MATHEMATICS OF OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0364-765X | EISSN 1526-5471 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Uniform turnpike theorems for finite Markov decision processes

Mark E. Lewis
School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, mark.lewis@cornell.edu

Anand Paul
Warrington College of Business, Information Systems and Operations Management Department, University of Florida, Gainesville, FL 32611-7169 anand.paul@warrington.ufl.edu

A turnpike integer is the smallest finite horizon for which an optimal infinite horizon decision is the optimal initial decision. An important practical question considered in the literature is how to bound the turnpike integer using only the problem inputs. In this paper, we consider turnpike integers as a function of the discount factor. While a turnpike integer is finite for any fixed discount factor, we show that it approaches infinity in the neighborhood of a specific set of discount rates (for all but some exceptional finite Markov decision processes). We completely characterize this taboo set of discount factors and find necessary and sufficient conditions for a set of turnpike integers to be unbounded. This finding provides a cautionary tale for practitioners using point estimates of the discount factor to manage the length of rolling horizon procedures.

Key words: Dynamic Programming/Optimal Control: Markov – Finite State, Decision Analysis: Theory, Inventory/production: Planning horizons
History:

1. Introduction.

The phrase "turnpike theorem" was first used by economists as an umbrella term for asymptotic results pertaining to the optimal path of capital accumulation over a long period. This idea has since branched off and enjoyed an independent development in the field of operations research, and in particular, in Markov decision process (MDP) theory. The importance of turnpikes in both application and theory is based on one simple fact: solving a decision problem to optimality is simplified under the assumption that the time horizon for decision-making is infinite. In essence, the problem facing the decision-maker is repeated *ad infinitum* under the infinite horizon assumption while each step changes the scenario when the horizon is finite.

Shapiro (1968) proved what is perhaps the first turnpike theorem for finite state and action, discounted Markov decision processes. The main result states that for any fixed discount factor $\alpha < 1$, there exists a positive integer $N(\alpha)$ called a *turnpike integer* such that for $n \geq N(\alpha)$, the first period decision rule in some optimal n-period policy is identical to the decision rule that defines a stationary optimal policy for the infinite horizon discounted expected return criterion. We may informally explain the significance of this finding as follows. When solving a finite horizon MDP over a sufficiently large number of periods, we choose actions for a certain number of periods as though we were optimizing over an infinite horizon. As the end of the horizon draws near actions may need to depart from the set of optimal infinite horizon actions to account for an end of the horizon effect. The optimal finite horizon strategy, then, is to follow a stationary policy for a

certain number of periods, and then leave it during the terminal phase of the horizon. This implies that if the optimal infinite horizon policy is unique, then the optimal rolling horizon strategy is a stationary strategy (Shapiro (1968), page 297). That is to say that the common business practice of forecasting out a certain number of periods, setting the control policy for that number of periods and then repeating this process at the end of every period can focus more on obtaining forecasts than on solving for the optimal control.

For practical applications, it is important to find bounds for turnpike integers. In this paper, we approach the question of finding said bounds by studying the turnpike integer associated with intervals of discount rates rather than with individual discount rates. We know that the interval of discount factors from $[0,1)$ can be partitioned with a different stationary policy being optimal over different sub-intervals (see for example Kallenberg (2009)). Starting from this partition of $[0,1)$ (which we call the canonical partition) we show the following:

- Precluding the possibility of what we term degenerate points, any proper subset of a subset in the canonical partition has associated it with a finite turnpike integer.
- The neighborhood of the boundary between intervals in the canonical partition is associated with an unbounded set of turnpike integers in all finite MDPs for which value iteration does not converge finitely.

