Algorithmic and Complexity Theoretic Aspects of Stochastic Games and Polystochastic Games

(Extended Abstract: Summary of Our Recent Results and Some New Results)

Nagarajan Krishnamurthy[1], Parthasarathy T.[2]

1. Indian Institute of Management, Indore, 2. Indian Statistical Institute, Chennai, India

nagarajan@iimidr.ac.in, pacha14@yahoo.com

1. Introduction

There is an increasing interest in the computational aspects of stochastic games. Recent years have seen various algorithms for solving different types of stochastic games as well as complexity theoretic results. In this paper, we provide a brief survey of some of these results including our recent results and we discuss some new results as well. We also mention some interesting and relevant open problems.

Solving a stochastic game involves computing its optimal value in the zero-sum case (Nash equilibrium payoffs in the non-zero-sum case) as well as optimal strategies (Nash equilibrium strategies) of the players. We start off with simple examples to illustrate how to solve a β-discounted stochastic game using Shapley's theorem (1953). We then discuss "The Big Match" with limiting time average (undiscounted) payoffs (Gillette, 1957; Blackwell and Ferguson, 1968) where one of the players does not have optimal strategies even if we allow behavioural strategies (strategies that depend on the history). On the other hand, discounted stochastic games always have stationary optimal strategies (Nash equilibrium strategies). (Stationary strategies depend only on the current state).

In the following sections, we discuss algorithms and complexity of solving stochastic games. We discuss our results on mixtures of stochastic games (Krishnamurthy, Parthasarathy and Ravindran, 2010) and make a few algorithmic observations which follow from this paper. We propose polynomial time algorithms for subclasses of Simple Stochastic Games (SSG), Perfect Information Stochastic

Games (PI) and Switching Control (SC) Stochastic Games (Krishnamurthy, Parthasarathy and Ravindran, 2012). Further, we discuss solving subclasses of multiplayer stochastic games via Linear Complementarity Problem (LCP) formulations (Krishnamurthy, Parthasarathy and Ravindran, 2011). We also discuss communication complexity of stochastic games where the players are at different nodes in a network and need to communicate in order to solve the game. We summarize results on the communication complexity of finding a pure equilibrium point (if one exists) for some classes of stochastic games (Krishnamurthy, Parthasarathy and Ravindran, 2009) and we solve a problem which was left open in that paper.

2. Algorithms and Complexity of Stochastic Games of Complete Information

We briefly outline a Newton-Raphson type iterative algorithm due to Pollatschek and Avi-Itzhak (1969) to solve two-person zero-sum discounted stochastic games. In general, we can use this algorithm only to find approximate optima as there is no guarantee on the number of steps such iterative algorithms take to converge. In fact, there are stochastic games (even in the discounted zero-sum case) for which no (finite arithmetic-step) exact algorithms exist. This is because such games have irrational optimal values though the inputs (payoffs, transition probabilities and the discount factor β in case of discounted games) are rational. We provide examples to illustrate this fact.

This motivates us to look for classes of stochastic games with rational inputs that are always guaranteed to have rational outputs (optimal value and a pair of optimal strategies of the players in the case of zero-sum games; a pair of Nash equilibrium payoffs and corresponding strategies in the case of non-zero-sum games). Such games are said to possess the orderfield property. To solve these games, there is hope of finding an exact algorithm. SSG (Simple Stochastic Games), One Player Control Stochastic Games, PI (Perfect Information Stochastic Games), SC (Switching Control Stochastic Games), SER-SIT (Separable Reward – State Independent Transition Stochastic Games) and ARAT (Additive Reward

Additive Transition Stochastic Games are some such classes with the orderfield property. We define these classes in the full version of the paper.

Algorithms are known for many of these classes where as, the problem remains open for some classes (even in the two-person case). We discuss some of these algorithms in the full version of the paper. Some of these algorithms reduce these stochastic games to a Linear Program, some of them reduce them to matrix or bimatrix games, some of them use policy-improvement techniques similar to those for MDPs (Markov Decision Processes) and others use Linear Complementarity Problems (LCP). For some classes of stochastic games, though algorithms for solving them are known, search is on for efficient algorithms to solve them. For example, there is no efficient algorithm (yet) to solve switching control stochastic games.

In many of these cases where algorithms have been proposed, nothing has been said about their complexity, though interesting complexity theoretic observations are just waiting to be mentioned. For example, two-person non-zero-sum SER-SIT games is in the complexity class PPAD (in both the discounted as well as the undiscounted case), two-person non-zero-sum one player control stochastic games is in PPAD (in the discounted case and in the undiscounted case when the transition matrix induced by any pair of pure strategies is irreducible), zero-sum SER-SIT as well as one player control games are in P, etc. Hardness results for many of these games remains open.

