stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 46/75 APRIL

JAC.M. ANTHONISSE AND H.C. TIJMS

ON WHITE'S CONDITION IN DYNAMIC PROGRAMMING

Prepublication

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS(MOS) subject classification scheme (1970): 90C40

On White's Condition in Dynamic Programming

by

Jac. M. Anthonisse and H.C. Tijms

ABSTRACT

This paper discusses White's condition on the stochastic matrices associated with the stationary policies in the familiar discrete-time dynamic programming model. We give an easily verifiable condition which implies White's one. Also, we present an example showing that White's condition is not equivalent to the assumption that the above stochastic matrices have each a single recurrent class and have a common aperiodic recurrent state.

KEY WORDS & PHRASES: *Dynamic programming, White's condition, single recurrent class, common aperiodic recurrent state.*

1. INTRODUCTION

Consider the familiar discrete-time dynamic programming model treated by HOWARD [3], where $\{1,\dots,N\}$ is the set of states, $A(i)$ is the finite set of actions available in state i, $r(i,a)$ is the immediate reward received from taking action a while in state i, and $q(j|i,a)$ is the probability that the next state of the system will be state j when action a is taken in state i.

A stationary policy f is a decision rule that adds to each state i a single action $f(i) \in A(i)$. Denote by F the class of all stationary policies. Associate with each $f \in F$ the $N \times N$ stochastic matrix $P(f)$ whose (i,j)th element to be denoted by $[P(f)]_{ij}$ equals $q(j|i,f(i))$. An important role in dynamic programming is played by the following condition:

WHITE'S CONDITION: There is some state r, an integer $\nu \geq 1$ and a number $\alpha > 0$ such that

$$[P(f_1)\dots P(f_\nu)]_{ir} \geq \alpha \quad \text{for all } 1 \leq i \leq N \text{ and all } f_1,\dots,f_\nu \in F.$$

This condition states that for each initial state there is a positive probability that the system will be in state r after ν transitions whatever sequence of actions is taken. Observe that this condition is equivalent to the condition that $[P(f_1)\dots P(f_n)]_{ir} \geq \alpha$ for all $1 \leq i \leq N$, all $n \geq \nu$ and all $f_1,\dots,f_n \in F$. Under the above condition WHITE [4] has shown that, for each fixed state i_0,

$$(1) \quad \lim_{n\to\infty} \{v_n(i) - v_n(i_0)\} \qquad \text{exists and is finite for all } 1 \leq i \leq N,$$

where $v_n(i)$ denotes the maximal total expected undiscounted reward for an n-stage process starting from state i. The convergence in (1) is exponentially fast at rate $0((1-\alpha)^{[n/\nu]})$ where $[x]$ is the largest integer less than or equal to x. Actually under the above condition the following stronger result holds

$$(2) \qquad \lim_{n\to\infty} \{v_n(i) - ng\} \qquad \text{exists and is finite for all } 1 \le i \le N,$$

where g denotes the maximal average expected reward per unit time for an infinite planning horizon (observe that g is independent of the initial state since each P(f) is unichained). Also, the convergence in (2) is exponentially fast at rate $0((1-\alpha)^{[n/\nu]})$. This result follows by making a minor modification of the proof of Theorem 4.3 in BATHER [1] and applying the fixed point theorem as stated on p.177 in DENARDO [2].

In practice White's condition may be difficult to verify, except when $\nu = 1$ applies. The purpose of this paper is to give an easily verifiable condition which implies White's condition. Also, we shall give an example showing that White's condition is not equivalent to the assumption that the stochastic matrices P(f) ($f \in F$) have each a single recurrent class and have a common *aperiodic* recurrent state.

2. RESULTS

The following theorem gives a condition implying White's condition.

THEOREM. *Assume that there is some state* r *such that*

(a) *The stochastic matrices* P(f) ($f \in F$) *have each a single recurrent class and have state* r *as common recurrent state*

(b) $q(r|r,a) > 0$ *for all* $a \in A(r)$.

Then, White's condition holds.