The pattern of turnpike integers as a function of discount rate that we discover expands significantly what is suggested in the literature. For instance, Shapiro (1968) found a finite upper bound for the turnpike integer at an arbitrary fixed discount rate and noted that the bound approached ∞ as the discount rate approached 1. This observation carries the tacit implication that the set of turnpike integers corresponding to any set of discount rates bounded away from unity is bounded. However, we show that in general a uniform bound cannot be provided even for subsets of discount rates bounded away from unity. Moreover, again under the assumption that there are no degenerate points, we show that the set of points where the turnpike integer may approach infinity from one side can be computed in polynomial time (see Section 3.3 for further discussion)

The rest of the paper is organized as follows. Section 2 provides a brief sketch of the background material for our research, making some fundamental definitions and recapitulating the main results from past work. Section 3 begins with a numerical example that we use to motivate the theoretical results that follow. In Section 3.1 we prove that the set of turnpike integers in a closed subset of an interval in the canonical partition is finite; further, we show that this set can be partitioned into intervals such that each interval is associated with a fixed turnpike integer. In Section 3.2 we show via an example that the set of turnpike integers in the neighborhood of the boundary point between contiguous intervals in the canonical partition may be unbounded. We then provide sufficient conditions for a set of turnpike integers to be unbounded. In Section 4 make some observations about turnpike integers in the Blackwell interval. We conclude with a discussion of some practical implications of our theoretical results in Section 5.

2. Background and Preliminaries. Consider a homogeneous Markov decision process (MDP) with finite state and action sets. Let $\mathbb{X}$ denote the state space of cardinality $|\mathbb{X}| = S < \infty$, and $\mathbb{A} := \bigcup_{x\in\mathbb{X}} A(x)$ be the action space, where $A(x)$ represents the (finite) set of available actions in state $x \in \mathbb{X}$. A decision-maker that finds the system in state x, chooses an action, say a from $A(x)$, accrues a non-negative reward $r(x,a)$ and then awaits the change in state that is governed by the transition probability $p(y|x,a)$ for $y \in \mathbb{X}$. This process continues over a finite or infinite planning horizon. The objective is to maximize the total discounted expected return or equivalently, minimize total expected cost. Thus, following the usual nomenclature (see Puterman (1994)) we define a *decision rule* $\{f_n(x), x \in \mathbb{X}\}$ as a vector of actions to choose at time n when in each state $x \in \mathbb{X}$. A *policy* is a sequence of decision rules $\pi = \{f_0, f_1, \ldots\}$ that specifies what action to use for each state for all time. Let F denote the set of all decision rules, and let Π be the set of all

non-anticipating policies. A policy is called stationary if it uses the same decision rule at each decision epoch. We denote such a policy (that uses decision rule f) as f^∞.

The total $n-$horizon $\alpha-$discounted expected reward under policy $\pi = \{f_0, f_1, \ldots\}$ given the starting state x is

$$v^\pi_{n,\alpha}(x) = \sum_{k=0}^{n-1} \alpha^k \, \mathbb{E}^\pi_x \, r(X_k, f_k(X_k)),$$

where $\{X_k, k \geq 0\}$ is the Markov chain generated by using policy π. When n is finite, we need only consider the first n decision rules of each policy. In the case that $n = \infty$ we simplify the notation to $v^\pi_\alpha(x)$. Define $\ell(x) := \max_{\pi \in \Pi}\{\ell^\pi(x)\}$ for $\ell = v_{n,\alpha}$ or v_α, depending on whether we are considering a finite or infinite horizon, respectively. The decision-maker seeks a policy say $\pi^* \in \Pi$ such that $v^{\pi^*}_{n,\alpha}(x) = v_{n,\alpha}(x)$ for all $x \in \mathbb{X}$. It is well-known that in the finite horizon case, the optimal value function satisfies the finite horizon optimality equations (FHOE) (for each $x \in \mathbb{X}$)

$$v_{n+1,\alpha}(x) = \max_{a \in A(x)} \{r(x,a) + \alpha \sum_{y \in X} p(y|x,a) v_{n,\alpha}(y)\}. \tag{1}$$

Similarly, in the infinite horizon case, the optimal value function satisfies the discounted reward optimality equations (DROE) (again for each $x \in \mathbb{X}$)

$$v_\alpha(x) = \max_{a \in A(x)} \{r(x,a) + \alpha \sum_{y \in X} p(y|x,a) v_\alpha(y)\}. \tag{2}$$

Remark 1. We recall that each value function is a bounded piecewise rational function (see Proposition 4.5.3 of Sennott (1999)) of α.

