Markov games : an annotated bibliography

TECHNOLOGICAL UNIVERSITY EINDHOVEN

Department of Mathematics

STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 75-09

Markov games: An annotated bibliography

by

J. van der Wal

Eindhoven, July 1975

Markov games: An annotated bibliography

J. van der Wal

The purpose of this memorandum is to list and abstract a number of papers on the subject of Markov games.

The number of papers listed here will be relatively small therefore we will order them chronologically on the date of publishing and not alphabetically.

It seems appropriate to formulate the concept of a Markov game before we give the abstracts. It also has the advantage that most of the notions mentioned in the abstracts have already been introduced.

A Markov game (many authors use the term stochastic game) is generally a game played by two players. Connected with the game is a set of states (or positions of the game). At some epochs the state of the game is observed: then both players select an action out of the set of actions available to them. As a result of these (two) actions the state of the game is changed and the players receive some (possibly negative) amount.

More formally we could describe the game as follows. Consider a dynamic system with state space S, the behavior of which is influenced by two players, P_1 and P_2 (the extension to more than two players is obvious). In each state $x \in S$ two nonempty sets of actions exist, one for each player, denoted by K_x for P_1 and L_x for P_2. At times $t = 0,1,2,\dots$ both players select an action out of the set available to them. As a joint result of the state x and the two selected actions, k for P_1 and ℓ for P_2, the system moves to a new state according to the probability distribution $p(\cdot|x,k,\ell)$. Moreover P_1 and P_2 receive some amount $r_1(x,k,\ell)$ and $r_2(x,k,\ell)$ respectively.

Most papers listed below deal with the zero-sum game i.e. the special case that $r_1(x,k,\ell) = -r_2(x,k,\ell)$ for all x, k and ℓ.

For convenience we will restrict ourselves from now on to the description of the zero-sum game.

Instead of considering the infinite horizon game one may also consider the truncated game, which is terminated at time T with P_1 obtaining a final payoff q(y) if the terminal state is y. This game we call the T-stage Markov game with final payoff q.

A strategy π for P_1 for a Markov game is any rule which prescribes for each state $x \in S$ and for each time t a randomization over the actions which may be taken as a function of the prior states of the system and previously taken actions. If these prescriptions do not depend on the prior states and actions the strategy is called a Markov strategy. If moreover they are independent of t the strategy is called stationary. Analogously for a strategy p for P_2. In order to compare strategies for the infinite horizon Markov game one may use one of the following three criteria:

i) total expected discounted rewards (future payoffs are discounted at a rate β, $0 \leq \beta < 1$) When this criterion is used we speak of the discounted game,

ii) average rewards per unit time,

iii) total expected rewards.

i) may be viewed at as a special case of iii) which is easily seen by interpreting $1 - \beta$ as a probability that the game vanishes and $\beta p(\cdot|x,k,\ell)$ as new probability distribution. Let us restrict ourselves to the discounted game. We may define a payoff function $V(\pi,\rho)$ for any pair of strategies π and ρ. $V(\pi,\rho)(x)$ denotes the total expected discounted income for P_1 over the duration of the game when at $t = 0$ the state of the game is x.
The game with starting state x is said to have a value v(x) if

$$\sup_{\pi} \inf_{\rho} V(\pi,\rho)(x) = v(x) = \inf_{\rho} \sup_{\pi} V(\pi,\rho)(x) .$$

For nonzero-sum games we have to replace the notion of a value by that of an, appropriately defined, equilibrium point.
A strategy π_ε for P_1 will be called ε-optimal if for all x and ρ

$$V(\pi_\varepsilon,\rho)(x) \geq v(x) - \varepsilon .$$

A 0-optimal strategy will be called optimal.
For the criteria ii) and iii) a payoff function, a value and (ε-)optimal strategies may be defined in a similar way.
Unless stated specifically the papers listed below deal with the infinite horizon two-person zero-sum game.

[1] SHAPLEY, L.S.; *Stochastic games*. Proc. Nat. Acad. Sci. USA 39 (1953), 1095-1100.

Shapley considers the Markov game when state and actions spaces are all finite. Moreover he demands $\sum_{y \in S} p(y|x,k,\ell) < 1$ for all x, k and ℓ. Using the criterion of total expected rewards Shapley proves that the game has a value and that optimal stationary strategies exist. The fact that the operator on $\mathbb{R}^N$ which maps a vector v on the value vector of the 1-stage game with final payoff v is contractive, results in a successive approximation algorithm. The discounted game is a special case of Shapley's game.

