Canonical forms of two-person zero-sum limit average payoff stochastic games

Research in Teams – 11rit176

Endre Boros[*] Khaled Elbassioni[†] Vladimir Gurvich[‡]
Kazuhisa Makino[§]

March 7, 2012

[*] RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003; (boros@rutcor.rutgers.edu)

[†] Max-Planck-Institute for Informatics; Campus E 1 4, 66123, Saarbruecken, Germany (elbassio@mpi-sb.mpg.de)

[‡] RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003; (gurvich@rutcor.rutgers.edu)

[§] Graduate School of Information Science and Technology, University of Tokyo, Tokyo, 113-8656, Japan; (makino@mist.i.u-tokyo.ac.jp)

Abstract

We consider two-person zero-sum stochastic games with perfect information and, for each $k \in \mathbb{Z}_+$, introduce a new payoff function, called the k-total reward. For $k = 0$ and 1 they are the so called mean and total rewards, respectively. For all k, we prove solvability of the considered games in pure stationary strategies, and show that the uniformly optimal strategies for the discounted mean payoff (discounted 0-reward) function are also uniformly optimal for k-total rewards if the discount factor is close enough (depending on k) to 1. We also demonstrate that the k-total reward games form a proper subset of the $(k+1)$-total reward games for each k. In particular, all these classes contain mean-payoff games. This observation implies that, in the non-zero-sum case, Nash-solvability fails for all k.

Keywords: stochastic game with perfect information, two person, zero sum, mean payoff, total payoff

1 Introduction

We consider two person zero sum stochastic games with perfect information and, for each positive integer k we define an effective payoff function, called the k-total reward, generalizing the classical *mean payoff* s [3] ($k = 0$), as well as *total rewards* [15, 16] ($k = 1$).

In this paper, we restrict ourselves by two person zero sum games, and the solution concept is Nash equilibrium, which is just a saddle point in the considered case.

We also restrict ourselves (and the players) to pure stationary strategies and deterministic states. Respectively, we call the considered family of games *k-total reward BW-games*, where B and W stand for the two players, BLACK the minimizer and WHITE the maximizer.

We denote by $\mathbb{R}$ the set of reals, by $\mathbb{Z}$ the set of integers, and by $\mathbb{Z}_+$ the set of nonnegative integers. For a subset $S \subseteq \mathbb{Z}_+$, let $\mathbb{R}^S$ denote the set of vectors indexed by the elements of S. We define $\mathcal{S} = \mathcal{S}(R) = \mathbb{R}^{\mathbb{Z}_+ \setminus \{0\}}$ as the set of infinite integral sequences with elements not larger in absolute value than R, and for $\mathbf{a} \in \mathcal{S}$ we write $\mathbf{a} = (a_1, a_2, ...)$. For $n \in \mathbb{Z}_+ \setminus \{0\}$ we define $[n] = \{1, 2, ..., n\}$ and write simply $\mathbb{R}^n$ instead of $\mathbb{R}^{[n]}$.

To describe BW-games, let us consider a directed graph (digraph) $G = (V, E)$, whose vertices (positions or states) are partitioned into two sets $V = B \cup W$, a fixed initial state $v_0 \in V$, a real-valued function $r : E \to \mathbb{Z}$ assigning an integer weights to the arcs (moves), and a mapping $\pi : \mathcal{S} \to \mathbb{R}$. We call the tuple (G, B, W) a *BW game form* and r its *local rewards*, while the tuple (G, B, W, r, π) is called a *BW game* and π is its *effective reward* (or *payoff function*). Two players, BLACK (the minimizer) and WHITE (the maximizer) control the positions of B and W, respectively. The game is played by starting at time $t = 0$ in the initial node $s_0 = v_0$. In a general step, in time t, we are at node $s_t \in V$. The player who controls s_t chooses an outgoing arc $e_{t+1} = (s_t, v) \in E$, and the game moves to node $s_{t+1} = v$. We assume, in fact without any loss of generality, that every vertex in G has an outgoing arc. (Indeed, if not, one can add loops to the corresponding vertices). We assume that an initial vertex v_0 is fixed. However, when we talk about solving a BW-game, we consider (separately) all possible initial vertices.

In the course of this game players generate an infinite sequence of edges $\mathbf{p} = (e_1, e_2, ...)$ (a *play*) and the corresponding real sequence $r(\mathbf{p}) = (r(e_1), r(e_2), ...) \in \mathcal{S}$ of local rewards. At the end (after infinitely many steps) BLACK pays WHITE $\pi(r(\mathbf{p}))$ amount. Naturally, WHITE's aim is to create a play which maximizes this payoff, while BLACK tries to minimize it. (Let us note that the local reward function $r : E \to \mathbb{R}$ may have negative values, and $\pi(r(\mathbf{p}))$ maybe negative too, in which case WHITE has to pay BLACK.)

