)|}\geq f^{\top}(s)\mathbf{R}^2(s)+\beta f^{\top}(s)\mathbf{T}(s,v^2),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\forall\, s\in S,\\
&(iii)\ (f,g)\in F_S\times G_S,\\
&(iv)\ f(s,a^1)\leq 1-h^1(s,a^1),\ \forall\, s\in S, a^1\in A^1(s),\\
&(v)\ g(s,a^2)\leq 1-h^2(s,a^2),\ \forall\, s\in S, a^2\in A^2(s),\\
&(vi)\ h^1(s,a^1)\in\{0,1\},\ \forall\, s\in S, a^1\in A^1(s),\\
&(vii)\ h^2(s,a^2)\in\{0,1\},\ \forall\, s\in S, a^2\in A^2(s),\\
&(viii)\ v^1(s)\mathbf{1}_{|A^1(s)|}-\mathbf{R}^1(s)g(s)-\beta\mathbf{T}(s,v^1)g(s)\\
&\qquad\qquad\qquad\qquad\leq K_1h^1(s),\forall\ s\in S,\\
&(ix)\ v^2(s)\mathbf{1}^{\top}_{|A^2(s)|}-f^{\top}(s)\mathbf{R}^2(s)-\beta f^{\top}(s)\mathbf{T}(s,v^2)\\
&\qquad\qquad\qquad\qquad\leq K_2\big(h^2(s)\big)^{\top},\ \forall\, s\in S,
\end{aligned} \tag{MIBP1}$$

where K_1 and K_2 are sufficiently large constants and $h^1(s) = (h^1(s,a^1))_{a^1 \in A^1(s)}$, $h^2(s) = (h^2(s,a^2))_{a^2 \in A^2(s)}$.

Theorem 2. *A strategy pair $(f,g) \in F_S \times G_S$ is a Nash equilibrium of the game G_β if and only if there exists (v^1, v^2, h^1, h^2) such that $(f, g, v^1, v^2, h^1, h^2)$ is a feasible solution of* (MIBP1).

Proof. Let (f,g) be a Nash equilibrium. Since f and g are the best response of each, it follows from Lemma 1 there exists v^1 and v^2 for which (v^1, v^2, f, g) satisfy $(i)-(iii)$ of (MIBP1). For all $s \in S$, $a^1 \in A^1(s)$ and $a^2 \in A^2(s)$, define integer variables as follows

$$h^1(s,a^1) := \begin{cases} 0 & \text{if } f(s,a^1) > 0 \\ 1, & \text{else.} \end{cases}$$
$$h^2(s,a^2) := \begin{cases} 0 & \text{if } g(s,a^2) > 0 \\ 1, & \text{else.} \end{cases} \tag{8}$$

The constraints $(iv)-(vii)$ of (MIBP1) are satisfied from the construction of integer variables given in (8). For all $s \in S$, $a^1 \in A^1(s)$ and $a^2 \in A^2(s)$ for which $f(s,a^1) > 0$ and $g(s,a^2) > 0$, it follows from (5), (7) and (8) that $(viii)$ and (ix) of (MIBP1) are satisfied. In all other cases, these constraints are satisfied because K_1 and K_2 are sufficiently large. To prove the converse part, let $(f, g, v^1, v^2, h^1, h^2)$ be a feasible solution of (MIBP1). It is enough to show that (v^1, v^2, f, g) satisfies (4), (5), (6), and (7). From constraint (i) of (MIBP1), for all $s \in S$, we have

$$\max_{a^1 \in A^1(s)} \left\{ r(s,a^1,g) + \beta \sum_{s' \in S} p(s'|s,a^1,g) v^1(s') \right\} \le v^1(s). \tag{9}$$

If $f(s,a^1) > 0$ for some $a^1 \in A^1(s)$, from (iv) of (MIBP1) $h^1(s,a^1) = 0$. Then, it follows from $(viii)$ of (MIBP1) that

$$v^1(s) = r(s,a^1,g) + \beta \sum_{s' \in S} p(s'|s,a^1,g) v^1(s').$$

Therefore, (9) becomes equality and (4), (5) are satisfied. Using the similar arguments as above, we can show that (6) and (7) are also satisfied. Thus, from Lemma 1 f and g are the best response of each other, which in turn implies that (f,g) is a Nash equilibrium of the game G_β. □

Now, we propose three different mixed integer bilinear optimization problems whose global optimization problem is equivalent to the Nash equilibrium problem of the game G_β.