[2] EVERETT, J.; *Recursive games.* Contributions to the theory of games III, ed. M. Dresher, A.W. Tucker and P. Wolfe. Princeton Univ. Press, Princeton, New Yersey, 1957, 47-78.

The recursive game is a Markov game with finite state and general state space. It is assumed that there is no payoff as long as the game continuous, there are only terminal payoffs. (One may formulate an equivalent of Shapley's game [1] which satisfies this condition.) There might be a positive probability that the recursive game goes on for ever. It is shown that, if for each vector v the 1-stage game with final payoff v has a value, the recursive game has a value and that there exist ε-optimal strategies.

[3] GILLETTE, D.; *Stochastic games with zero stop probabilities.* Contributions to the Theory of Games III, ed. M. Dresher, A.W. Tucker and P. Wolfe. Princeton Univ. Press. Princeton, New Yersey, 1957, 179-187.

Gillette considers the game with finite state and action spaces, using the criterion of average rewards. In order to prove that two special types of this game have a value and that optimal stationary strategies exist the author uses an extension of a Hardy and Littlewood theorem which is incorrect as is shown by Liggett and Lippman [16].

[4] TAKAHASHI, M.; *Stochastic games with infinitely many strategies.* J. Sci. Hiroshima Univ. Ser A. I <u>26</u> (1962), 123-134.

This paper considers the Markov game with finite state space and general action spaces. It is assumed that $\sup_{x,k,\ell} |r(x,k,\ell)| < \infty$ and that

$$\sup_{x,k,\ell} \int_S dp(y|x,k,\ell) = s < 1 \ .$$

It is shown that when K_x and L_x are compact and $r(x,k,\ell)$ and $p(y|x,k,\ell)$ continuous on $K_x \times L_x$ for all $x \in S$ the game has a value and both players have optimal strategies, and that under a little weaker conditions the game has a value.

[5] BENIEST, W.; *Jeux stochastiques totalement cooperatifs arbitres*. Cahiers du Centre d'Etudes de Recherche Operationnelle 5 (1963), 124-138.

This paper considers the non-zero sum Markov game when state and action spaces are all finite and the transition probabilities satisfy $\sum_{y \in S} p(y|x,k,\ell) < 1$ for all x, k and ℓ. For two different cooperation schemes it is shown that the game has an equilibrium point.

[6] ZACHRISSON, L.E.; *Markov games*. Advances in Game Theory, ed. M. Dresher, L.S. Shapley and A.W. Tucker. Princeton Univ. Press. Princeton, New Yersey, 1964, 211-253.

The first part of this paper deals with the T-stage Markov game, with finite state space and compact action spaces. Moreover the author demands that the immediate payoffs do not depend on the actions taken and that the transition probabilities $p(y|x,k,\ell)$ are continuously in the actions k and ℓ simultaneously. It is shown that this game has a value, that optimal strategies exist and that they may be determined by a dynamic programming approach. In the case that the action spaces are finite and the transition probabilities satisfy $\sum_{y \in S} p(y|x,k,\ell) < 1$ for all x, k and ℓ. Zachrisson obtains the same results as Shapley [1] by letting T tend to infinity.

[7] HOFFMAN, A.K. and R.M. KARP; *On nonterminating stochastic games*. Management Science 12 (1966), 359-370.

The authors consider the Markov game with finite state and action spaces and they use the criterion of average rewards. For the case that every pair of stationary strategies yields an irreducible Markov chain the following algorithm is presented, which approximates the value of the game and determines ε-optimal strategies.

Algorithm:

i) Select an initial stationary strategy for P_2.

ii) Solve the resulting Markov decision process, i.e. determine the gain and the relative value vector v.

iii) Determine an optimal strategy for P_2 for the 1-stage Markov game with final payoff v. Let P_2 play the corresponding stationary strategy and continue with ii).

[8] RIOS, S. and I. YANEZ; *Programmation Sequentielle en Concurrence.* Research papers in Statistics. Edited by F.N. David. John Wiley and Sons. London, New York, Sydney 1966, 289-299.

This paper deals with the Markov game with time average payoffs when state and action spaces are all finite. The authors show that when $p(y|x,k,\ell) > 0$ holds for all y, x, k and ℓ one may use a successive iteration technique in order to approximate the value of the game and to obtain ε-optimal strategies. The approach is similar to White's in J. Math. Anal. Appl. 6 (1963), 373-376 for Markov decision processes.