In this paper, we restrict ourselves, and the players, to stationary strategies. It means that each player chooses, in advance, a move in every position that he/she controls and makes this move whenever the play comes to this position. Then, the play is uniquely determined by the one time selection of arcs and by the initial position. Such a play always looks like a "lasso": it consists of an initial path entering a directed cycle, which is then repeated infinitely many times.

Thus, the effective reward is defined for very specially structured (quasi-periodical) sequences, which have only finitely many different local values (in fact, no more then $n = |V|$), and in which we repeat cyclically, infinitely many times, the values following an initial segment.

Given two sequences, $\mathbf{x} \in \mathbb{Z}^p$ and $\mathbf{y} \in \mathbb{Z}^q$, let us denote by $\mathbf{a} = (\mathbf{x}(\mathbf{y}))$ the infinite sequence obtained by listing first the elements of $\mathbf{x}$ and then repeating $\mathbf{y}$ cyclically, infinitely many times. Let us call such an $\mathbf{a} \in \mathcal{S}$ a *lasso sequence*, and let us denote the set of lasso sequences that can arise from a graph on n vertices by

$$\mathcal{S}_n(R) \;=\; \left\{ \mathbf{a} = (\mathbf{x}(\mathbf{y})) \;\middle|\; \begin{array}{cc} p, q \in \mathbb{Z}_+, & p + q \leq n \\ \mathbf{x} \in [-R, R]^p, & \mathbf{y} \in [-R, R]^q \end{array} \right\},$$

where $[-R, R]$ denotes the set of integers of absolute value not exceeding R. Note that a BW-game with n states in stationary strategies always produces a play $\mathbf{p}$ such that the corresponding rewards sequence $r(\mathbf{p})$ belongs to $\mathcal{S}_n(R)$, where R is an upper bound on the absolute value of the integral arc rewards. We shall simply write $\mathcal{S}_n$ when R is not specified.

Mean payoff (undiscounted) stochastic games, introduced in [3], are BW games (see also [12, 11, 2, 5]) with payoff function $\pi = \phi$:

$$\phi(\mathbf{a}) \;=\; \liminf_{T \to \infty} \frac{1}{T} \sum_{j=1}^{T} a_j. \tag{1}$$

This family of games is known to have a value (Nash equilibrium) in pure stationary strategies [3, 9] and this value and the corresponding optimal stationary strategies can be computed in pseudo-polynomial time [18]. Let us remark that the problem of deciding if the value of a mean payoff BW-game is below (or above) a given threshold belongs to both NP and co-NP ([5, 8, 18]). The exact complexity of this problem is however still not known; the best known algorithms are either pseudo-polynomial [18] or subexponential [1, 6, 17].

Discounted mean payoff stochastic games were in fact introduced earlier in [14] and have payoff function $\pi = \phi_\beta$:

$$\phi_\beta(\mathbf{a}) \;=\; (1-\beta)\sum_{j=1}^{\infty} \beta^{j-1} a_j. \tag{2}$$

As a consequence of the classical Hardy-Littlewood tauberian theorems [7] we have the equality

$$\phi(\mathbf{a}) \;=\; \lim_{\beta\to 1} \phi_\beta(\mathbf{a}). \tag{3}$$

Discounted games, in general, are easier to solve, due to the fact that a standard value iteration is in fact a fast converging contraction. Hence, they are widely used in the literature of stochastic games together with the above limit equality. In fact, for BW-games it is known [18] that for two sequences $\mathbf{a}, \mathbf{b} \in \mathcal{S}(R)$ we have $\phi_\beta(\mathbf{a}) < \phi_\beta(\mathbf{b})$ if and only if $\phi(\mathbf{a}) < \phi(\mathbf{b})$ whenever $1-\beta \leq \frac{1}{4n^3R}$.

Total reward, introduced in [15] and considered in more detail in [16], is defined by

$$\psi(\mathbf{a}) \;=\; \liminf_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T}\sum_{j=1}^{i} a_j. \tag{4}$$

It was shown in [16] that a total reward game is equivalent with a mean payoff game having countably many states. The authors derive from this that total reward games have values. Furthermore, ϵ-optimal stationary strategies can be constructed by solving a discounted mean payoff game on the same graph with the same rewards using a discount factor β close enough to 1. The proof of the latter is claimed in [16] to be analogous to the proof in [10].

2 New Results

We extend and generalize the above results by introducing a complete hierarchy of payoff functions.

For every $\ell \in \mathbb{Z}_+$, we define the ℓ-total effective reward, which coincide with the mean payoff when $\ell = 0$ and with the total reward when $\ell = 1$.

