BEBR

FACULTY WORKING
PAPER NO. 780

Axioms and Examples Related to Ordinal Dynamic Programming

Charles E. Blair

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 780

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

June 1981

Axioms and Examples Related to Ordinal Dynamic Programming

Charles E. Blair, Associate Professor
Department of Business Administration

Abstract

We continue the work of Sobel on axioms for preferences in discrete Markov processes. Sufficient conditions for optimality are presented, and the logical interrelation with previous axiomations is discussed.

Axioms and Examples Related to Ordinal Dynamic Programming

by Charles E. Blair

We consider deterministic sequential Markov process. Let X be a set of states. For each $x \varepsilon X$, $M(x) \subset X$ is the set of states that can be reached in one step from x. Define Δ to be the set of mappings $\delta : X \to X$ such that $\delta(x) \varepsilon M(x)$ for every $x \varepsilon X$. A policy is an infinite sequence $\delta_1 \delta_2 \ldots$ where $\delta_i \varepsilon \Delta$. A stationary policy has all δ_i equal. For each policy $\pi = \delta_1 \delta_2 \ldots$ and each $x \varepsilon X$ there is a unique sequence $x_0 x_1 x_2 \ldots$ such that $x_0 = x$ and $x_n = \delta_n(x_{n-1})$, $n = 1, 2, \ldots$ We will denote this sequence by $P(\pi, x)$. For $x \varepsilon X$, Φ_x is defined to be the set of sequences $P(\pi, x)$ that arise as π varies over all possible policies. Φ_x is the set of all posterities with initial state x.

Sobel [1] studied situations in which orderings are assigned to the sets Φ_x, which satisfied various axioms. For p, q $\varepsilon \Phi_x$ we thus have an ordering under which either $p \geq q$ or $q \geq p$. The ordering on posterities induces a partial ordering on policies: $\pi_1 \geq \pi_2$ if and only if $P(\pi_1, x) \geq P(\pi_2, x)$ for all $x \varepsilon X$. An optimal policy π is one such that $\pi \geq \pi'$ for all policies π'.

[1,2] showed that, provided certain axioms hold with regard to the orderings on posterities and policies these results hold:

(1) If there exists an optimal policy, then there exists an optimal stationary policy.

(2) If $\pi = \delta_1 \delta_2 \delta_3 \ldots$ and for every $\delta \varepsilon \Delta$ $\pi \geq \delta \delta_1 \delta_2 \delta_3 \ldots = \delta \pi$, then π is optimal.

(3) If X is finite there is a stationary optimal policy.

We follow [1] in assuming throughout that the orderings on Φ_x satisfy

(4) if p_1, $p_2 \varepsilon \Phi_x$ and $x_0 \ldots x_n$ is a sequence such that $x_i \varepsilon M(x_{i-1})$ $1 \leq i \leq n$ and $x \varepsilon M(x_n)$ then $x_0 \ldots x_n p_1 \geq x_0 \ldots x_n p_2$ if and only if $p_1 \geq p_2$.

Here $x_0 \ldots x_n p$ is the sequence of states formed by concatenating $x_0 \ldots x_n$ and p. The hypotheses imply that these two sequences are members of Φ_{x_0}. The intuitive content is that if one sequence is preferable to another when x is the starting state, then the same holds if x is reached at a later time.

(4) is satisfied by most criteria that one would want to use in a dynamic programming problem. However additional assumptions must be made in order to obtain (1)-(3).

[1] proposes the "countable transitivity" axiom

(5) Let $p_i \varepsilon \Phi_x$ $i = 0, 1, 2, \ldots$. If for $i \geq 1$, the first i terms of p_i coincide with the first i terms of p_0 and $p_1 \leq p_2 \leq p_3 \leq \ldots$, then $p_0 \geq p_i$ for all i.