[9] CHARNES, A. and R.G. SCHROEDER; *On some stochastic tactical antisubmarine games.* NRLQ 14 (1967), 291-311.

This paper deals with the Markov game with finite state and action spaces when the transition probabilities satisfy $\sum_{y\in S} p(y|x,k,\ell) < 1$ for all x, k and ℓ. The authors use the criterion of total expected rewards. For Shapley's successive approximation algorithm [1] bounds for the value of the game are derived. Also some attention is paid to the case that for some x, k and ℓ $\sum_{y\in S} p(y|x,k,\ell) = 1$ holds and to the finite-stage Markov game.

[10] DENARDO, E.V.; *Contraction mappings in the theory underlying dynamic programming.* SIAM Review 9 (1967), 165-177.

This paper analyzes a broad class of optimization problems including many dynamic programming problems. Three properties of optimization problems are considered called "contraction", "monotonicity" and "N-stage contraction". Shapley's stochastic game (see [1]) is reviewed as an example to illustrate Denardo's analysis.

[11] BLACKWELL, D. and T.S. FERGUSON; *The big math.* Ann. Math. Statist. 39 (1968), 159-163.

The big match is a Markov game with three states and finitely many actions. It is played as follows in state 1 both players have to select a number, 0 or 1. If they choose the same number P_1 wins a unit, otherwise there is no payoff. The game stays in state 1 as long as P_1 choose 0. If he choose 1 and P_2 0 the game moves to state 2 and if P_2 choose 1 the games moves to state 3. In states 2 and 3 the system will remain for ever, in 2 there is no payoff in 3 P_1 receives 1 every unit of time. It is shown that with the criterion of average rewards the game has a value, that P_2 has an optimal stationary strategy and that P_1 has an ε-optimal strategy but no optimal one.

[12] OWEN, G.; *Game Theory*, W.B. Saunders Company. Philadelphia, London, Toronto, 1968, 98-112.

In this part of chapter V: "multi-stage games", several types of Markov games with finite state and action spaces are considered. Amoung them Shapley's stochastic game and Everett's recursive game [2]. Also some examples are presented.

[13] FRID, E.B.; *The optimal stopping rule for a two-person Markov chain with opposing interests*. Theory Prob. Applications 14 (1969), 714-716.

Frid considers a special Markov game with an arbitrary state space S, on which two disjoint subsets S_1 and S_2 are defined, and finite action spaces. The players have no influence on the state of the system; they can only quit playing: P_1 on S_1 and P_2 on S_2. Thus the game has perfect information. If the game is stopped in state x P_2 receives a final payoff g(x). All other immediate payoffs are zero. It is shown that this game, using the criterion of total expected rewards, has a pure value and that pure stationary optimal strategies exist.

[14] KIFER, T.I.; *Optimal strategy in games with an unbounded sequence of moves*. Theory Prob. Applications 14 (1969), 279-286.

A Markov game with finite state and action spaces is considered in which in each state x either K_x or L_x consists of only one element and the transition probabilities $p(y|x,k,\ell)$ are either 0 or 1. It is shown that the discounted game and the game with average payoffs have a value and that pure stationary optimal strategies for both players exist.

[15] KUSHNER, H.J. and S.G. CHAMBERLAIN; *Finite state stochastic games: Existence theorems and computational procedures*. IEEE Trans. Automatic Control. 14 (1964), 248-255.

This paper considers the Markov game with finite state and compact actions spaces under the criterion of total expected rewards. The following assumptions are considered. A_0: $\sup_{x,k,\ell} |r(x,k,\ell)| < \infty$, A_1: for each pair of strategies the game terminates with probability p_1 before time N (= |S|).

A_2: $\inf_{x,k,\ell} r(x,k,\ell) \geq \delta > 0$ and P_2 can stop the game before time N with probability $p_2 > 0$. A_3:

$$\sup_k \inf_\ell [r(x,k,\ell) + \sum_{y \in S} p(y|x,k,\ell)q(y)] =$$

$$= \inf_\ell \sup_k [r(x,k,\ell) + \sum_{y \in S} p(y|x,k,\ell)q(y)] \quad \text{for any } q \in \mathbb{R}^N .$$

A_4: $p(y|x,k,\ell)$ and $r(x,k,\ell)$ are continuous in k and ℓ for all $x \in S$, $y \in S$. It is shown that under $A_0A_1A_3A_4$ or $A_0A_2A_3A_4$ the game has a value and that pure optimal strategies exist. Moreover that successive approximation yields an approximation of the value and ε-optimal strategies. The assumption of compactness may be weakened.