First, we show that ℓ-total reward games have uniformly optimal stationary strategy for every ℓ.

We also prove that ℓ-total reward games form a proper subset of $(\ell{+}1)$-total reward games for each $\ell \in \mathbb{Z}_+$. In particular, mean payoff games form a proper subset of ℓ-total reward games for all $\ell \in \mathbb{Z}_+$. This containment and the example of [4] prove that for each $\ell \in \mathbb{Z}_+$, there is a (non-zero sum) ℓ-total reward game without a Nash equilibrium.

We show that the optimal stationary strategy for an ℓ-total reward game can be computed by solving a discounted mean payoff game with a discount factor close enough to 1. This provides us with a pseudo-polynomial algorithm to determine the Nash equilibria of ℓ-total reward games, whenever ℓ is fixed. We must add that the complexity depends exponentially on ℓ, in the worst case.

We also prove that 1-total reward games with nonnegative arc rewards are polynomially solvable. This contrasts the fact that mean payoff games with nonnegative rewards are as hard as general mean payoff games, and that the fastest known algorithms for mean payoff BW-games are either pseudo-polynomial [5, 13, 18] or subexponential [1, 6, 17].

Let us finally note that the 1-total reward for a play terminating in a zero-reward loop is just the sum of the local rewards (the cost or length) of this play, and WHITE is a maximizer, while BLACK is minimizer of these costs. In contrast, if a play ends with a cycle then the corresponding 1-total reward is $+\infty$ (respectively, $-\infty$) whenever the sum of the local rewards along this cycle is positive (respectively, negative), and the 1-total reward takes a finite value if and only if the above sum is 0.

References

[1] H. Björklund and S. Vorobyov. Combinatorial structure and randomized subexponential algorithms for infinite games. *Theoretical Computer Science*, 349(3):347–360, 2005.

[2] A. Ehrenfeucht and J. Mycielski. Positional strategies for mean payoff games. *International Journal of Game Theory*, 8:109–113, 1979.

[3] D. Gillette. Stochastic games with zero stop probabilities. In *Contributions to the Theory of Games, Vol. III*, volume 39 of *Annals of Mathematics Studies*, pages 179–187. Princeton, N.J., 1957.

[4] V. Gurvich. A stochastic game with complete information and without equilibrium situations in pure stationary strategies. *Russian Mathematical Surveys*, 43(2):171–172, 1988.

[5] V.A. Gurvich, A.V. Karzanov, and L.G. Khachiyan. Cyclic games and an algorithm to find minimax cycle means in directed graphs. *USSR Comput. Math. Math. Phys.*, 28:8591, 1988.

[6] N. Halman. Simple stochastic games, parity games, mean payoff games and discounted payoff games are all LP-type problems. *Algorithmica*, 49(1):37–50, 2007.

[7] G.H. Hardy and J.E. Littlewood. Notes on the theory of series (xvi): two tauberian theorems. *J. of London Mathematical Society*, 6:281–286, 1931.

[8] A.V. Karzanov and V.N. Lebedev. Cyclical games with prohibition. *Mathematical Programming*, 60:277–293, 1993.

[9] Thomas M. Liggett and Steven A. Lippman. Stochastic games with perfect information and time average payoff. *SIAM Review*, 11(4):604–607, 1969.

[10] J. F. Mertens and A. Neyman. Stochastic games. *International Journal of Game Theory*, 10:53–66, 1981.

[11] H. Moulin. Extension of two person zero sum games. *Journal of Mathematical Analysis and Aplication*, 55:2:490–507, 1976.

[12] H. Moulin. Prolongement des jeux à deux joueurs de somme nulle. une théorie abstraite des duels. *Mémoires de la Soc. Math. France*, 45:5–111, 1976.

[13] N.N. Pisaruk. Mean cost cyclical games. *Mathematics of Operations Research*, 24:4:817–828, 1999.

[14] L. Shapley. Stochastic games. *Proc. Nat. Acad. Sci. USA*, 39:1095–1100, 1953.

[15] F. Thuijsman and O. J. Vrieze. The bad match, a total reward stochastic game. *Operations Research Spektrum*, 9:93–99, 1987.

[16] F. Thuijsman and O. J. Vrieze. Total reward stochastic games and sensitive average reward strategies. *Journal of Optimization Theory and Applications*, 98:175–196, 1998.

[17] S. Vorobyov. Cyclic games and linear programming. *Discrete Applied Mathematics*, 156(11):2195–2231, 2008.

[18] Uri Zwick and Mike Paterson. The complexity of mean payoff games on graphs. *Theoretical Computer Science*, 158(1-2):343 – 359, 1996.