БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ . BULGARIAN ACADEMY OF SCIENCES

КИБЕРНЕТИКА И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ · Том **4**, № 1
CYBERNETICS AND INFORMATION TECHNOLOGIES · Volume **4**, No 1

София . 2004 . Sofia

Characteristic Features of Markov Flow Games with Linear Constraints

Vassil Sgurev

Institute of Information Technologies, 1113 Sofia

Abstract: *The paper discusses a new class of the multi-step stochastic games – Markov flow games, consisting of a set of Markov decision making processes with different, in the general case not equal to one, own values and common linear constraints, and on the other hand – some models of the games theory.*

It is shown that the combining of the two mathematical structures leads to a new class of games – Markov flow games.

Some characteristic features of these games with constraints are presented and the possible ways of their future study.

Keywords: *multi-step stochastic games, Markov flows, theory of games*

Introduction

During the 40-ies of the last century the scientific direction “Markov decision making processes” (MDMP) [5] was founded and soon gained wide distribution.
At the end of the 60-ies of the same century, a specific class of Markov decision making processes with additional constraints [10] attracted the attention of the researchers, and was later generalized as Markov flows [12].

The present paper discusses a class of models, called Markov-flow games, in which the characteristic features of two scientific directions – Markov flows and theory of games, are combined. An attempt has been made to show the variety of these models and the directions for their study and application.

Discrete models will be considered only. Hence, the adjective “discrete” will be omitted in the text.

I. From Markov decision making processes to Markov flows

The topic of Markov processes focused the constant interest of a great number of researchers during XX century. Later on Acad. A. Kolmogorov proposed the use of the tools of measure theory [1] in defining stochastic values, which made possible the establishing of probability theory on fundamental mathematical grounds. The new possibilities were soon transferred to Markov chains and processes [2, 3].

Up to the beginning of the 60-ies of the last century, mainly homogenous Markov chains and processes with a fixed matrix of the transition probabilities were investigated. The Markov processes considered are with a finite and infinite number of states, with discrete and continuous time, homogenous and non-homogenous. Efficient applications of these processes in different areas of natural and technical sciences were found and they were widely used in the research and updating of different processes in biology, demography, sociology and other scientific directions [4].

At the end of the 50-ies of the last century, R. Howard defined for the first time the notion "Markov decision making process" (MDMP) (or "controllable Markov process" – in another interpretation [5].

In the classical Markov chain [2, 3] the square matrix of transition probabilities $\|p \times n\|_{n\times n}$ shows the probability distribution $\{p_{ij}\}$ of passing from any current state $i \in S$ into an arbitrary state $j \in S$, where S is the set of states, in which the Markov process can be found. At that, for every $i \in S$ and $j \in S$:

$$\sum_{j\in S} p_{ij} = 1;\ 0 \le p_{ij} \le 1.$$

The Markov decision making process (MDMP) is characterized by the following [5, 6]:

a) From each state $i \in S$ the decision maker (DM) can choose one of the possibilities $k \in K_i$, and the transition of the Markov process will be realized according to the distribution $\{p^k_{ij} \mid j \in S\}$, where K_i are the possible policies from state $i \in S$. Hence, differing from the classical Markov process, in which all transition probabilities $\{p_{ij} \mid i \in S, j \in S\}$ are fixed, in MDMP the word strategy will denote a sequence of policies $\Pi = \{h_n \mid n = 0, 1, ...\}$, where n is a discrete parameter of time.

b) For every transition from state i into state j of the Markov process, an income r_{ij} is obtained. If a policy $k \in K_i$ be selected, the profit expected will be equal to $r^k_i = \sum_{j\in S} p_{ij} r_{ij}$ The purpose is to maximize (minimize) it.

We shall consider further on MDMP with a finite number of states only.

For stationary MDMP with re-evaluation of the income β, R. Howard has developed an efficient interaction algorithm for β-optimal strategy. F. D?Epenoax [7] suggests the use of linear optimizing in determining the same strategy. For MDMP with one ergodic class and without re-evaluation of the profit, R. Howard proposes another interaction algorithm to define an optimal stationary strategy [5]. The same result may be obtained applying the method used by De Chellinck [8], based on linear optimizing.