The interval $[0,1)$ of discount factors can be partitioned into a finite number of subintervals $\{I_i, i = 0, 1, \ldots \ell\}$ of the form $[a_i, a_{i+1})$ - where $a_i < a_{i+1}$, $a_0 = 0$ and $a_\ell = 1$ such that a fixed stationary infinite horizon policy - say $(d^*_i)^\infty$- is optimal for all the discount factors in I_i (see Kallenberg (2009), Theorem 5.9, page 141). We assume the partition is chosen in such a way that for all sufficiently small $\epsilon > 0$ we have $v_\alpha^{(d^*_i)^\infty} > v_\alpha^{(d^*_{i-1})^\infty}$ for all $\alpha \in (a_i, a_i + \epsilon)$ and $v_\alpha^{(d^*_{i-1})^\infty} > v_\alpha^{(d^*_i)^\infty}$ for all $\alpha \in (a_i - \epsilon, a_i)$.

Definition 1. We shall refer to the partition of $[0,1)$ into subintervals I_i as described above as the **canonical partition** and the intervals that comprise it as the **partition intervals**. Further, we shall refer to any of the points a_i marking the boundary between contiguous partition intervals as **break points**.

Remark 2. It may be that there are two or more stationary optimal policies for the infinite horizon problem for all $\alpha \in I_i$ for some I_i in the canonical decomposition. Note that the value function associated with any stationary infinite horizon policy is an analytic function of α. It is a standard fact (called analytic continuation) that if two or more analytic functions defined on $(0,1)$ assume the same values for any subinterval of $(0,1)$, then they must be identical throughout $(0,1)$. It follows that if distinct policies f^∞ and g^∞ are associated with the same optimal value function for some proper subset of I_i, then these policies are associated with the same optimal value function for all $\alpha \in (0,1)$. In this sense, f^∞ and g^∞ are in the same equivalence class and $(d^*_i)^\infty$ as defined above is unique up to this class. However, for the sake of precision we are careful to finesse the proofs of our theorems to account for the possibility of multiple stationary infinite horizon optimal policies.

Note that by construction and continuity of the value function, there are (at least) two optimal policies at each end point of every interval in the canonical partition. Furthermore, each partition interval I_i of the canonical partition is associated with an optimal value function $v_\alpha(x)$ for all $x \in \mathbb{X}$ and for all $\alpha \in I_i$. This value function is *unique up to an equivalence class*, as explained in Remark 2. The rightmost interval is associated with a particular optimality criterion.

DEFINITION 2. A stationary policy $(d^*)^\infty$ is called **Blackwell optimal** if there exists a discount factor α_∞ such that $(d^*)^\infty$ is infinite horizon $\alpha-$discounted reward optimal for all $\alpha \in [\alpha_\infty, 1)$. We refer to the subinterval $[a_{\ell-1}, a_\ell) = [\alpha_\infty, 1)$ as the **Blackwell interval**.

Turnpike integer and turnpike interval: We conclude this section with the formal definition of a turnpike integer and a turnpike interval. Informally, recall that a turnpike integer at a given discount factor is the smallest finite horizon for which an optimal infinite horizon decision is the optimal initial decision for the finite horizon problem. Let $F^\infty(\alpha)$ denote the set of decision rules that achieve the maximum in the DROE. Note $F^\infty(\alpha)$ also defines the set of optimal infinite horizon stationary policies at discount rate α. Similarly, suppose $F^n(\alpha)$ is the set of all initial optimal decision rules for the n-horizon α-discounted problem (without terminal reward).

