A suboptimality test for two person zero sum Markov games

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics

PROBABILITY THEORY, STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 76-19

A suboptimality test for two person zero sum Markov games

by

Dieter Reetz and Jan van der Wal

Revised October 1977

Eindhoven, October 1976

The Netherlands

A suboptimality test for two person
zero sum Markov games

by

Dieter Reetz and Jan van der Wal

Abstract. This paper presents a games version of the nonoptimality test given by Hastings for Markov decision processes. A pure action will be eliminated if compared to some randomized action it performs worse against any of the opponents possible actions.

1. Introduction and preliminaries

For Markov decision processes (MDP) several authors have proposed tests to eliminate suboptimal actions, a.o. [4,3,1,2]. In this note we give a test for the elimination of suboptimal actions in two person zero sum Markov games with finite state and action spaces.
Following the notation in [7] the Markov game is characterized by the state space $S := \{1,2,\dots,N\}$, for each state $x \in S$ two finite nonempty sets of actions K_x for player 1 (P_1) and L_x for P_2, and if in state x actions k and ℓ are taken, an immediate payoff from P_2 to P_1 $r(x,k,\ell)$ and transition probabilities $p(y|x,k,\ell)$, $y \in S$. We assume $\sum_{y\in S} p(y|x,k,\ell) < 1$ for all x, k and ℓ.
As criterion we use total expected rewards. Shapley [6] showed that this game has a value, which we will denote by v^*, as well as optimal stationary strategies.

A policy f for P_1 specifies the probabilities $f(x,k)$ by which action k is taken in state x. The randomized action in state x is denoted by $f(x)$.
In order to simplify the expressions in the remainder we define for all $v \in \mathbb{R}^N$, x, k and ℓ

$$r(x,k,\ell,v) := r(x,k,\ell) + \sum_{y\in S} p(y|x,k,\ell)v(y) .$$

Let $\{v_n\}$ be determined by the standard successive approximation method and λ_n, μ_n, a_n and b_n be defined as follows (cf. [7])

$$\lambda_n := \min_{x\in S} (v_n - v_{n-1})(x)$$

$$\mu_n := \max_{x\in S} (v_n - v_{n-1})(x)$$

$$a_n := \begin{cases} \max\limits_{x,k,\ell} \sum\limits_{y \in S} p(y|x,k,\ell) & \text{if } \lambda_n < 0 , \\ \min\limits_{x,k,\ell} \sum\limits_{y \in S} p(y|x,k,\ell) & \text{if } \lambda_n \geq 0 ; \end{cases}$$

$$b_n := \begin{cases} \max\limits_{x,k,\ell} \sum\limits_{y \in S} p(y|x,k,\ell) & \text{if } \mu_n \geq 0 , \\ \min\limits_{x,k,\ell} \sum\limits_{y \in S} p(y|x,k,\ell) & \text{if } \mu_n < 0 . \end{cases}$$

And let f_n be an optimal policy for P_1 in the 1-stage game with terminal payoff v_{n-1}, i.e.

$$\min_{\ell \in L_x} \sum_{k \in K_x} f_n(x,k)\, t(x,k,\ell,v_{n-1}) = v_n(x), \quad x \in S .$$

An action $k_0 \in K_x$ will be called *suboptimal at stage n* if no optimal policy f_n, satisfying the equality above, can have $f_n(x,k_0) > 0$. An action $k_0 \in K_x$ is called *suboptimal* if no optimal strategy $f^{*(\infty)}$, thus satisfying

$$\min_{\ell \in L_x} \sum_{k \in K_x} f^*(x,k)\, t(x,k,\ell,v^*) = v^*(x), \quad x \in S$$

can have $f^*(x,k_0) > 0$.
In the next section we present a test for eliminating actions for one or more stages which is a straightforward extension to Markov games of tests of Hübner [3], Hastings [1], Hastings and van Nunen [2] proposed for MDP.

2. The suboptimality test

First we prove an auxilary result which says when it is possible to eliminate actions.

Lemma 1. Let $v \in \mathbb{R}^N$ be given arbitrarily. And let there exist a probability distribution $\hat{f}(x)$ on K_x with $\hat{f}(x,k_0) = 0$, and

$$\sum_{k \in K_x} \hat{f}(x,k)\, t(x,k,\ell,v) > t(x,k_0,\ell,v) \quad \text{for all } \ell \in L_x .$$

Then action k_0 is suboptimal in the 1-stage game with terminal payoff v.

Proof. We will prove this by contradiction. Let $f^*(x)$ be an optimal randomized action for P_1 in the game above with $f^*(x,k_0) > 0$.
Now define the randomized action $\tilde{f}(x)$ by

$$\tilde{f}(x,k_0) = 0$$

$$\tilde{f}(x,k) = f^*(x,k) + f^*(x,k_0)\hat{f}(x,k), \quad k \neq k_0 .$$

Then we have for all $\ell \in L_x$

$$\sum_{k\in K_x} \tilde{f}(x,k)t(x,k,\ell,v) =$$

$$\sum_{k\neq k_0} f^*(x,k)t(x,k,\ell,v) + f^*(x,k_0)\sum_{k\in K_x} \hat{f}(x,k)t(x,k,\ell,v) >$$

$$\sum_{k\in K_x} f^*(x,k)t(x,k,\ell,v) .$$

But this contradicts the optimality of $f^*(x)$, hence $f^*(x,k_0) = 0$ for all optimal $f^*(x)$. I.e. k_0 is suboptimal in the 1-stage game with terminal payoff v. □