MDMP with additional linear constraints was proposed in 1969 in [10]. The same paper notes that this class of MDMP has got new properties and that in MDMP with linear constraints the use of mixed (randomized) strategies leads to higher average incomes than pure strategies use. A similar result, but for semi-Markov decision making processes, is reported by Gertzbach.

In 1977, in [11] the control of several MDMP, called independent MDMP, on one and the same set of states and with common linear constraints, was suggested. It was proved that in this case the use of a set of mixed strategies leads to higher average profits than pure strategies use only.

In "Howard" MDMP it is assumed that, similarly to classical Markov chains, the stochastic process has not its own "values", or put in another way, it has a single value, which is not accounted during its running. Some problems arise in practice, when it is necessary to consider sets of simultaneously controlled MDMP with common linear constraints, in which each MDMP has its own value, not obligatory equal to one. This leads to a principally new approach for MDMP interpretation as a specific flow on a dynamic graph [12, 13]. A method is suggested in the same works for the presentation of the set of MDMP with common linear constraints and non-single own values as a network flow with linear capacity constraints (LCC-flow) and additional linear constraints (ALC), generally called a Markov flow [12]. The basis for such network-flow interpretation lies in the recurrent relations, according to which the probability to go to any state from n-th step is equal to the sum probability for leaving or remaining in this state [6]. Thanks to this a number of results for the network flows [13] can be referred to the class of Markov flows described. The latter can be regarded as a more generalized model than the different classes of "Howard" MDMP.

The Markov flows are rich stochastic structures and combined with other approaches of operational research they enable the receiving of new relevant results, useful in theoretic and applied aspect.

2. Game structures with linear constraints

The situations in which the interests of separate groups (sides) coincide partially or do not coincide, or prove to be opposite are of interest in real-life activity. The last case leads to a conflict situation. If the minor details are removed and only the essential part remains, a formal mathematic structure, that can be regarded as a game will be obtained.

In 1944 J. Von Neuman and O. Morgenstern realize the systematic analysis and research of the discipline, established as theory of games [14]. Later on it is applied in wide areas of economy, techniques, biology, etc. [15, 16, 17, 18].

Within the frames of this theory the interested sides (players) act according to apriori known rules with the help of specific strategies. In a conflict, every player chooses their strategy, as a result of this a sequence of strategies is formed, called usual situation. The interests of the players in selecting a situation concern the acquiring of an income, called profit of the player in this situation.

The analysis of the games is realized from different points of view, namely: development of optimality principles through determining a reasonable behaviour of the players, analysis of the possibility for realization of these optimality principles, defining an optimal behaviour and searching for an equilibrium situation.

Very often when pure strategies only are applied, no equilibrium situation between the two players can be reached. In this case there is a possibility for stochastic assortment of the strategies, each one of them selected with a definite probability. Thus mixed (randomized) strategies are accomplished, in which there is always an equilibrium situation, considered optimal for both players.

The selection of the separate strategies in games theory likewise in MDMP may be realized in the presence of additional linear constraints also.

The most broadly spread and studied class of games is that of the matrix antagonistic games [14, 16]. Another important class of games, describing conflicts are the positional games. Their dynamics influences the behaviour of the separate players. The positional games are non-coalitional. They model the processes of successive decision making by the players at incomplete and time altering information. Sequential transition is done from one game state to another either by rational selection of one of the possible strategies – satisfying the established game rules, or by stochastic selection.

The third class of well-studied games are the bi-matrix games, in which the interests of the players do not always coincide, but still they are not antagonistic. Every player in them, unlike in matrix games, has a personal payment matrix and strives to maximize the profit [19]. Given that the interests of the players do not coincide, it is necessary to find a compromise solution, which will satisfy both players to an extent. It is proved that this can be achieved on the basis of an equilibrium situation, using mixed strategies. This does not lead to antagonism of the interests, as in matrix games, but to behavioural antagonism.