DEFINITION 3. For each discount factor $\alpha \in [0, 1)$ let $N^*(\alpha)$ denote the smallest positive integer $N(\alpha)$ such that for all $n \geq N(\alpha)$

$$F^n(\alpha) \subseteq F^\infty(\alpha). \tag{3}$$

We call $N(\alpha)$ the **turnpike integer at discount rate** α. If I is an interval of α-values, we define $N^*(I)$ - **the turnpike integer for the interval I** - to be $sup_{\alpha \in I} N(\alpha)$.

The existence and finiteness of $N(\alpha)$ was established by Shapiro (1968).

DEFINITION 4. We say that I is a **turnpike interval** if $N(\alpha_i) = N(\alpha_j)$ for all $\alpha_i, \alpha_j \in I$.

Note the difference between the turnpike integer for an interval (which always exists, but may be infinite) and turnpike intervals; we provide conditions for their existence (see Theorem 1).
Having made these definitions, we may now restate the mission of this paper as follows. For a given finite MDP, we wish to study the structure of the set of turnpike integers for discount rates in $[0, 1)$. One would perhaps expect the graph of $N(\alpha)$ to be a bounded step function as α varies over $[0, \alpha_0]$ for any $\alpha_0 < 1$. We shall see that this intuitive picture is almost invariably false, in the sense that the corresponding necessary and sufficient conditions hold true only for exceptional classes of Markov decision processes. In fact, at the break points of the canonical decomposition an optimal policy is optimal as the discount factors approaches from one side, but not the other. This leads to the turnpike integer approaching a particular point from one direction, but infinity from the other. See Figure 1. Thus, if the estimate of the discount factor near one of these points is inaccurate (even by a little) the bound used for the turnpike integer can lead to dire consequences.

To conclude this section, we need to account for the possible existence of a point in the interior of an interval of the canonical decomposition where $N(\alpha)$ exhibits similar behavior to that depicted in Figure 1. Suppose α^* is in the interior of a subinterval I_i and a decision rule $d^* \neq d_i^*$ is such that $(d^*)^\infty$ is infinite horizon optimal at the point α^* - as is $(d_i^*)^\infty$ - but strictly sub-optimal for all other points in I_i. We shall refer to a point satisfying this property to be a **degenerate point**. There may be more than one point with this property, in a single sub-interval or in more than one sub-interval. We have not observed any examples that fit this definition, but we are unable to rule it out.

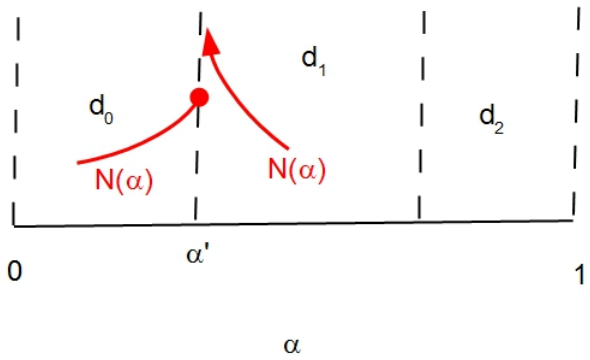

FIGURE 1. Convergence of turnpike integers to the endpoints of the canonical decomposition.

2.1. Turnpikes in deterministic optimal control theory. We briefly describe the turnpike problem in deterministic control theory, based on Zaslavski (2014), pages 15-16 (we have reformulated it here for a finite state space). The concept of a turnpike in this context is similar to that in Markov decision processes, but it gives rise to mathematically distinct problems. Consider a discrete-time control system with the state space $\mathbb{X}$ with metric ρ. The objective to be maximized is a function $v : \mathbb{X} \times \mathbb{X} \mapsto \mathbb{R}$. Let $\Omega \subset \mathbb{X} \times \mathbb{X}$ denote a class of admissible trajectories. Given an integer time horizon $T \geq 1$ and boundary conditions $x_0 = z, x_T = y$, the problem is to find $x_1, ..., x_{T-1}$ so as to maximize

$$\sum_{i=0}^{T-1} v(x_i, x_{i+1}).$$

The simplest analogy is to consider a routing problem where the user must start at time zero in state z and end at time T in state y choosing the pathway of states to visit in between.