Now we can formulate the suboptimality test. Define $y_n(x,k_0)$ by

$$y_n(x,k_0) := \min_{\ell\in L_x} [\sum_{k\in K_x} f_n(x,k)t(x,k,\ell,v_{n-1}) - t(x,k_0,\ell,v_{n-1})] .$$

Now we may prove the following theorem:

<u>Theorem 1.</u> (cf. [2]).

i) If $y_n(x,k_0) - \sum_{\ell=n}^{n+m-1} (\mu_\ell b_\ell - \lambda_\ell a_\ell) > 0$ then action k_0 is suboptimal at stage $n+m$.

ii) If $y_n(x,k_0) - \sum_{\ell=n}^{\infty} (\mu_\ell b_\ell - \lambda_\ell a_\ell) > 0$ then action k_0 is suboptimal in the ∞-stage game.

<u>Proof.</u>

i)
$$\sum_{k\in K_x} f_n(x,k)t(x,k,\ell,v_{n+m-1}) - t(x,k_0,\ell,v_{n+m-1}) =$$

$$\sum_{k\in K_x} f_n(x,k)t(x,k,\ell,v_{n-1}) - t(x,k_0,\ell,v_{n-1}) +$$

$$\sum_{k\in K_x} f_n(x,k)[t(x,k,\ell,v_n) - t(x,k,\ell,v_{n-1})] - [t(x,k_0,\ell,v_n) - t(x,k_0,\ell,v_{n-1})]$$

$$\ldots + \sum_{k\in K_x} f_n(x,k)[t(x,k,\ell,v_{n+m-1}) - t(x,k,\ell,v_{n+m-2})] - [t(x,k_0,\ell,v_{n+m-1}) - t(x,k_0,\ell,v_{n+m-2})]$$

$$\geq y_n(x,k_0) + a_n\lambda_n - b_n\mu_n + \ldots + a_{n+m-1}\lambda_{n+m-1} - b_{n+m-1}\mu_{n+m-1} > 0 .$$

Hence with lemma 1 k_0 is suboptimal at stage $n+m$.

ii) From i) with $m \to \infty$. □

As we do not know a_ℓ, b_ℓ, λ_ℓ and μ_ℓ in advance there are two possible ways of using this test.

i) Eliminate action k_0 for as many stages as is possible at stage n. This means that k_0 is eliminated at stage n until stage $n+m$ where m is the largest integer (possibly ∞) for which

$$y_n(x,k_0) - \sum_{\ell=n}^{n+m-1} (b_n^{\ell+1-n}\mu_n - a_n^{\ell+1-n}\lambda_n) > 0$$

(where we use $b_n^{\ell+1-n}\mu_n - a_n^{\ell+1-n}\lambda_n \geq b_\ell\mu_\ell - a_\ell\lambda_\ell$) .

ii) Eliminate k_0 for one stage (if possible) after which you test whether it can be eliminated for another state in such a way that an action eliminated at stage n will return at stage $n+m$ where m is the first integer for which

$$y_n(x,k_0) - \sum_{\ell=n}^{n+m-1} (\mu_\ell b_\ell - \lambda_\ell a_\ell) < 0 .$$

3. Some final remarks

i) If we apply the suboptimality test we get exactly the same successive approximations v_n as in algorithm without the test (cf. Karlin [4], pp. 38-39).

ii) In the preceding sections we only treated the suboptimality test for actions of P_1 but the case for P_2 is completely symmetric.

iii) The test can be used also in other successive approximation algorithms, for example Jacobi, Gauss-Seidel (in this case the definitions of a_n and b_n must be adapted, cf. [7]).

iv) If at stage n_0 action $\ell_0 \in L_x$ is eliminated for all future iterations then in the definition of $y_n(x,k_0)$, $n > n_0$ we can take the minimum over $\ell \neq \ell_0$ instead of $\ell \in L_x$.

v) In the test for suboptimality at stage n the assumption $\sum_{y\in S} p(y|x,k,\ell) < 1$ plays no role at all, so this test can be used also in the finite horizon and average reward cases.

4. References

[1] Hastings, N.A.J., A test for nonoptimal actions in undiscounted Markov decision chains, Management Science 23 (1976), 87-92.

[2] Hastings, N.A.J. and J.A.E.E. van Nunen, The action elimination algorithm for Markov decision processes, Markov Decision Theory, eds. H.C. Tijms, J. Wessels, Amsterdam, Mathematisch Centrum (Mathematical Centre Tract no. 93), 1977, 161-170.

[3] Hübner, G., Improved procedures for eliminating suboptimal actions in Markov programming by the use of contraction properties, to appear in: Transactions of the seventh Prague Conf. 1976.

[4] Karlin, S., Mathematival Methods and Theory in Games, Programming and economics, Vol. 1, Addison-Wesley. Publishing Company, Reading, Massachusetts-London, 1959.

[5] MacQueen, J., A test for suboptimal actions in Markov decision problems, Operations Research 15 (1967), 559-561.

[6] Shapley, L.S., Stochastic games, Proc. Nat. Acad. Sci. USA 39 (1953), 1095-1100.

[7] Van der Wal, J., Discounted Markov games: successive approximation and stopping times, Intern. J. of Game Theory 6 (1977), 11-22.