Mixed integer formulation II

$$\min \sum_{s \in S} \Bigg[\sum_{a^1 \in A^1(s)} (\alpha^1(s,a^1) - K_1 h^1(s,a^1)) + \sum_{a^2 \in A^2(s)} (\alpha^2(s,a^2) - K_2 h^2(s,a^2)) \Bigg]$$
s.t. $(i)-(vii)$ of (MIBP1),
$$v^1(s)\mathbf{1}_{|A^1(s)|} - \mathbf{R}^1(s) g(s) - \beta \mathbf{T}(s, v^1) g(s) \le \alpha^1(s),\ \forall\ s \in S,$$
$$v^2(s)\mathbf{1}^\top_{|A^2(s)|} - f^\top(s)\mathbf{R}^2(s) - \beta f^\top(s)\mathbf{T}(s, v^2) \le \left(\alpha^2(s)\right)^\top,\ \forall\, s \in S,$$
$$\alpha^1(s) \ge K_1 h^1(s),\ \alpha^2(s) \ge K_2 h^2(s),\ \forall\ s \in S, \tag{MIBP2}$$

where $\alpha^i(s) = (\alpha^i(s,a^i))_{a^i \in A^i(s)}$, $i = 1, 2$ for all $s \in S$.

Theorem 3. *A strategy pair $(f,g) \in F_S \times G_S$ is a Nash equilibrium of the game G_β if and only if there exists $(v^1, v^2, h^1, h^2, \alpha^1, \alpha^2)$ such that $(f, g, v^1, v^2, h^1, h^2, \alpha^1, \alpha^2)$ is a global minimum of* (MIBP2) *with objective function value zero.*

Proof. Let (f,g) be a Nash equilibrium of the game. Define vectors h^1 and h^2 using (8). For all $s \in S$, $a^1 \in A^1(s)$ and $a^2 \in A^2(s)$, let $\alpha^1(s,a^1) = K_1 h^1(s,a^1)$ and $\alpha^2(s,a^2) = K_2 h^2(s,a^2)$. It follows from the same arguments used in the proof of Theorem 2 that there exist v^1 and v^2 such that $(f, g, v^1, v^2, h^1, h^2, \alpha^1, \alpha^2)$ is a feasible solution of (MIBP2) with zero objective function value. Such a point is also a global minimum because for any feasible solution the objective function of (MIBP2) is nonnegative. To prove converse part suppose $(f, g, v^1, v^2, h^1, h^2, \alpha^1, \alpha^2)$ is a global minimum of (MIBP2) with objective function value zero. Then, $\alpha^1(s,a^1) = K_1 h^1(s,a^1)$ and $\alpha^2(s,a^2) = K_2 h^2(s,a^2)$ for all $s \in S$, $a^1 \in A^1(s)$ and $a^2 \in A^2(s)$. This implies that $(f, g, v^1, v^2, h^1, h^2)$ is a feasible solution of (MIBP1). Thus, from Theorem 2 (f,g) is a Nash equilibrium of the game G_β. □

Mixed integer formulation III

$$\min \sum_{s \in S} \Bigg[\sum_{a^1 \in A^1(s)} (\gamma^1(s,a^1) - (1 - h^1(s,a^1))) + \sum_{a^2 \in A^2(s)} (\gamma^2(s,a^2) - (1 - h^2(s,a^2))) \Bigg]$$
s.t. $(i)-(iii), (vi)-(ix)$ of (MIBP1),
$$\gamma^1(s,a^1) \ge 1 - h^1(s,a^1),\ \forall\, s \in S, a^1 \in A^1(s),$$
$$\gamma^1(s,a^1) \ge f(s,a^1)\ \forall\, s \in S, a^1 \in A^1(s),$$
$$\gamma^2(s,a^2) \ge 1 - h^2(s,a^2), \forall\, s \in S, a^2 \in A^2(s),$$
$$\gamma^2(s,a^2) \ge g(s,a^2)\ \forall\, s \in S, a^2 \in A^2(s). \tag{MIBP3}$$