The objective function is said to possess the *turnpike property* if there exists a point $\bar{x} \in \mathbb{X}$ such that the following conditions holds: for each $\epsilon > 0$ there exists an integer $L \geq 1$ such that for each integer $T \geq 2L$ and each solution $x_0, x_1, ..., x_T$ of the maximization problem described above, we have that $\rho(x_i, \bar{x}) \leq \epsilon$ for all $i = L, ..., T - L$. Note that L does not depend on T, y or z. The point $\bar{x}$ is defined to be the turnpike. Zaslavski (2014) explains that the term turnpike arises from the fact that in the routing example above, en route to the destination one should move to the turnpike and then stay there until the end of the trip approaches. This turnpike interpretation is similar to the stationary policy interpretation in the current paper.

The same concept arises in an infinite horizon setting. Suppose v is a concave function. The turnpike $\bar{x}$ is the unique solution to the problem of maximizing $v(x, x)$ over all $(x, x) \in \Omega$. Further, it is the case that any sequence $\{x_i, i \geq 0\}$ converges to the turnpike $\bar{x}$ or else the sequence $\{\sum_{i=0}^{T-1} v(x_i, x_{i+1} - Tv(\bar{x}, \bar{x})\}$ diverges to $-\infty$. In recent years the turnpike property has been extended from its original setting of a convex space and a convex objective to more general spaces and objectives.

If we consider maximizing

$$\sum_{i=0}^{\infty} \alpha^t v(x_t, x_{t+1})$$

we obtain a discounted problem for which several turnpike theorems have been obtained (see Khan and Piazza (2011) for a recent overview). However, we have not found any research that studies the turnpike property as a function of the discount rate, which is the focus of the present paper in the setting of finite Markov decision processes.

3. Turnpike theory. Throughout this section, we assume that the canonical partition consists of more than a single interval. This of course covers all but the most trivial of MDPs. The example below illustrates several insights for turnpike integers as a function of the discount factor.

EXAMPLE 1. Consider an MDP with $\mathbb{X} = \{0,1,2,3\}$. We consider an expected cost minimization problem (the ideas are precisely the same for an expected reward maximization problem, with merely a reversal in sign). There is exactly one action in states 1,2, and 3, while there are 2 actions in state zero. Let $p(1|1,a_1) = p(3|2,a_2) = p(2|3,a_3) = 1$, while the costs are $c(1,a_1) = 0$; $c(2,a_2) = 1$; $c(3,a_3) = 1$. In state 0 we have actions b,c where

$$p(1|0,b) = p(3|0,b) = \frac{1}{2} \qquad c(0,b) = 1$$

$$p(1|0,c) = \frac{1}{8}, \quad p(3|0,c) = \frac{7}{8} \qquad c(0,c) = \frac{1}{8}.$$