Known results for algorithms and complexity that we discuss include the results due to Filar (1981), Vrieze (1981), Parthasarathy and Raghavan (1981), Parthasarathy, Tijs and Vrieze (1984), Condon (1992, 1993), Nowak and Raghavan (1993), Conitzer and Sandholm (2003), Gartner and Rust (2005), Dieckelmann (2007), Gimbert and Horn (2009), Ganzfried and Sandholm (2009) etc. For a class of multi-player stochastic games, we discuss an algorithm by Mohan et al. (1997). Algorithms have also been proposed for mixtures of classes of stochastic games. For example, Neogy et al. (2008) propose an algorithm to solve a mixture of SC and ARAT states. Krishnamurthy, Parthasarathy and Ravindran (2010) discuss new sufficient conditions for mixtures of classes of stochastic games to possess the orderfield property. We discuss these in detail in the full version of the paper.

3. Polynomial Time Algorithms for Some Classes of Stochastic Games

Krishnamurthy, Parthasarathy and Ravindran (2012) propose polynomial time algorithms for subclasses of Simple Stochastic Games (SSG), Perfect Information (PI) Stochastic Games and Switching Control (SC) Stochastic Games with discounted as well as undiscounted payoffs.

We reproduce the following for the sake of completeness. Note that, as shown in Krishnamurthy, Parthasarathy and Ravindran (2012), though the algorithm seems to be simple backward induction, it does not work for mixtures of some classes of stochastic games and hence requires proof for classes where it works.

Polynomial Algorithm for Subclasses of SC:

Theorem: Let Γ_β = (S, A_1, A_2, r, q, β) be a finite zero-sum discounted switching control (SC) stochastic game with rational inputs and let G = (V, E) be the dependency graph of Γ_β where V = S. Let C_1, C_2, …, C_k be the strongly connected components of G with sets of vertices (states) V_1 (S_1), V_2 (S_2), …, V_k (S_k) respectively. That is, $S = S_1 \cup S_2 \cup \ldots \cup S_k$, ($S_{k_1} \cap S_{k_2} = \phi$, $k_1 \neq k_2$, $1 \leq k_1, k_2 \leq k$) such that the subsets S_1, S_2, …, S_k are cycle-free. Assume that S_k is a sink and that there are no transitions from S_{k_2} to S_{k_1} whenever $k_2 > k_1$. That is,

$$\sum_{s_k \in S_k} q(s_k \mid s, i, j) = 1 \quad \text{for all } s \in S_k,\ i \in A_1,\ j \in A_2,$$

and $q(s_{k_1} \mid s_{k_2}, i, j) = 0$ for all $s_{k_1} \in S_{k_1}$, $s_{k_2} \in S_{k_2}$, ($k_2 > k_1$), $i \in A_1$, $j \in A_2$.

Furthermore, assume that each S_h ($1 \leq h \leq k$) satisfies one of the following:

(i) S_h is player 1 controlled or

(ii) S_h is player 2 controlled or

(iii) S_h consists of both player 1 and player 2 controlled states but all states in S_h are those of perfect information. Furthermore

(a) one of the players has O(log N) states with constant number of actions and a constant number of states with a non-constant number of actions or

(b) states in S_h resemble an SSG such that if S_h has all 3 types of states (namely, player 1, player 2 and random states), then S_h has either O(log N) player 1 states or O(log N) player 2 states or a constant number of random vertices. (By "resemble an SSG" we mean the following. When player 1 or player 2 has 2 or more actions in a state, the probabilities of transition from that state are 0/ 1).

Then Γ_β can be solved in polynomial time.

[Note:

(1) We do not require the game to already be partitioned into $S_1, S_2, \ldots, S_k$.

(2) We need to show that, unlike the SER-SIT case, we can recursively plug in values here without altering the overall structural properties. Krishnamurthy, Parthasarathy and Ravindran (2010) show that such games have the orderfield property. Here, we require that the structure is not altered and further, that the resulting game is solvable in polynomial time].

Proof: We prove the theorem by induction on k, the number of subsets.

We shall prove the theorem when each subset S_h ($1 \leq h \leq k$) is player 1 controlled (assumption (i) in the theorem) or player 2 controlled (assumption (ii) in the theorem). Using polynomial time algorithms for subclasses of PI and SSG that we shall be proving in subsequent sections, using the fact that SSG $\subset$ PI $\subset$ SC and using techniques similar to those in the proof below, we can prove the theorem by allowing assumptions (iii) (a) and (iii) (b) as well.