Theorem 4. *A strategy pair $(f,g) \in F_S \times G_S$ is a Nash equilibrium of the game G_β if and only if there exists $(v^1, v^2, h^1, h^2, \gamma^1, \gamma^2)$ such that $(f, g, v^1, v^2, h^1, h^2, \gamma^1, \gamma^2)$ is a global minimum of* (MIBP3) *with objective function value zero.*

Proof. Let (f,g) be a Nash equilibrium. Define h^1 and h^2 using (8), and let $\gamma^1(s,a^1) = 1 - h^1(s,a^1)$, $\gamma^2(s,a^2) = 1 - h^2(s,a^2)$ for all $s \in S$, $a^1 \in A^1(s)$, $a^2 \in A^2(s)$. Then using similar arguments used in Theorem 2 there exists v^1, v^2 such that $(f, g, v^1, v^2, h^1, h^2, \gamma^1, \gamma^2)$ is a feasible solution of (MIBP3) with zero objective function value which in turn implies that it is a global minimum of

(MIBP3). Suppose $(f, g, v^1, v^2, h^1, h^2, \gamma^1, \gamma^2)$ is a global minimum of (MIBP3) with objective function value zero. Then $\gamma^1(s, a^1) = 1 - h^1(s, a^1)$, $\gamma^2(s, a^2) = 1 - h^2(s, a^2)$ for all $s \in S$, $a^1 \in A^1(s)$, $a^2 \in A^2(s)$ which in turn implies that $(f, g, v^1, v^2, h^1, h^2)$ is a feasible solution of (MIBP1). Thus, (f, g) is a Nash equilibrium of the game. □

Mixed integer formulation IV

$$
\begin{aligned}
\min \sum_{s\in S} \Bigg[& \sum_{a^1\in A^1(s)} (\alpha^1(s,a^1) + \gamma^1(s,a^1)) \\
& + \sum_{a^2\in A^2(s)} (\alpha^2(s,a^2) + \gamma^2(s,a^2)) \Bigg] \\
\text{s.t. } & (i)-(iii),(vi)-(vii) \text{ of (MIBP1)}, \\
& \gamma^1(s,a^1) \geq 1 - h^1(s,a^1),\ \forall\, s \in S, a^1 \in A^1(s), \\
& \gamma^1(s,a^1) \geq f(s,a^1)\ \forall\, s \in S, a^1 \in A^1(s), \\
& \gamma^2(s,a^2) \geq 1 - h^2(s,a^2), \forall\, s \in S, a^2 \in A^2(s), \\
& \gamma^2(s,a^2) \geq g(s,a^2)\ \forall\, s \in S, a^2 \in A^2(s), \\
& \alpha^i(s,a^i) \geq h^i(s,a^i),\ \forall\ s \in S, a^i \in A^i(s), i = 1,2, \\
& v^1(s)\mathbf{1}_{|A^1(s)|} - \mathbf{R}^1(s)g(s) - \beta\mathbf{T}(s,v^1)g(s) \\
& \qquad \leq K_1\alpha^1(s),\ \forall\ s \in S, \\
& v^2(s)\mathbf{1}^\top_{|A^2(s)|} - f^\top(s)\mathbf{R}^2(s) - \beta f^\top(s)\mathbf{T}(s,v^2) \\
& \qquad \leq K_2\big(\alpha^2(s)\big)^\top,\ \forall\, s \in S.
\end{aligned}
\tag{MIBP4}
$$

Theorem 5. *A strategy pair $(f, g) \in F_S \times G_S$ is a Nash equilibrium of the game G_β if and only if there exists $(v^1, v^2, h^1, h^2, \gamma^1, \gamma^2, \alpha^1, \alpha^2)$ such that $(f, g, v^1, v^2, h^1, h^2, \gamma^1, \gamma^2, \alpha^1, \alpha^2)$ is a global minimum of* (MIBP4) *with objective function value $\sum_{s\in S} |A^1(s)| + |A^2(s)|$.*