[16] LIGGETT, T.M. and S.A. Lippman; *Stochastic games with perfect information and time average payoff.* SIAM Review 11 (1969), 604-607.

This paper considers the Markov game with finitely many states and actions. A counterexample to an alleged extension of the Hardy-Littlewood theorem (see Gillette [3]) is given and the optimality of stationary strategies for stochastic games of perfect information with time average payoffs is established.

[17] POLLATSCHEK, M.A. and B. AVI-ITZHAK; *Algorithms for stochastic games with geometrical interpretation.* Management Science 15 (1969), 399-415.

Two algorithms are presented for games with finite state and finite action spaces, when the transition probabilities satisfy $\sum_{y \in S} p(y|x,k,\ell) < 1$ for all x, k and ℓ. The first algorithm is essentially analogous to the algorithm suggested by Hoffmand and Karp [7] for the game with the average reward per unit time criterion. The second algorithm is an extension of Howards algorithm for Markov decision processes to Markov games. The authors prove convergence of the latter algorithm under very strong conditions only.

[18] ROGERS, P.D.; *Nonzero-sum stochastic games.* Report ORC 69-8, Operations Research Center, University of California, Berkeley (1969).

Rogers considers noncooperative two-person Markov games when state and action spaces are finite. It is shown that equilibrium points exist for the discounted game as well as for the game with average payoffs. Extensions to n-person games and underlying semi-Markov processes are discussed.

[19] FOX, B.L.; *Finite-state approximations to denumerable-state dynamic programs*. Rand Corporation, RM-6195-RIZ February 1970.

In this paper it is shown how to find a sequence of policies for essentially finite state dynamic programs such that the corresponding vector of optimal returns converges pointwise to that of a denumerable state dynamic program. The corresponding result for discounted Markov games is also given.

[20] MAITRA, A. and T. PARTHASARATHY; *On stochastic games*. J. Opt. Theory Appl. 5 (1970), 289-300.

The authors consider the discounted Markov game when state and action spaces are all compact metric spaces. Moreover it is assumed that the action spaces are identical in each state, that $r(x,k,\ell)$ is continuous in x, k and ℓ and that $(x_n,k_n,\ell_n) \to (x_0,k_0,\ell_0)$ implies that $p(\cdot|x_n,k_n,\ell_n)$ converges weakly to $p(\cdot|x_0,k_0,\ell_0)$. It is shown that the game has a value and that both players have optimal strategies.

[21] MAITRA, A. and T. PARTHASARATHY; *On stochastic games*, II. J. Opt. Theory Appl. 8 (1971), 154-160.

This paper deals with positive stochastic games, i.e. Markov games in which $r(x,k,\ell)$ is nonnegative for all x, k and ℓ. State and action sets are all compact metric spaces and the criterion used is that of total expected rewards. Again the action spaces are assumed to be identical in each state. The same conditions are imposed on the functions r and p as in [20]. Moreover the function of total expected rewards $V(\pi,\rho)(x)$ is assumed to be uniformly bounded in π, ρ and x. It is shown that the game has a value, that the maximizing player has an ε-optimal stationary strategy and that the minimizing player has an optimal strategy.

[22] PARTHASARATHY, T.; *Discounted and positive stochastic games*. Bull. Amer. Math. Soc. 77 (1971), 134-136.

This note announces a few results on Markov games. Under the assumptions that the state space is a complete separable metric space, that the actions spaces are finite or compact metric (and some measurability conditions on the reward functions) three theorems are stated without proof about the value and optimal stationary strategies.

[23] PARTHASARATHY, T. and T.E.S. RAGHAVAN; *Some topics in two-person games*. American Elsevier, New York 1971, 238-252.

In chapter ten: "Stochastic games", the authors consider the Markov game with infinite state space and finite action spaces. It is shown that the discounted game with uniformly bounded payoffs $r(x,k,\ell)$ has a value and that optimal stationary strategies exist. They also consider the game with time average payoffs, using Gillette's [3] argument (see also Liggett and Lippman [16]).

[24] SOBEL, M.J.; *Noncooperative stochastic games*. Ann. Math. Statist. 42 (1971), 1930-1935.

In this paper the game with finite state and action spaces is played by N players each having his own payoff function $r_i(x,k_1,\ldots,k_N)$, $i = 1,\ldots,N$ (not necessarily $\sum_i r_i(x,k_1,\ldots,k_N) = 0$). Sobel shows that the discounted game has an equilibrium point and provides a sufficient condition for the game with time average payoffs to have an equilibrium point.