PROOF. We shall first decompose the set of all states into a finite number of disjoint sets. Define

$$S_0 = \{r\}$$

and, for $k \geq 1$, define S_k recursively by

$$S_k = \{i \mid i \notin \bigcup_{h=0}^{k-1} S_h \text{ and for each } a \in A(i) \text{ there is some } j \in \bigcup_{h=0}^{k-1} S_h \text{ such that } q(j \mid i, a) > 0\}.$$

Hence each state belonging to S_k $(k \geq 1)$ has the property that whatever action is chosen in that state there is a positive probability that the next state will belong to $\bigcup_{h=0}^{k-1} S_h$. We shall now prove that for each $k \geq 1$ holds that

$$(3) \qquad \bigcup_{h=0}^{k-1} S_h \neq \{1,\ldots,N\} \quad \text{implies} \quad S_k \neq \emptyset.$$

To prove this, assume to the contrary that $S_k = \emptyset$. Then we can construct a policy $f \in F$ such that

$$q(j|i,f(i)) = 0 \quad \text{for all } i \notin \bigcup_{h=0}^{k-1} S_h \text{ and all } j \in \bigcup_{h=0}^{k-1} S_h$$

which says that the set consisting of the states $i \notin \bigcup_{h=0}^{k-1} S_h$ forms a closed set for the stochastic matrix $P(f)$. This contradicts part (a) of the assumption since state $r \in \bigcup_{h=0}^{k-1} S_h$. Hence (3) holds. Since $S_k \cap S_m = \emptyset$ for $k \neq m$, it follows from (3) that there is an integer $\nu(<N)$ such that

$$S_k \neq \emptyset \quad \text{for } 0 \leq k \leq \nu \quad \text{and} \quad \bigcup_{k=0}^{\nu} S_k = \{1,\ldots,N\}.$$

We are now in a position to prove that

(4) $[P(f_1)\ldots P(f_\nu)]_{ir} > 0$ for all $1 \le i \le N$ and all $f_1, \ldots f_\nu \in F$.

Clearly, (4) implies White's condition since the class F is finite. To prove (4), choose $f_1, \ldots f_\nu \in F$ and fix state i. When $i = r$, (4) holds by part (b) of the assumption. Suppose now that $i \neq r$. Then $i \in S_m$ for some $1 \le m \le \nu$. Now, by the construction of the S_k's, we have for some $1 \le s \le m$

$$[P(f_1)\ldots P(f_s)]_{ir} > 0.$$

Together this and part (b) of the assumption imply (4) which ends the proof. □

It is easy to see that White's condition implies that the stochastic matrices P(f) (f∈F) have each a single recurrent class and have a common aperiodic recurrent state. We shall now give an example showing that the converse may be false when $N \ge 3$ (for the case of $N = 2$ states the converse is true as can be directly verified by considering all possible combinations). Consider the example in which

$$N = 3, \quad A(1) = A(3) = \{a_1\}, \quad A(2) = \{a_1, a_2\},$$
$$q(2|1,a_1) = 1 = q(2|2,a_1) \quad q(1|2,a_2) = q(3|2,a_2) = \tfrac{1}{2},$$
$$q(1|3,a_1) = 1.$$

The class F consists of two policies f_1 and f_2 where $f_1(2) = a_1$ and $f_2(2) = a_2$. Let $P_i = P(f_i)$ for $i = 1,2$, then

$$P_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}$$

Clearly, P_1 and P_2 have each a single recurrent class and have state 2 as common aperiodic recurrent state. However,

$$P_1^2P_2 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = (P_1^2P_2)^n \qquad \text{for all } n \geq 1$$

and

$$P_1^n = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad \text{for all } n \geq 2.$$

REFERENCES

[1] BATHER, J., *Optimal Decision Procedures for Finite Markov Chains, Part II: Communicating Systems,* Adv. App. Prob., Vol. 5 (1973), 521-540.

[2] DENARDO, E.V., *Contraction Mappings in the Theory Underlying Dynamic Programming,* Siam Review, Vol. 9 (1967), 165-177.

[3] HOWARD, R.A., *Dynamic Programming and Markov Processes,* The M.I.T. Press, Cambridge, Massachusetts (1960).

[4] WHITE, D.J., *Dynamic Programming, Markov Chains, and the Method of Successive Approximations,* J. Math. Anal. and Appl., Vol. 6 (1963), 373-376.