However (4) and (5) do not imply (2).*

Example: Let $X = \{0,1\}$. $M(0) = X$. $M(1) = \{1\}$. Φ_1 consists of the single posterity 1111... Φ_0 consists of the posterities 0000... and $0^k111\ldots$ for $k \geq 1$. Define 000... > 01111... > 0011... > etc. (4) is easy to verify. (5) is satisfied because $p_1 \leq p_2 \leq p_3 \ldots$ implies (in this example) that $p_i = p_{i+1}$ for all sufficiently large i. Let

*This corrects theorem 3 of [1]. Sobel had discovered this independently while writing [3]. This motivated the use of the alternative axiom (6) in [2].

$\delta_1(0) = 1 = \delta_1(1)$ Then the policy $\pi = \delta_1\delta_1\delta_1 \ldots = \delta_1^\infty$ satisfies $\pi \geq \delta\pi$ for any $\delta\varepsilon\Delta$. But if $\delta_2(0) = 0$ and $\pi' = \delta_2^\infty$ then $P(\pi,0) = 01111\ldots$ $\not\geq P(\pi',0) = 000\ldots$, hence $\pi \not\geq \pi'$ and π is not optimal.

It can be shown that (4) and (5) imply strengthened versions of (1) and (3).

<u>Theorem 1</u>: Assume (4) and (5) hold. Suppose that there is a $\delta\varepsilon\Delta$ such that, for every $x\varepsilon X$, if $p\varepsilon\Phi_x$ there is a $p'\varepsilon\Phi_x$ whose first two terms are $x, \delta(x)$ with $p' \geq p$. Then δ^∞ is an optimal policy.

<u>Proof</u>: Let $x\varepsilon X$, $p\varepsilon\Phi_x$. We will construct a sequence of $p_i\varepsilon\Phi_x$ such that $p_1 = p \leq p_2 \leq p_3 \leq \ldots$ and the first i members of p_i coincide with the first i members of $P(\delta^\infty, x)$. We start with $p_1 = p$ and continue by induction. If $p_1, \ldots p_n$ have already been constructed let $p_n = x_0x_1 \ldots$ By hypothesis, there is a $q\varepsilon\Phi_{x_{n-1}}$ such that $q \geq x_{n-1}x_n \ldots$ and the first two terms of q are x_{n-1} and $\delta(x_{n-1})$. By (4), $p_{n+1} = x_0x_1\ldots x_{n-2}q$ $\geq p_n$. This completes the construction of the p_i. (5) implies that $P(\delta^\infty,x) \geq p$. Since x, p were arbitrary δ^∞ is optimal. Q.E.D.

Theorem 1 has a converse in the sense that if no δ exists satisfying the hypothesis then no policy is optimal.

<u>Corollary 2</u>*: If $\pi = \delta_1\delta_2\delta_3 \ldots$ is an optimal policy, then δ_1^∞ is an optimal policy.

<u>Proof</u>: In this case p' in the hypothesis of Theorem 1 is $P(x,\pi)$. Q.E.D.

---

*This result is established in the proof of Theorem 2 of [1]

Corollary 3: If X is finite there is a stationary optimal policy.

Proof: For each $x \varepsilon X$, $\Phi_x = \bigcup_{y \varepsilon M(X)} Q_y$, where Q_y consists of those posterities whose first two terms are x,y. If an ordered set is the union of finitely many subsets at least one of the subsets is such that, for each point of the set, there is a point of the subset at least as large. If $\delta(x)$ is chosen so that $Q_{\delta(x)}$ is such a subset, then the hypothesis of Theorem 1 is satisfied and δ is a stationary optimal policy.

An alternative to (5) was proposed in [2]:

(6) Let $\pi = (\delta_1 \delta_2 \ldots)$ and ξ be two policies.

then $\xi \geq \delta_1 \ldots \delta_k \xi$ for all k implies $\xi \geq \pi$

$\xi \leq \delta_1 \ldots \delta_k \xi$ for all k implies $\xi \leq \pi$.

(4) and (6) together imply (1), (2) and (3). However there are two objections to (6). First, it discusses the partial ordering on policies rather than the total ordering on posterities, and is thus somewhat indirect. Second, (6) excludes lexicographic discounted-return criteria, a fairly natural class of preference orderings (example 3 of [1]).