3. Some classes of Markov flow games with linear constraints

The application of game approaches to Markov decision making processes is offered at the beginning of 70-ies of XX century [20], after obtaining the most important results in MDMP. Many publications are available nowadays in that field – a good idea for this can be found in [21, 22, 23, 24]. Some cases of Markov games in Boreal spaces are discussed there.

The game situation is usually defined for two players, every one of them having the possibility to select one of the policies possible for him, after that the controllable Markov process goes to a new state, conveying incomes to the players in accordance with their payment matrices. In case the profit (income) of the first one is a loss for the second, an antagonistic Markov game with a zero sum is realized. Very frequently the first player is active, while the second one personalizes the enemy nature, acting most unfavourably against the first one. Games with a set of players are also defined, every one receiving their own income when passing from one state to another.

It is evident, that the Markov games thus constructed are multi-step positional games. They can be antagonistic or non-antagonistic, with re-evaluation or without re-evaluation of the income, with or without constraints on the profits gained, continuous or discrete.

The Markov flow described in the previous chapter is a specific structure, which is not enclosed in the examples above described. It is possible to simultaneously control not one, but several MDMP with non-single own values, on the transition probabilities of which common additional linear equalities and inequalities are implied.

The definition of the following classes of Markov-flow games (MFG) is possible:

1. At a single Markov flow an antagonistic Markov-flow game with a single value of the flow and with two players with opposite interests can be determined – one of them attempts to maximize the income expected, while the other one – to minimize it. The

controllable Markov flow has fixed transition probabilities and a set of possible policies. Each one of the players influences the process selecting his mixed strategies from a state $i \in S$, namely $\{q^k_1(i)\}$ for the first one and $\{q^k_2(i)\}$ – for the second. After such selection of mixed strategies the Markov process will pass from state *i* to *j* at *n*-th step with a probability

$$P_{ij}(n) = \sum_{k \in K_i} ((q^k_1(i)\, p^k_{ij})\, m_1 + (q^k_2(i)\, p^k_{ij})\, m_2),$$

where $0 \le m_1 \le 1$; $0 \le m_2 \le 1$, $m_1 + m_2 = 1$.

The parameters m_1 and m_2 reflect the influence of the mixed strategy selected by the first and by the second player respectively. The income expected from *n*-th step for a discrete time interval for the strategies selected is equal to $r_i(n) = \sum_{j \in S} p_{ij}(n)\, r_{ij}$, where r_{ij} is the income, obtained at the process transition from state *i* to *j*.

When selecting a mixed strategy it is necessary to satisfy the typical for Markov flow common linear constraints (equalities and inequalities) on the transition probabilities, and also the possible constraints on the income gained.

A separate payment matrix can be constructed for each state for the income expected at one step, its rows and columns corresponding to the pure strategies that can be chosen by every player. When mixed strategies are selected, an equilibrium status for one step of the process is attained.

The situation is still more complicated than in classical matrix games, since the selection of mixed strategies at each step must be realized depending on the complete process, i.e. it is impossible to make an optimal choice for one forward step only. Selection of mixed strategies is necessary, accounting the possible behaviour of the Markov process till its completing – if it has a fixed end, or of the optimal stationary mixed strategy otherwise.

It is obvious that the definition of the optimal mixed strategies in MFG is a more complex procedure than their computation in non-game Markov flows or in matrix games.

The essence of the model discussed is not altered by the assumption that the second player is of enemy disposition, acting against the first one in the most unfavourable way.

2. The previous model of MFG can be modified, if assumed that every one of the two players gains different incomes in the process transition from one state to another. Then a non-antagonistic game with bi-matrix game elements is expected.

In this case together with behavioural antagonism, there also appear supplementary contradictions, since the two players control one and the same stochastic process with common linear inequalities on the transition probabilities, which limits their possibilities for choice of mixed strategies.

