J. Math. Anal. Appl. 457 (2018) 1038–1064

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Generic uniqueness of the bias vector of finite zero-sum stochastic games with perfect information ☆

Marianne Akian [a], Stéphane Gaubert [a,*], Antoine Hochart [b]

[a] CMAP, Inria, Ecole polytechnique, CNRS, Université Paris-Saclay, 91128 Palaiseau, France
[b] Toulouse School of Economics, Université Toulouse I, 31015 Toulouse, France

ARTICLE INFO

Article history:
Received 1 October 2016
Available online 24 July 2017
Submitted by H. Frankowska

Keywords:
Zero-sum games
Ergodic control
Nonexpansive mappings
Fixed point sets
Policy iteration

ABSTRACT

Mean-payoff zero-sum stochastic games can be studied by means of a nonlinear spectral problem. When the state space is finite, the latter consists in finding an eigenpair (u, λ) solution of $T(u) = \lambda e + u$, where $T : \mathbb{R}^n \to \mathbb{R}^n$ is the Shapley (or dynamic programming) operator, λ is a scalar, e is the unit vector, and $u \in \mathbb{R}^n$. The scalar λ yields the mean payoff per time unit, and the vector u, called *bias*, allows one to determine optimal stationary strategies in the mean-payoff game. The existence of the eigenpair (u, λ) is generally related to ergodicity conditions. A basic issue is to understand for which classes of games the bias vector is unique (up to an additive constant). In this paper, we consider perfect-information zero-sum stochastic games with finite state and action spaces, thinking of the transition payments as variable parameters, transition probabilities being fixed. We show that the bias vector, thought of as a function of the transition payments, is generically unique (up to an additive constant). The proof uses techniques of nonlinear Perron–Frobenius theory. As an application of our results, we obtain an explicit perturbation scheme allowing one to solve degenerate instances of stochastic games by policy iteration.

1. Introduction

1.1. The ergodic equation for stochastic games

Repeated zero-sum games describe long-term interactions between two agents, called players, with opposite interests. In this paper, we consider *perfect-information zero-sum stochastic games*, in which the players

☆ This work was performed when A. Hochart was with Inria and CMAP, Ecole polytechnique, CNRS, Université Paris-Saclay. The authors are also partially supported by the PGMO program of EDF and FMJH and by the ANR (MALTHY Project, number ANR-13-INSE-0003).

* Corresponding author.
E-mail addresses: marianne.akian@inria.fr (M. Akian), stephane.gaubert@inria.fr (S. Gaubert), antoine.hochart@polytechnique.edu (A. Hochart).

http://dx.doi.org/10.1016/j.jmaa.2017.07.017
0022-247X/

choose repeatedly and alternatively an action, being informed of all the events that have previously occurred (state of nature and chosen actions). These choices determine at each stage of the game a payment, as well as the next state by a stochastic process. Given a finite horizon k and an initial state i, one player intends to minimize the sum of the payments of the k first stages, while the other player intends to maximize it. This gives rise to the value of the k-stage game, denoted by v_i^k.

A major topic in the theory of zero-sum stochastic games is the asymptotic behavior of the mean values per time unit (v_i^k/k) as the horizon k tends to infinity. The limit, when it exists, is referred to as the *mean payoff*. This question was first addressed in the case of a finite state space by Everett [27], Kohlberg [36], and Bewley and Kohlberg [14]. See also Rosenberg and Sorin [50], Sorin [53], Renault [47], Bolte, Gaubert and Vigeral [15] and Ziliotto [56] for more recent developments. We refer the reader to [44] for more background on stochastic games.

A way to study the asymptotic behavior of the values $(v_i^k)_k$ consists in exploiting the recursive structure of the game. This structure is encompassed in the *dynamic programming* or *Shapley operator* of the game. In this paper, since the state space is assumed to be finite, say $\{1,\dots,n\}$, the latter is a map $T:\mathbb{R}^n\to\mathbb{R}^n$. Then, a basic tool to study the asymptotic properties of the sequence $(v_i^k)_k$ is the following nonlinear spectral problem, called the *ergodic equation*:

$$T(u)=\lambda e+u\ , \tag{1.1}$$

where e is the unit vector of $\mathbb{R}^n$. Indeed, if there exist a vector $u\in\mathbb{R}^n$ and a scalar $\lambda\in\mathbb{R}$ solution of (1.1), then, not only the sequence $(v_i^k/k)_k$ converges as the horizon k tends to infinity, but also the limit is independent of the initial state i, and equal to λ. This scalar, which is unique, is called the *ergodic constant* or the *(additive) eigenvalue* of T, and the vector u, called *bias vector* or *(additive) eigenvector*, gives optimal stationary strategies in the mean-payoff game.

A first question is to understand when the ergodic equation has a solution. In [6], we considered the Shapley operator $T:\mathbb{R}^n\to\mathbb{R}^n$ of a game with finite state space and bounded transition payment function, and gave necessary and sufficient conditions under which the ergodic equation is solvable for all the operators $g+T$ with $g\in\mathbb{R}^n$, or equivalently for all perturbations g of the payments that only depend on the state. Moreover, assuming the compactness of the action sets of players and some continuity of the transition functions, these conditions can be characterized in terms of reachability in directed hypergraphs.

A second question concerns the structure of the set of bias vectors. For one-player problems, i.e., for discrete optimal control, the ergodic equation (1.1), also known as the *average case optimality equation*, has been much studied, either in the deterministic or in the stochastic case (Markov decision problems). Then, the representation of bias vectors and their relation with optimal strategies is well understood.

Indeed, in the deterministic case, the analysis of the ergodic equation relies on max-plus spectral theory, which goes back to the work of Romanovsky [49], Gondran and Minoux [32] and Cuninghame-Green [23]. Kontorer and Yakovenko [39] deal specially with infinite horizon optimization and mean-payoff problems. We refer the reader to the monographies [11,38,18] or surveys [12,1] for more background on max-plus spectral theory. One of the main result of this theory shows that the set of bias vectors has the structure of a max-plus (tropical) cone, i.e., that it is invariant by max-plus linear combinations, and it has a unique minimal generating family consisting of certain "extreme" generators, which can be identified by looking at the support of the maximizing measures in the linear programming formulation of the optimal control problem, or at the "recurrence points" of infinite optimal trajectories. A geometric approach to some of these results, in terms of polyhedral fans, has been recently given by Sturmfels and Tran [54].

The eigenproblem (1.1) has been studied in the more general infinite-dimensional state space case, see Kolokoltsov and Maslov [38], Mallet-Paret and Nussbaum [42], and Akian, Gaubert and Walsh [7] for an