Proof. Let (f, g) be a Nash equilibrium. Define h^1 and h^2 using (8), let $\gamma^i(s, a^i) = 1 - h^i(s, a^i)$, $\alpha^i(s, a^i) = h^i(s, a^i)$, for all $s \in S$, $a^i \in A^i(s)$, $i = 1, 2$. Then using similar arguments used in Theorem 2 there exists v^1, v^2 such that $(f, g, v^1, v^2, h^1, h^2, \gamma^1, \gamma^2, \alpha^1, \alpha^2)$ is a feasible solution of (MIBP4) with objective function value $\sum_{s\in S} |A^1(s)| + |A^2(s)|$. This implies that it is a global minimum of (MIBP4) because for every feasible solution of (MIBP4) the objective function value is at least $\sum_{s\in S} |A^1(s)| + |A^2(s)|$. Suppose $(f, g, v^1, v^2, h^1, h^2, \gamma^1, \gamma^2, \alpha^1, \alpha^2)$ is a global minimum of (MIBP4) whose objective value is $\sum_{s\in S} |A^1(s)| + |A^2(s)|$. Since each term of the objective function is at least 1, $\alpha^i(s, a^i) + \gamma^i(s, a^i) = 1$ for all $s \in S$, $a^i \in A^i(s)$, $i = 1, 2$. This implies that $\alpha^i(s, a^i) = h^i(s^i, a^i)$ for all $s \in S$, $a^i \in A^i(s)$, $i = 1, 2$. Using this $(f, g, v^1, v^2, h^1, h^2)$ becomes a feasible solution of (MIBP1) which implies that (f, g) is a Nash equilibrium. □

IV. Computational Results

In this section, we present computational experiments to assess the performance of the optimization formulations mentioned in Section III.

A. Experimental Setup and Instances

All experiments have been carried out on a workstation with Intel(R) Xeon(R) Silver 4114 processor with 40 CPUs sharing 64 GB RAM and a speed of 2.20 GHz. The optimization formulations described in Section III are implemented using AMPL[1] [7] and solved using the Octeract Engine v4.6.0 [22] and Baron v23.3.11 [18] which are both state-of-art global mixed-integer nonlinear optimization solvers. Octeract Engine implements a distributed branch-and-bound algorithm which is designed to run effectively on distributed-memory and multicore shared-memory processors. Baron deploys a branch-and-reduce algorithm using advanced techniques like constraint propagation, interval analysis, duality, etc., and also works in parallel on shared-memory parallel systems. For more information on different optimization solvers and their benchmarking, we refer the readers to [12].

We generated 8 instances with two players for our experiments. Each instance has 2 to 20 states (see the first column of Table I and Table II). For each state $s' \in S$, both players were assigned random integer payoff matrices with values ranging from 0 to 500. These values were generated using the `randint` function in Python. Also, the number of actions for both players is generated randomly, ranging from 1 to 5, again, using the `randint` function in Python. The transition probabilities $p(s'|s, a^1, a^2)$ are taken to be a fraction p/q such that the only prime factors of q are 2 and 5 for all $s', s \in S, a^1 \in A^1(s)$ and $a^2 \in A^2(s)$. The parameter β was set to 0.75 for all the instances, and the parameters K_i were set to $\max_{s\in S, a^1\in A^1(s), a^2\in A^2(s)} r^i(s, a^1, a^2)/(1 - \beta)$ for $i = 1, 2$ where the rewards $r^i(s, a^1, a^2)$ were positive. The wall clock time limit for each run was set to 2 hours.

B. Results and conclusions

Table I and Table II shows the detailed experimental results. The different mixed integer bilinear formulations are indicated using the acronym MIBP followed by the numbers 1 to 4, representing the formulations mentioned in Section III, respectively. Similarly, the columns for the two trilinear formulations are indicated by (T1) and (T2), respectively. The entries in Table I and Table II represent one of the following values: (a) wall clock time taken in seconds to solve the instance - shown in normal font, (b) the best upper bound value when the solver reached the time limit for all formulations except (MIBP4) - shown in bold font within parentheses, (c) the percentage gap - shown in parenthesis for (MIBP4); this is because the optimality gaps (with respect to the best lower bound on the optimal value) for only (MIBP4) are reasonably low. Otherwise, the symbol '–' is mentioned, indicating that the time limit was reached and no feasible solution was found.