3. A MFG is possible when two players control two separate Markov flows connected in-between by common linear constraints on the transition probabilities. The expression $0 \le m_1^1 \le 1$; $0 \le m_2^2 \le 1$, then will denote the influence of the strategies being selected by the first player in controlling the first and the second Markov flow respectively. For the second game the analogous values are $1 - m_1^1$ and $1 - m_2^2$ respectively. The final values 1 and 0 correspond to the case, when either the first player

completely controls the first Markov flow, or he does not participate in this control. Some cases are possible when one of the players does not participate in the control of the first Markov flow, but the two players simultaneously control the second flow.

The two controllable Markov flows may have either equal policies and transition probabilities connected with them, or different ones. They could possess either one and the same incomes for a transition from one state to another, or different for the two players. In the first case an antagonistic game is played, in the second one – non-antagonistic game with behavioural antagonism and contradictions with respect to the common linear constraints on the transition probabilities. Another case may also be considered when the two players have antagonistic interests in the control of the first flow and non-antagonistic (but behaviouraly antagonistic) in the control of the second flow, keeping the common linear constraints.

4. The following generalized models are for a set of mutually connected Markov flows *V*, controlled by several players *T*, where the number of the flows and players are arbitrary finite numbers. In this case the influence of every player $t \in T$ on each flow $s \in V$ is expressed by the values $0 \leq m_t^s \leq 1$, for which

$$\sum_{t \in T} m_t^s = 1; \text{ for each } s \in V.$$

All possible modifications of MFG, discussed in previous models of Ch. 1 up to 3 can be transferred to the cases when the players and flows are more than two. There is a new possibility to construct coalitional MFG, when among the different coalitions of players there exist antagonistic and non-antagonistic interests. The players may be chasing common interests inside the coalitions (for example common coalition maximization of the income), observing their own interests to certain limits.

5. Still greater variety of MFG can be achieved in the cases when personal (as a rule different than 1) values of the controllable Markov processes are considered. All the models described in Ch. 1 up to 4 could be generalized in a different way and with greater diversity for the cases when the Markov processes have non-single own values.

6. For every model of MFG with more than one controllable Markov flow (or with one flow, but with altering own value), the behaviour of the flows can be investigated parallel to the incomes growth realized by them with the increase of their number or of their own values. In this case due to the requirement for common linear constraints on the transition probabilities, saturation of these constraints will be reached and realization of the capacity values for the corresponding cuts – notions, which can be introduced to MFG similarly to the usual Markov flows [12]. It is almost sure that in the case discussed an analog can be found with the famous mincut-maxflow theorem of L. Ford and D. Fullkerson, which in its game interpretation will signify that the maximal number of the dependent Markov flows participating in the game (or of the maximal sum of their own values) in MFG, is equal to the minimal value of the capacity on any of the possible cuts.

The introduction of the notions cut and capacity function in the theory of games enables the study of these games with the help of network flows tools.

7. All models of MFG from Ch. 3 up to Ch. 5 may be adapted also for the cases when the separate Markov flows are independent, i.e., when there are not any common linear constraints on their transition probabilities. Then the number of the controllable

Markov flows, or the sum of their own values is not above bounded and there do not exist capacity functions on the separate cuts.

The models above described do not exhaust the entire variety of MFG. In general it is greater than the different classes of Markov flows and of games types.

The complexity of the different MFG, especially of those with a set of flows and games with non-single own values and with common linear constraints on the transition probabilities is quite great.

In order to define the optimal mixed strategies, it is necessary to develop a set of appropriate optimization procedures with acceptable computing complexity. More efficient results can be obtained in some of the most complicated MFG, using approximate methods for quasi-optimal decision making.

It is difficult to describe the whole sphere of practical problems, which can get adequate description and efficient solution applying the described class of multi-step stochastic processes – the Markov flow games. These applications must be searched for in broad areas of production and services and in social-economic practice. In order to implement them, it is necessary to study in detail the whole spectrum of MFG in advance.