We briefly outline the proof for assumptions (i) and (ii) below.

For $k = 2$, $S = S_1 \cup S_2$, $S_1 \cap S_2 = \phi$, such that

$\sum_{s_2 \in S_2} q(s_2 \mid s,i,j) = 1$ for all s $\in$ S_2, i $\in$ A_1, j $\in$ A_2.

In this case, it follows that $q(s_1 \mid s_2, i, j) = 0$ for all $s_1 \in S_1$, $s_2 \in S_2$, i $\in$ A_1, j $\in$ A_2.

If both S_1 and S_2 are controlled by the same player, then Γ_β is a one-player control game and can be reduced to a matrix game or equivalently an LP (Parthasarathy and Raghavan, 1981).

Now, without loss of generality, let S_1 be controlled by player 1 and S_2 by player 2. That is,

$q(s \mid s_1, i, j) = q(s \mid s_1, i)$, for all $s_1 \in S_1$, $s \in S$, $i \in A_1$, $j \in A_2$,

$q(s \mid s_2, i, j) = q(s \mid s_2, j)$, for all $s_2 \in S_2$, $s \in S$, $i \in A_1$, $j \in A_2$.

As there are no transitions from S_2 to S_1, $\Gamma_\beta| S_2$ is a one player controlled game, possesses the orderfield property and can be solved by reduction to LP. Let (f_2^*, g_2^*) be a pair of optimal strategies of the players in the sub-game $\Gamma_\beta| S_2$. Using Shapley's theorem, value of the stochastic game $\Gamma_\beta| S_2$ starting at state $s_2 \in S_2$ is

$$v'_\beta(s_2) = r(s_2, f_2^*(s_2), g_2^*(s_2)) + \beta \sum_{s \in S_2} q(s \mid s_2, g_2^*(s_2)) v_\beta(s).$$

Now, define a new game $\Gamma_\beta' = (S' = S \cup \{s^*\}, A_1, A_2, r', q', \beta)$ where s^* is a new absorbing state such that

$r'(s^*, i, j) = 0$, $\forall\, i \in A_1$, $\forall\, j \in A_2$.

$$r'(s,i,j) = r(s,i,j) + \beta \sum_{s_2 \in S_2} q(s_2 \mid s,i) v_\beta(s_2), \forall s \in S_1, \forall i \in A_1, \forall j \in A_2$$

$r'(s, i, j) = r(s, i, j)$, $\forall\, s \in S_2$, $\forall\, i \in A_1$, $\forall\, j \in A_2$.

$$q'(s^* \mid s,i) = 1 - \sum_{s_1 \in S_1} q(s_1 \mid s,i), \forall s \in S_1, \forall i \in A_1.$$

$q'(s_1 \mid s, i) = q(s_1 \mid s, i)$, $\forall\, s \in S_1$, $\forall\, s_1 \in S_1$, $\forall\, i \in A_1$.

$q'(s_2 \mid s, i) = 0$, $\forall\, s \in S_1$, $\forall\, s_2 \in S_2$, $\forall\, i \in A_1$.

$q'(s_2 \mid s, i, j) = q\,(s_2 \mid s, i, j), \forall s, s_2 \in S_2, \forall i \in A_1, \forall j \in A_2.$

$q'(s^* \mid s^*, i, j) = 1, \forall\, i \in A_1, \forall\, j \in A_2.$

This new game Γ_β' consists of two independent sub-games, $\Gamma_\beta' | S_1$ and $\Gamma_\beta' | S_2$. $\Gamma_\beta' | S_2$ is the same player 2 controlled game as $\Gamma_\beta | S_2$, and hence their optimal values are equal and (f_2^*, g_2^*) is a pair of optimal strategies for the players in $\Gamma_\beta' | S_2$ as well.

$\Gamma_\beta' | S_1 \cup \{s^*\}$ is also a one player control game (controlled by player 1) with rational entries (r' is rational as $v_\beta(s_2)$ is rational for all $s_2 \in S_2$ and as r is rational, q' is rational by definition of q' as q is rational and β is already given to be rational). Hence this sub-game can be solved in polynomial time by reducing it to a matrix game (or LP). (Note that it is important that the inputs to $\Gamma_\beta' | S_1$ are rational and it is also important that the resulting game $\Gamma_\beta' | S_1$ is controlled by the player who controls S_1, otherwise our claim does not hold).

