NONLINEAR SPECTRAL RADII OF ORDER-PRESERVING MAPS AND INFINITE HORIZON ZERO-SUM TWO-PLAYER STOCHASTIC GAMES

M. AKIAN*, S. GAUBERT†, AND R. NUSSBAUM‡

Key words. Nonlinear Perron-Frobenius theory, zeros-sum two player games, dynamic programming operators, Value iteration.

AMS subject classifications. 37N35, 49L20, 91A05

EXTENDED ABSTRACT. Order preserving, additively (sub)-homogeneous self-maps of R^n can be thought of as Shapley or dynamic programming operators of (discounted) zero-sum two-player stochastic games with n states and discrete time. Hence the fixed point or the additive eigenvalue of such maps are expected to coincide with the value function of some infinite horizon or mean-payoff zero-sum two-player stochastic game. We give several sufficient conditions under which the latter properties hold true, and obtain in that case an explicit formula for the rate of convergence of value iteration. These properties are based on Collatz-Wielandt type characterizations of nonlinear spectral radii of order preserving maps. The present results extend in particular a result by Bertsekas and Tsitsiklis (91) concerning stochastic control problems with an absorbing state.

1. Order-preserving maps as dynamic programming operators of zero-sum two-player stochastic games. The Shapley or dynamic programming operator of a discounted zero-sum two-player stochastic game with n states and discrete time is a map $f : \mathbb{R}^n \to \mathbb{R}^n$ of the form:

$$f(v)_i := \max_{a\in\mathcal{A}} \left(\min_{b\in\mathcal{B}} \left(\sum_{j=1}^{n} P_{ij}^{(a,b)}\, v_j + r_i^{(ab)} \right)\right) \qquad \forall i \in [1:n], \tag{1.1}$$

where $[1:n] := \{1,\dots,n\}$ is the set of states, $\mathcal{A}$ and $\mathcal{B}$ are the sets of actions of the two players, and for all $i,j \in [1:n]$, $a \in \mathcal{A}$, $b \in \mathcal{B}$, $r_i^{(ab)} \in \mathbb{R}$ and $P_{ij}^{(a,b)} \geq 0$ with $\sum_{j=1}^n P_{ij}^{(a,b)} \leq 1$. Here $r_i^{(ab)}$ is the reward of the first player, which is also the opposite of the reward of the second player, at each time when the game is in state i, and $P_{ij}^{(a,b)} \geq 0$ is the transition probability from state i to state j, multiplied by some discount factor (which is ≤ 1) depending on state i; $r_i^{(ab)}$ and $P_{ij}^{(a,b)} \geq 0$ depend on the action a of the first player and the action b of the second player at the current time. The aim of each player is to maximize his own objective function, which is for instance the sum of the rewards he received at all steps, by choosing at each step his action as a function of past states and actions of the game. If we assume that the first player plays before the second player, that the action spaces are finite sets, and that the discount factor are uniformely bounded from above by a constant $\mu < 1$, then the

*M. Akian, INRIA Saclay & Centre de Mathématiques Appliquées (CMAP), École Polytechnique, 91128 Palaiseau, France (marianne.akian@inria.fr).

†S. Gaubert, INRIA Saclay & Centre de Mathématiques Appliquées (CMAP), École Polytechnique, 91128 Palaiseau, France (stephane.gaubert@inria.fr).

‡Roger Nussbaum, Mathematics Department, Hill Center, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey, U.S.A. 08854-8019 (nussbaum@math.rutgers.edu).

value of the game exists and is the unique fixed point of the map f. Moreover, the value iteration of this game coincides with the fixed point iteration associated to f.

Fixed points equations of maps f of the form (1.1) can also be obtained as discretizations of Isaacs partial differential equations associated to differential stochastic games, when using a monotone scheme.

Under the above assumptions, the map f is order-preserving for the partial order of $\mathbb{R}^n$, that is $f(v) \leq f(v')$ for all $v, v' \in \mathbb{R}^n$ such that $v \leq v'$, and additively subhomogeneous, that is $f(\lambda + v) \leq \lambda + f(v)$ for all $v \in \mathbb{R}^n$ and $\lambda \geq 0$, where we use the notation $\lambda + v$ for the vector with entries $(\lambda + v)_i = \lambda + v_i$, hence it is necessarily nonexpansive for the sup-norm.

If we have in addition $\sum_{j=1}^n P_{ij}^{(a,b)} = 1$, for all $i \in [1:n]$, $a \in \mathcal{A}$, and $b \in \mathcal{B}$, then the map f is additively homogeneous, that is $f(\lambda + v) = \lambda + f(v)$ for all $v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. If such a map has an additive eigenvalue $\rho \in \mathbb{R}$, that is if there exists $v \in \mathbb{R}^n$ such that $f(v) = \rho + v$, then ρ is the value of the game where the objective of first player is now the mean payoff (in time in the long term).

2. Uniqueness of a fixed point. We assume now that $\mathcal{A}$ and $\mathcal{B}$ are compact, that $(a,b) \in \mathcal{A} \times \mathcal{B} \mapsto P_{ij}^{(a,b)}$, and $(a,b) \in \mathcal{A} \times \mathcal{B} \mapsto r_i^{(a,b)}$ are continuous, for all $i, j \in [1:n]$, and that

$$f(v)_i = \max_{a \in \mathcal{A}} \left(\min_{b \in \mathcal{B}} \left(\sum_{j=1}^n P_{ij}^{(a,b)} v_j + r_i^{(ab)} \right) \right) = \min_{b \in \mathcal{B}} \left(\max_{a \in \mathcal{A}} \left(\sum_{j=1}^n P_{ij}^{(a,b)} v_j + r_i^{(ab)} \right) \right)$$

for all $i \in X$ and $v \in \mathbb{R}^n$.