References

1. H a l m o s, P. Measure Theory. Princeton, Van Nostrant Co., 1950.
2. K e m e n y, J., J. S n e l l. Finite Markov Chains. Princeton. Van Nostrant Co., 1960.
3. C h u n g, K. L. Markov Chains with Stationary Transition Probabilities. Berlin, Springer-Verlag, 1960.
4. F e l l e r, W. An Introduction to Probability Theory and Its Applications. N. Y., J. Wiley and Sons, 1957.
5. H o w a r d, R. Dynamic Programming and Makov Processes. Technology Press MIT and J. Wiley and Sons, Inc., 1960.
6. M i n e, H., S. O s a k i. Markovian Decision Processes. N.Y., American Elsevier Publishing Com., Inc., 1970.
7. D?E p e n o a x, F. Sur un probleme de production et de stockage dans l?aetoir. – Rev. Franc Rech. Operationelle, **14**, 1960, 3-16.
8. D e G h e l l i n c k, G. Le probleme de decision sequentielles. – Cahiers Centre Etude Rech. Operationelle, R., 1960, 161-179.
9. B e l l m a n, R. Dynamic Programming. Princeton Univ. Press, 1962.
10. S g u r e v, V. Optimal control of Markov processes with additional constraints. – Technicheska misul, Sofia, 1970, No 3 (in Bulgarian).
11. S g u r e v, V. Control of a finite number of Markov processes with additional constraints. – Preprints of Symposium "Applied problems of large control systems". BAS, 1980 (in Russian).
12. S g u r e v, V. Markov Flows. Sofia, BAS, 1993 (in Russian).
13. S g u r e v, V. Markov Flows with Common Constraints. Sofia, BAS, 1991 (in Russian).
14. V o n N e u m a n n, J., O. M o r g e n s t e r n. Theory of Games and Economic Behaviour. Princeton, N. J., Princeton Univ. Press, 1953.
15. W i l l i a m s, J. The Complete Strategist. N. Y., J. Wiley and Sons, 1954.
16. L u c e, R., H. R a i f f a. Games and Decisions. N. Y., J. Wiley and Sons, 1957.
17. H e r m e e r, J. B. Games with Non-contradictive Interests. M., Nauka, 1976 (in Russian).
18. V o r o b i o v, N. Bases of Games Theory. Non-coalitional games. M., Nauka, 1984 (in Russian).
19. S h i k i n, E. From games to games. M., URSS, 2003 (in Russian).
20. M a i t r a, A., P a r t h a s a r a n t h y. On Stochastic Games. – J. Optim. Theory Appl., **5**, 1970, 289-300.
21. W e s s e l s. Markov Games with Unbounded Rewards. – Bonner Mathematische Schriftenn, No 98, Bonn, 1977.

22. A l t m a n, E., A. H o d j i k, F. S p i e k s m a. Constraint conditions for average and discounted optimality in countable state Markov games with unbounded Rewards. – Math. Oper. Res., **22**, 1997, 588-618.
23. L a l, A., S. S i n h a. Zero-sum two person semi-Markov games. – J. Appl. Prob., **29**, 1992, 56-72.
24. A l t m a n, E., G. K o o l. Stochastic Scheduling Games with Markov Decsison Arrival Processes. – J. Computers and Mathematics with Appl., **26**, 1993, No 6, 141-148.

Особености на Марковските потокови игри с линейни ограничения

Васил Сгурев

Институт по информационни технологии, 1113 София

(Р е з ю м е)

В работата се разглежда един нов клас на многостъпковите стохастични игри – Марковски потокови игри с линейни ограничения и доходи.

В тяхната основа лежи, от една страна, понятието за Марковски поток, състоящо се от множество Марковски процеси за вземане на решение с различни, в общия случай неравни на единица собствени стойности и общи линейни ограничения, а от друга – модели на теорията на игрите.

Показано е, че съчетанието на двете математически структури води до нов клас игри – Марковски потокови игри.

Дадени са някои особености на тези игри с ограничения и възможните начини на тяхното бъдещо изследване.