Now, we show that the optimal value and optimal strategies coincide for all $s \in S_1$ as well. For all $s \in S_1$, Shapley equations for $\Gamma_\beta' | S_1$ are

$$v'_\beta(s) = val(r'(s,i,j) + \beta \sum_{s_1 \in S_1} q(s_1 \mid s,i) v'_\beta(s_1))$$

and Shapley equations for the game Γ_β are,

$$v_\beta(s) = val(r(s,i,j) + \beta \sum_{s' \in S} q(s' \mid s,i) v_\beta(s')).$$

Plugging in the values of r' from above equation, namely

$$r'(s,i,j) = r(s,i,j) + \beta \sum_{s_2 \in S_2} q(s_2 \mid s,i) v_\beta(s_2), \forall s \in S_1, \forall i \in A_1, \forall j \in A_2$$

it is easy to see that $v_\beta(s) = v'_\beta(s)$ for all $s \in S_1$.

We formally describe the algorithm (algorithm 2) below and it is easy to see that it runs in polynomial time. $\square$

Algorithm 1 below is a pre-processor, that accepts a stochastic game Γ_β, constructs its dependency graph G, finds the strongly connected components of G, topologically sorts these components and checks that the conditions of the above theorem hold. All these are linear time operations. Algorithm 2 solves Γ_β (given that the conditions of the above theorem hold) by reducing Γ_β to Γ'_β (which is a linear time operation) and solves Γ'_β (in polynomial time). We can solve subclasses of undiscounted SC games using similar techniques and using the reduction to LP by Vrieze (1981).

Input: 2-Player Zero-Sum Discounted Stochastic game $\Gamma_\beta = (S, A_1, A_2, r, q, \beta)$.

Output: Outputs “yes” if the conditions of the above theorem are satisfied, “no” otherwise.

(1) Construct the dependency graph G of Γ_β.
(2) Find the strongly connected components of G. Let them be $C_1, C_2, \ldots, C_k$ corresponding to subsets of states $S_1, S_2, \ldots, S_k$ respectively.
(3) Topologically sort these components.
(4) If each component satisfies the conditions of the theorem 2, return “yes”. Otherwise “no”.

Algorithm 1: To partition a stochastic game into strongly connected components and check if conditions of the above theorem are satisfied.

Input: 2-Player Zero-Sum Discounted Switching Control Stochastic game $\Gamma_\beta = (S, A_1, A_2, r, q, \beta)$ that satisfies the conditions (i) or (ii) of theorem 2.

Output: Optimal value vector of Γ_β and a pair of optimal stationary strategies (f^*, g^*) of players 1 & 2.

(1) Solve the sub-games corresponding to sinks via reduction to LP.
(2) Plug-in these values in their predecessor classes and solve recursively using backward induction. (Plug-in as discussed in the proof of theorem).

Algorithm 2: To solve subclasses of SC

Refer to Krishnamurthy, Parthasarathy and Ravindran (2012) for polynomial time algorithm for subclasses of PI and SSG.

4. Solving Polystochastic Games via Linear Complementarity Problem (LCP) Formulations

Krishnamurthy, Parthasarathy and Ravindran (2011) prove that certain new subclasses and mixtures of multi-player (or n-person) stochastic games can be solved via LCP formulations. Mohan, Neogy and Parthasarathy (1997) proposed an LCP formulation of β-discounted (multi-player) polystochastic games where the transitions are controlled by one player, and proved that this LCP is processible by Lemke's algorithm. Using this formulation repeatedly, we prove that we can solve a subclass of β-discounted switching control polystochastic games. As our proof is constructive, we have an algorithm for solving this subclass. This algorithm only involves iteratively solving different LCPs and hence, it follows that this subclass has the orderfield property, a question left open in the paper on orderfield property of mixtures of stochastic games by Krishnamurthy, Parthasarathy and Ravindran (2010). Furthermore, we use results from Krishnamurthy, Parthasarathy and Ravindran (2010) to solve some mixture classes using LCP (or VLCP) formulations. We also propose two different VLCP formulations for β-discounted zero-sum perfect information stochastic games, the underlying matrices of both formulations being R_0. As a result, we also have an alternative proof of the orderfield property of such games.

5. Communication Complexity of Stochastic Games

Krishnamurthy, Parthasarathy and Ravindran (2009) study the problem of determining the existence of pure Nash equilibria when each player knows only his or her payoffs and not that of the opponent. The aim of the players is to communicate with each other according to some pre-defined protocol and find whether the game has a pure strategy Nash equilibrium or not. The paper discusses finding the communication complexity for SER-SIT games and single controller stochastic games. We summarize these results and we extend these results to undiscounted SER-SIT games, a case left open in the paper.

6. Acknowledgements

The authors would like to thank G. Ravindran for useful inputs.