Example 2: Let $X = \{0,1\}$. $M(0) = M(1) = X$. For a posterity $P = x_0 x_1 x_2 \ldots$ define $v_i(P) = \sum_{n=1}^{\infty} (\frac{1}{2})^n r_i(x_{n-1}, x_n)$, $i = 1,2$. $r_1(0,0) = 1$; $r_1(1,1) = 2$; $r_1(0,1) = r_1(1,0) = 0$. $r_2(0,1) = 1$; $r_2(0,0) = r_2(1,1) = r_2(1,0) = 0$. For p, $p' \varepsilon \Phi_x$ $p \geq p'$ iff $v_1(p) > v_1(p')$ or $v_1(p) = v_1(p')$ and $v_2(p) \geq v_2(p')$. Let $\xi = \delta_1^{\infty}$, where $\delta_1(0) = \delta_1(1) = 0$. Let $\pi = \delta_2 \delta_3^{\infty}$ where $\delta_2(0) = 1$, $\delta_2(1) = 0$; $\delta_3(0) = 0$, $\delta_3(1) = 1$. $v_1(P(\xi,0)) = v_1(0^{\infty}) = 1$, $v_1(P(\xi,1)) = \frac{1}{2}$. $v_1(P(\delta_2 \xi, 0)) = v_1(010^{\infty}) < v_1(P(\xi,0))$.

Since $P(\delta_2\xi,1) = P(\xi,1)$, it follows that $\xi \geq \delta_2\xi$. Similarly, it can be verified that $\xi \geq \delta_2\delta_3{}^k\xi$ for every positive k. Since $v_1(P(\pi,0)) = v_1(01^\infty) = 1 = v_1(P(\xi,0))$ and $v_2(P(\pi,0)) = \frac{1}{2} > v_2(P(\xi,0)) = 0$, it follows that $\xi \not\geq \pi$, which contradicts (6).

An alternative to (6) is the "dual" to (5).

(5′) Let $p_i \varepsilon \Phi_x$ $i = 0,1,2,\ldots$ If for $i \geq 1$, the first i terms of p_i coincide with the first i terms of p_0 and $p_1 \geq p_2 \geq p_3 \geq \ldots$, then $p_0 \leq p_i$ for all i.

<u>Theorem 2</u>: (4) and (5′) imply (2).

<u>Proof</u>: Suppose $\pi \geq \delta\pi$ for every $\delta\varepsilon\Delta$ and let $\xi = \delta_1\delta_2\delta_3 \ldots$ and $x\varepsilon X$. Then repeated application of (4) gives $\pi \geq \delta_1\pi \geq \delta_1\delta_2\pi \geq \delta_1\delta_2\delta_3\pi \geq \ldots$ hence $P(\pi,x) \geq P(\delta_1\pi,x) \geq P(\delta_1\delta_2\pi,x) \geq \ldots$ Hence (5′) implies $P(\pi,x) \geq P(\xi,x)$. Since x and ξ were arbitrary this implies π is optimal. Q.E.D.

<u>Corollary</u>: If the orderings on Φ_x are given by lexicographic discounted-return criteria then (1), (2), (3) hold.

<u>Proof</u>: It suffices to verify that (5) and (5′) both hold. This is easily established by noting that $v_i(p_0) = \lim_{n\to\infty} v_i(p_n)$. Q.E.D.

It seems that (5) and (5′) are preferable to (6). [1] suggests that there are several problems still to be addressed in the case in which X is infinite. We would like to mention this issue: in those cases in which there is no optimal stationary policy (hence no optimal policy by (1)) when is it the case that for every policy π there is a stationary policy δ^∞ such that $\delta^\infty \geq \pi$?

Acknowledgement

These problems were suggested to me by Professor Sobel, with whom I have had many helpful discussions.

References

1. M. J. Sobel, "Ordinal Dynamic Programming," Management Science vol. 21 (1975) pp. 967-975.

2. M. J. Sobel, "Ordinal Sequential Games," Georgia Institute of Technology, College of Management, Technical Report MS-80-3.

3. D. P. Heyman and M. J. Sobel, Stochastic Models in Operations Research, Vol. 2, McGraw-Hill, New York, to appear.