For all policies $\alpha \in \mathcal{A}^n$ and $\beta \in \mathcal{B}^n$ of the two players, let us denote by $P^{(\alpha,\beta)}$ and $r^{(\alpha,\beta)}$ the associated matrices and reward: $P_{i,j}^{(\alpha,\beta)} = P_{i,j}^{(\alpha_i,\beta_i)}$, and $r_i^{(\alpha,\beta)} = r_i^{(\alpha_i,\beta_i)}$. The previous equation can also be written in vectorial form:

$$f(v) = \max_{\alpha \in \mathcal{A}^n} \left(\min_{\beta \in \mathcal{B}^n} \left(P^{(\alpha,\beta)} v + r^{(\alpha,\beta)} \right) \right) = \min_{\beta \in \mathcal{B}^n} \left(\max_{\alpha \in \mathcal{A}^n} \left(P^{(\alpha,\beta)} v + r^{(\alpha,\beta)} \right) \right), \quad v \in \mathbb{R}^n \quad .$$

By a classical result of game theory, it is known that f admits directional derivatives at any $v \in \mathbb{R}^n$:

$$f'_v(x) = \max_{\alpha \in \mathcal{A}(v)} \left(\min_{\beta \in \mathcal{B}(v)} \left(P^{(\alpha,\beta)} x \right) \right) = \min_{\beta \in \mathcal{B}(v)} \left(\max_{\alpha \in \mathcal{A}(v)} \left(P^{(\alpha,\beta)} x \right) \right),$$

for $x \in \mathbb{R}^n$, where

$$\mathcal{A}(v) = \left\{ \alpha \in \mathcal{A}^n \mid f(v) = \min_{\beta \in \mathcal{B}^n} \left(P^{(\alpha,\beta)} v + r^{(\alpha,\beta)} \right) \right\}$$

$$\mathcal{B}_i(v) = \left\{ \beta \in \mathcal{B}^n \mid f(v) = \max_{\alpha \in \mathcal{A}^n} \left(P^{(\alpha,\beta)} v + r^{(\alpha,\beta)} \right) \right\} \quad .$$

Since f'_v is positively homogenous, one can consider for it several notions of spectral radius, see [4]. For instance for any positively homogenous map of $\mathbb{R}^n$, we shall denote by $r(h)$ the maximal modulus of an eigenvalue of h.

In the finite dimensional case, one shows in [1] that for order-preserving nonexpansive maps such as f'_v, all the usual spectral notions coincide with $r(h)$. Moreover, applying the results of [1] to the positively homogeneous map f'_v, and the results of [2] to f, and using the fact that $\mathbb{R}^n$ is finite dimensional, we get:

THEOREM 2.1. *The following properties hold for f'_v:*

1. $r(f'_v) = \max(\max_{\alpha \in \mathcal{A}(v)} \inf_{\beta \in \mathcal{B}(v)} r(P^{(\alpha,\beta)}), \max_{\beta \in \mathcal{B}(v)} \inf_{\alpha \in \mathcal{A}(v)} r(P^{(\alpha,\beta)}))$;
2. $r(f'_v) < 1$ *if and only if* f'_v *has a unique fixed point, that is* 0.

Assume that v *is a fixed point of* f *and that* $r(f'_v) < 1$*, then* v *is the unique fixed point of* f*. Moreover the value iteration converges with rate* $r(f'_v)$*:*

$$\limsup_{k \to \infty} \|f^k(w) - v\|^{1/k} \leq r(f'_v) \ .$$

When f si convex, that is in the case of one player, Property 1 reduces to $r(f'_v) = \max_{\alpha \in \mathcal{A}(v)} r(P^{(\alpha)})$, hence $r(f'_v) < 1$ if $r(P^{(\alpha)}) < 1$ for all $\alpha \in \mathcal{A}$. This last property holds in particular when the matrices $P^{(\alpha)}$ correspond to the restriction of irreducible stochastic matrices $Q^{(\alpha)}$ with same cemetery point, to the complementary of this cemetery point. Hence, one recovers the result of Bertsekas and Tsitsiklis [3], that under the assumption that all transition matrices of the stochastic control problem are irreducible with the same cemetery point, the value iteration converges with some rate $r < 1$, although we also give an explicit formula of this rate.

Other results of [2] concern the uniqueness of additive eigenvector and the convergence of value iterations in the undiscounted case.

REFERENCES

[1] M. Akian, S. Gaubert, and R. Nussbaum. A Collatz-Wielandt characterization of the spectral radius of order-preserving homogeneous maps on cones. 2011. arXiv:1112.5968.

[2] M. Akian, S. Gaubert, and R. Nussbaum. Uniqueness of the fixed point of nonexpansive semidifferentiable maps. 2012. arXiv:1201.1536.

[3] D.P. Bertsekas and J.N. Tsitsiklis. An analysis of stochastic shortest path problems. *Math. Oper. Res.*, 16(3):580–595, 1991.

[4] J. Mallet-Paret and Roger Nussbaum. Eigenvalues for a class of homogeneous cone maps arising from max-plus operators. *Discrete and Continuous Dynamical Systems*, 8(3):519–562, July 2002.