Our computational experiments show that the trilinear formulation (T2) performs the best, with all instances being solved in less than 5 minutes, outperforming (T1) as well

[1]The AMPL models (.mod files), the data (.dat files) and other scripts are available at https://github.com/ppalkar/nashEquilibriumComputation.

TABLE I

PERFORMANCE OF DIFFERENT FORMULATIONS FOR TWO-PLAYER STOCHASTIC GAME INSTANCES USING OCTERACT ENGINE

# states	Mixed Integer Bilinear				Trilinear	
	MIBP1	MIBP2	MIBP3	MIBP4	T1	T2
2	0.04	0.18	0.24	0.28	40.39	0.03
3	–	**(7.58)**	**(6.00)**	*(0.04%)*	**(2.00e-04)**	0.42
4	–	**(1.68e+02)**	**(8.00)**	*(0.37%)*	**(2.27e-04)**	4.24
5	8.88	**(6.63e+02)**	**(1.0e+01)**	*(0.56%)*	**(5.92e-07)**	0.32
6	–	**(8.32e+02)**	**(1.2e+01)**	*(0.78%)*	**(3.36e-04)**	1.35
7	–	**(7.73e+02)**	**(1.4e+01)**	*(1.03%)*	**(1.12e+01)**	100.69
10	–	**(1.47e+03)**	**(2.0e+02)**	*(0.99%)*	**(8.48e+01)**	86.56
20	–	**(4.60e+03)**	**(4.0e+01)**	*(1.62%)*	**(1.76e+02)**	217.64

TABLE II

PERFORMANCE OF DIFFERENT FORMULATIONS FOR TWO-PLAYER STOCHASTIC GAME INSTANCES USING BARON SOLVER

# states	Mixed Integer Bilinear				Trilinear	
	MIBP1	MIBP2	MIBP3	MIBP4	T1	T2
2	0.03	0.16	0.28	0.06	0.11	0.03
3	–	**(1.64e+02)**	**(1.00)**	*(1.80%)*	**(4.54e-13)**	0.22
4	–	**(1.76e+02)**	2.33e+02	*(0.43%)*	**(-1.27-06)**	1.56
5	–	**(1.05e+02)**	**(1.58)**	*(0.08%)*	**(-7.74e-07)**	0.22
6	–	**(9.28e+01)**	**(1.97)**	*(1.03%)*	**(-2.15e-06)**	0.59
7	–	**(4.75e+02)**	**(2.97)**	*(0.57%)*	**(4.17e+01)***	106.55
10	–	**(1.55e+03)**	**(5.39)**	*(1.55%)*	**(4.30)***	43.75
20	–	**(4.15e+03)**	**(9.99)**	3.60e+02*	**(4.34)***	174.98

* Terminated within time limit dute to numerical issues in the solver.

as all the mixed integer bilinear formulations. The poor performance of (T1) for some instances could be attributed to its weaker relaxation compared to (T2), where the solver reached the optimal solution but spent a lot of time in verifying optimality. Among the mixed-integer bilinear formulations, (MIBP1) was solved for only a couple of instances by both solvers. Feasible solutions were obtained using (MIBP2), but most of them were quite far from optimality. (MIBP3) found better solutions, but the optimality gap could not be closed. However, (MIBP4), which could solve only one instance to optimality, seems promising as it exhibits optimality gaps close to 0 for the remaining instances. When comparing Octeract and Baron solvers, the latter performs better on (MIBP3) in terms of getting better solutions, while the former solves one more instance on (MIBP1). Baron is also slightly faster on (T2) but encounters numerical issues on some instances of (MIBP4) and (T1). Otherwise, there is no significant difference between the performance of the two solvers in our experiments.