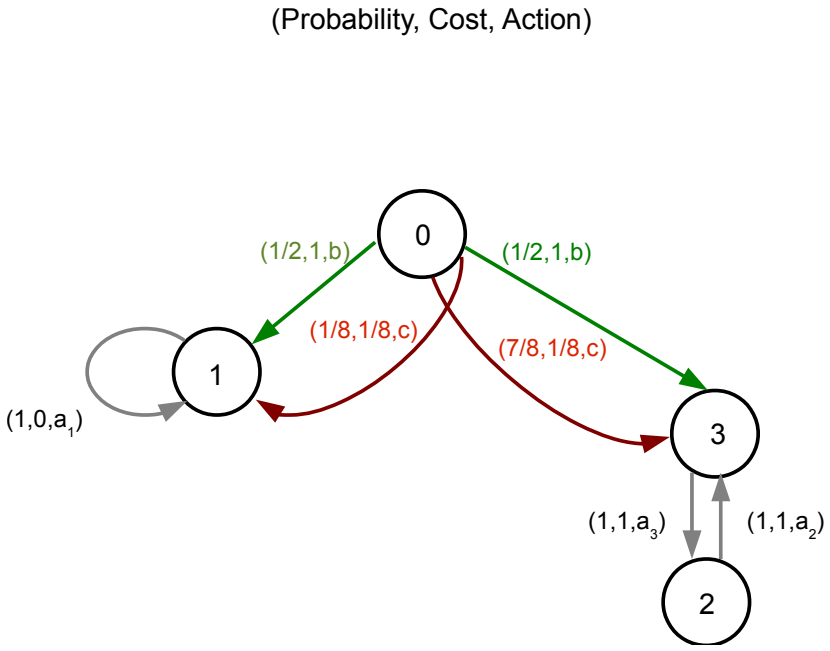

FIGURE 2. Example 1 data.

It should be clear that $v_\alpha(1) = 0$. Using the optimality equations yields,

$$v_\alpha(2) = 1 + \alpha v_\alpha(3)$$
$$v_\alpha(3) = 1 + \alpha v_\alpha(2)$$

so that

$$v_\alpha(3) = 1 + \alpha(1 + \alpha v_\alpha(3))$$
$$\Rightarrow (1-\alpha)^2 v_\alpha(3) = 1 + \alpha.$$
$$\Rightarrow v_\alpha(3) = v_\alpha(2) = \frac{1}{1-\alpha}$$

Now to get $v_\alpha(0)$, note

$$\begin{aligned} v_\alpha(0) &= \min\left\{\overbrace{1+\frac{\alpha}{2}[v_\alpha(1)+v_\alpha(3)]}^{\text{use action b}}, \overbrace{\frac{1}{8}+\alpha[\frac{1}{8}v_\alpha(1)+\frac{7}{8}v_\alpha(3)]}^{\text{use action c}}\right\} \\ &= \min\left\{1+\frac{\alpha}{2}[v_\alpha(3)], \frac{1}{8}+\alpha[\frac{7}{8}v_\alpha(3)]\right\} \\ &= \min\left\{\frac{7}{8}+\frac{\alpha}{2(1-\alpha)}, \frac{7\alpha}{8(1-\alpha)}\right\}+\frac{1}{8} \\ &= \min\left\{\frac{7}{8}, \frac{3\alpha}{8(1-\alpha)}\right\}+\frac{1}{8}+\frac{\alpha}{2(1-\alpha)}. \end{aligned}$$

Notice that for $\alpha \in (0,1)$ we know $\frac{\alpha}{(1-\alpha)}$ is continuous, non-decreasing and ranges from zero to ∞. We should use action c for $\alpha \leq .7$ and action b for $\alpha \geq .7$. Thus, the canonical partition is $[0,.7)$, $[.7,1)$.

Now, to compute the finite horizon optimal policy, note that $v_{n,\alpha}(1)$ remains zero. Since we have a geometric series we also know that

$$v_{n,\alpha}(2) = v_{n,\alpha}(3) = \frac{1}{1-\alpha} - \frac{\alpha^n}{1-\alpha} = \frac{1-\alpha^n}{1-\alpha}.$$

Using the same logic as the infinite horizon case, we have

$$\begin{aligned} v_{n+1,\alpha}(0) &= \min\left\{\overbrace{1+\frac{\alpha}{2}[v_{n,\alpha}(1)+v_{n,\alpha}(3)]}^{\text{use action b}}, \overbrace{\frac{1}{8}+\alpha[\frac{1}{8}v_{n,\alpha}(1)+\frac{7}{8}v_{n,\alpha}(3)]}^{\text{use action c}}\right\} \\ &= \min\left\{1+\frac{\alpha}{2}[v_{n,\alpha}(3)], \frac{1}{8}+\alpha[\frac{7}{8}v_{n,\alpha}(3)]\right\} \\ &= \min\left\{\frac{7}{8}+\frac{\alpha(1-\alpha^n)}{2(1-\alpha)}, \frac{7\alpha(1-\alpha^n)}{8(1-\alpha)}\right\}+\frac{1}{8} \\ &= \min\left\{\frac{7}{8}, \frac{3\alpha(1-\alpha^n)}{8(1-\alpha)}\right\}+\frac{1}{8}+\frac{\alpha(1-\alpha^n)}{2(1-\alpha)} \\ &= \frac{3}{8}\min\left\{\frac{7}{3}, \frac{\alpha(1-\alpha^n)}{(1-\alpha)}\right\}+\frac{1}{8}+\frac{\alpha(1-\alpha^n)}{2(1-\alpha)}. \end{aligned} \tag{4}$$