[25] BARON, S., D.L. KLEINMAN and S. SERBIN; *A study of the Markov game approach to tactical maneuvering problems*. NASA, Langley Research Center prepared by Bolt Beranek and Newman, inc. Cambridge Mass. nr. NASA CR-1979 (1972).

This report presents the results of a study to apply a Markov game approach to planar air combat problems. The Markov game has finite state and action spaces. Numerical results for a highly idealized version of the problem are presented.

[26] ORKIN, M.; *Recursive matrix games*. J. Appl. Prob. 9 (1972), 813-820.

Orkin considers the following game. Starting with a fixed matrix game (belonging to a finite set) player 1 chooses a row and P_2 chooses simultaneously a column. Then according to the transition probabilities either a new matrix game is selected while there is no payoff or the game is terminated with a final payoff. If the play goes on infinitely long payoff is defined 0. The author shows that this game has a value and that both players have ε-optimal stationary strategies.

[27] PARTHASARATHY, T.; *Discounted, positive and noncooperative stochastic games*. Int. J. of Game Theory 2 (1973), 25-37.

The author considers the discounted Markov game under the same conditions as in [20] but now allows the action spaces to depend on x, the positive game when state and action spaces (identical in each state) are finite and the nonzero-sum noncooperative discounted game when the state space is countable and the action spaces are finite. Moreover in the positive game (as in [21]) $V(\pi,\rho)(x)$ is assumed to be bounded and in the nonzero-sum game the functions r_1 and r_2 are assumed to be uniformly bounded. It is shown that these games have a value and that both players have stationary optimal strategies.

[28] RAO, S.S., R. CHANDRASEKARAN, and K.P.K. NAIR; *Algorithms for discounted stochastic games*. J. Opt. Theory Appl. 11 (1973), 627-637.

This paper considers the discounted game with finite state and action spaces. Two algorithms are given for this game: i) the Hoffman and Karp [7] algorithm which is shown to converge ii) the algorithm given by Pollatschek and Avi-Itzhak [17]. Rao et al. also tried to prove that the latter algorithm would always converge, however the proof they supplied is incorrect.

[29] SATIA, J.K. and R.E. LAVE; *Markovian decision processes with uncertain transition probabilities*. O.R. 21 (1973), 728-740.

This paper considers a discounted Markov decision process with finitely many states and actions when the transition probabilities are not known with certainty. One of the approaches given is a game theoretic one in which a Markov game is considered with finite state space, finite action space for P_1 and compact action space for P_2. This game is solved with the discounted version of the Hoffman and Karp [7] algorithm.

[30] SOBEL, M.L.; *Continuous stochastic games*. J. Appl. Prob. 10 (1973), 597-604.

In this paper Sobel considers discounted non-zero sum noncooperative Markov games when the sets of states, actions and players are given by metric spaces. The existence of an equilibrium point is proven under assumptions of continuity and compactness (see also [18] and [24]).

[31] VAN DER WAL, J.; *The method of successive approximations for the discounted Markov game*. Memorandum COSOR 75-02. Technological University Eindhoven 1975 (Department of Mathematics).

This paper presents a number of successive approximation algorithms for the discounted Markov game with finite state and action space. It is shown that each algorithm provides upper and lower bounds on the value of the game and nearly optimal strategies for both players.

[32] VAN DER WAL, J.; *Note on the optimal strategies for the finite-stage Markov game*. Memorandum COSOR 75-06. Technological University Eindhoven 1975 (Department of Mathematics).

This note considers the finite-stage Markov game when state and action spaces are all finite. Zachrisson [6] (silently) assumes that both players use only Markov strategies when he proves that the game has a value. Here a simple proof is given which shows this restriction to be irrelevant.

[33] VAN DER WAL, J.; *The solution of Markov games by successive approximation*. Master's thesis. Technological University Eindhoven, 1975 (Department of Mathematics).

This thesis deals with Markov games with finite state and action spaces. Besides the results in [31] and [32] it contains a class of algorithms for the discounted game to which Hoffmann and Karp's algorithm [7] belongs (see also [17] and [28]). Moreover for two games at the criterion of total expected rewards a successive approximation algorithm is given yielding upper and lower bounds. In the first game assumption A_1 of Kushner and Chamberlain [15] is satisfied, in the other game it is assumed that $r(x,k,\ell) > 0$ for all x, k and ℓ and that P_2 has in each state an action which terminates the game immediately (compare assumption A_2 in [15]).