Overall, (T2) seems to be the most effective and reliable formulation for practical purposes. Our results are consistent with those in [6] (where the authors study static games) in the sense that the feasibility multilinear formulation performs better compared to the other mixed-integer formulations for stochastic games. The mixed integer bilinear formulations could possibly be improved by using primal heuristics, tighter relaxations, cuts, etc., specific to stochastic games and could be explored further. So far, the research has only focused on random integer games with a uniform distribution. Subsequent studies might explore games with correlated outcomes, which may provide more authentic situations and understanding. Moreover, a future goal could be to design algorithms and techniques (e.g., cuts, primal heuristics, presolving, bound tightening, etc.) tailored to solve the game formulations. That would possibly enhance the performance of solvers and enable handling more intricate scenarios effectively. Furthermore, using machine learning methods to predict the success of different formulations based on the game's features might be a viable and advantageous strategy. This will enable better judgments in formulating strategies customized to particular attributes of the game being studied.

REFERENCES

[1] G. Debreu. A social equilibrium existence theorem. *Proceedings of National Academy of Sciences*, 38:886–893, 1952.

[2] K. Fan. Applications of a theorem concerning sets with convex sections. *Mathematische Annalen*, 163:189–203, 1966.

[3] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.

[4] J. A. Filar, T. A. Schultz, F. Thuijsman, and O. J. Vrieze. Nonlinear programming and stationary equilibria in stochastic games. *Mathematical Programming*, 50:227–237, 1991.

[5] A. M. Fink. Equilibrium in a stochastic n-person game,. *Journal of Science of Hiroshima University Series A-I Math*, 28:89–93, 1964.

[6] M. Fischer and A. Gupte. Multilinear formulations for computing a Nash equilibrium of multi-player games. In *International Symposium on Experimental Algorithms*, page 1–14, 2023.

[7] Robert Fourer, David M Gay, and Brian W Kernighan. *AMPL: A mathematical programming language*. Citeseer, 1987.

[8] N. Krishnamurthy and S. K. Neogy. On lemke processibility of lcp formulations for solving discounted switching control stochastic games. *Annals of Operations Research*, 295:633–644, 2020.

[9] C. Lemke. Bimatrix equilibrium points and mathematical programming. *Management Science*, 11:681–689, 1965.

[10] C. E. Lemke and J. T. Howson. Equilibrium points of bimatrix games. *Journal of the Society for Industrial and Applied Mathematic*, 12:413–423, 1964.

[11] O. L. Mangasarian and H. Stone. Two-person nonzero-sum games and quadratic programming. *Journal of Mathematical Analysis and Applications*, 9:348–355, 1964.

[12] Hans Mittelmann. Benchmarks for optimization software. https://plato.la.asu.edu/bench.html, accessed: June 03, 2024.

[13] J. F. Nash. Equilibrium points in n-person games. *Proceedings of National Academy of Sciences*, 36:48–49, 1950.

[14] T. Parthasarathy and T. E. S. Raghavan. An orderfield property for stochastic games when one player controls transition probabilities. *Journal of Optimization Theory and Applications*, 33:375–392, 1981.

[15] M.L. Puterman. *Markov Decision Processes*. John Wiley & Sons, Inc., New York, 1994.

[16] T. E. S. Raghavan and J. A. Filar. Algorithms for stochastic games–a survey. *Mathematical Methods of Operations Research*, 35:437–472, 1991.

[17] T. E. S. Raghavan and Z. Syed. Computing stationary Nash equilibria of undiscounted single-controller stochastic games. *Mathematics of Operations Research*, 27:384–400, 2002.

[18] Nikolaos V Sahinidis. Baron: A general purpose global optimization software package. *Journal of global optimization*, 8:201–205, 1996.

[19] T. Sandholm, A. Gilpin, and V. Conitzer. Mixed-integer programming methods for finding Nash equilibria. In *Proceedings of the 20th national conference on Artificial intelligence*, AAAI '05, pages 495–501, 2005.

[20] L.S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39:1095–1100, 1953.

[21] M. Takahashi. Equilibrium points of stochastic non-cooperative n-person games. *Journal of Science of Hiroshima University Series A-I Math*, 28:95–99, 1964.

[22] Octeract Team. Octeract engine. https://octeract.gg/octeract-engine/, accessed: June 03, 2024.