7. References

Condon, A. The Complexity of Stochastic Games. *Info. and Computation*, 96, (1992), pp. 203-224.

Condon, A. On algorithms for simple stochastic games. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 13, (1993), pp. 51–71.

Blackwell, D. & Ferguson, TS. The Big Match. *Ann. of Math. Statistics*, 39, (1968), pp. 159-168.

Conitzer, V. & Sandholm, T. Complexity Results about Nash Equilibria, *Proc. of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03)*, (2003), pp. 765-771.

Dieckelmann, J. The Complexity of Simple Stochastic Games, *eprint*, arXiv:0704.2779v1 [cs.CC], 2007.

Filar, J. A. Orderfield Property of Stochastic Games when the Player who Controls the Transition Changes from State to State. *J.O.T.A.*, 34, (1981), pp. 505-517.

Ganzfried, S. & Sandholm, T. Computing equilibria in multiplayer stochastic games of imperfect information, *Proceedings of the 21st International Joint Conference on Artifical Intelligence (IJCAI)*, (2009), pp.140-146.

Gartner, B. and Rust, L. Simple stochastic games and P-matrix generalized linear complementarity problems. *Fundamentals of Computation Theory*, 15, 2005.

Gillette, D. Stochastic Games with Zero-Stop Probabilities. *Contributions to the Theory of Games, Ann. of Math. Studies, Princeton University Press*, 39, (1957), pp. 179-188.

Gimbert, H. & Horn, F. Solving Simple Stochastic Games with Few Random Vertices. *Logical Method in Computer Science*, 5(2:9), (2009), pp. 1-17.

Krishnamurthy, N.; Parthasarathy, T. & Ravindran, G. Communication Complexity of Stochastic Games. *Proc. of the Int. Conf. on Game Theory for Networks (GameNets), IEEEXplore*, (2009), pp. 411-417.

Krishnamurthy, N.; Parthasarathy, T. & Ravindran, G. Orderfield Property of Mixtures of Stochastic Games. *Sankhya: The Indian Journal of Statistics*, Springer, 72-A(1), (2010), pp. 246-275.
Retrieved from http://sankhya.isical.ac.in/search/72a1/final16.pdf

Krishnamurthy, N., Parthasarathy, T. & Ravindran, G. On Solving Subclasses of Multi-Player Stochastic Games via Linear Complementarity Problem Formulations: A Survey and Some New Results. *Journal of Optimization and Engineering*, Springer, (2011). doi: 10.1007/s11081-011-9163-1

Krishnamurthy, N.; Parthasarathy, T. & Ravindran, G. Polynomial time algorithms for sub-classes of simple stochastic games and switching control stochastic games of perfect as well as imperfect information. *Proc. of the Int. Conf. on Applications of Game Theory in Policies and Decisions (in Honor of Professor C. R. Rao)*, Hyderabad, India. (2012).

Mohan, SR.; Neogy SK. & Parthasarathy, T. Linear Complementarity and Discounted Polystochastic Game when One Player Controls Transitions. *Proceedings of the International Conference on Complementarity Problems*, Eds. Michael C. Ferris, Johg-Shi Pang, *SIAM*, (1997), pp. 284-294.

Neogy, SK.; Das, AK.; Sinha, S. & Gupta, A. On a Mixture Class of Stochastic Game with Ordered Field Property. S. K. Neogy, R. B. Bapat, A. K. Das and T. Parthasarathy, eds. *Mathematical Programming and Game Theory for Decision Making, World Scientific,* Vol. 1. (2008), pp. 451-477.

Nowak, AS. and Raghavan, T. E. S. A finite step algorithm via a bimatrix game to a single controller non-zero sum stochastic game. *Math. Programming.* 59, (1993), pp. 249-259.

Parthasarathy, T. and Raghavan, T. E. S. An Orderfield Property for Stochastic Games when One Player Controls Transition Probabilities. *J.O.T.A.,* 33(3), (1981), pp. 375-392.

Parthasarathy, T.; Tijs, SJ. & Vrieze, OJ. Stochastic Games with State Independent Transitions and Separable Rewards. *Lecture Notes in Economics/ Math. Systems*, 226, (1984), pp. 262-271.

Pollatschek, MA. & Avi-Itzhak, B. Algorithms for Stochastic Games with Geometrical Interpretation, *Management Science*, 15(7), (1969), pp. 399-415.

Shapley, L. Stochastic Games. *Proceedings of the National Academy of Sciences*, 39, (1953), pp. 1095-1100.

Vrieze, OJ. Linear Programming and Undiscounted Stochastic Game in which One Player Controls Transitions. *O.R. Spektrum*, 3, (1981), pp. 29-35.