Notice that at $\alpha = .7$ we have

$$\frac{\alpha(1-\alpha^n)}{(1-\alpha)} = \frac{.7(1-(.7)^n)}{.3} = \frac{7}{3}(1-(.7)^n) < \frac{7}{3}.$$

Thus, at $\alpha = .7$, where in the infinite horizon case a decision-maker is indifferent, here it is optimal (for any n) to use action c. By contrast, consider $\alpha = .8$ (where b should eventually be optimal)

$$\frac{\alpha(1-\alpha^n)}{(1-\alpha)} = \frac{.8(1-(.8)^n)}{.2} = 4(1-(.8)^n).$$

Thus, for $n \leq 3$, action c is optimal, but for $n \geq 4$ we have that action b is optimal. A detailed numerical analysis (see Table 1) reveals the following turnpike intervals:

Interval	Turnpike integer
(0,0.700]	1
[0.7045,0.7065]	11
[0.707,0.7095]	10
[0.71,0.7145]	9
[0.715,0.722]	8
[.723,0.734]	7
[0.735,0.755]	6
[0.757,0.795]	5
[0.796,0.878]	4
[0.88,1)	3

TABLE 1. Turnpike integers for Example 1.

Example 1 provides several insights. We formalize the following insights in the next section.

- If we approach the boundary of the first set in the canonical partition from the left ($\alpha \in [0, .7)$), the turnpike integer is finite (in fact 1 throughout).
- On the other hand, if we approach it from the right ($\alpha \in (.7, 1)$), the turnpike integer grows to 11 but is bounded as long we stay away from .7.
- Also note that it is possible that the set of turnpike intervals is finer than the canonical partition.

3.1. Turnpike integers in the interior of the canonical partition. Example 1 shows that the endpoints of the canonical partition are potentially points of interesting behavior. We show later that degenerate points - if they happen to exist - are also critical. We expect the turnpike integers for intervals bounded away from these points to be finite. Before we establish these assertions we set the stage by stating several lemmas, most of which are standard results. The first follows directly from Remark 1.

LEMMA 1. *For each fixed $x \in \mathbb{X}$, distinct value functions $v_\alpha^{\pi_1}(x)$ and $v_\alpha^{\pi_2}(x)$ associated with stationary infinite horizon policies π_1 and π_2, respectively (regarded as functions of α) intersect finitely many times.*

Next we establish Proposition 1, a result which may be of independent interest and is an important step in the proof of Theorem 1. For each $x \in \mathbb{X}$ and a fixed discount rate α, the pointwise convergence to zero of the difference $v_\alpha(x) - v_{n,\alpha}(x)$ is well-known (e.g. Shapiro (1968), Corollary 3; Sennott (1999)). Here if $(d_i^*)^\infty$ is optimal over an interval in the canonical partition, we are interested in investigating the uniform convergence (in α) to zero of $v_{n,\alpha}^{d_i^*}(x) - v_{n,\alpha}(x)$ over this (or a proper subset of this) interval. We use the following criterion for uniform convergence:

DEFINITION 5. A sequence of non-negative valued functions $\{f_n, n \geq 1\}$ converges uniformly to 0 on a set E if and only if $\lim_{n\to\infty} \sup_{x\in E} f_n(x) = 0$.

REMARK 3. It is a standard fact (cf. Equation (1) on p. 257, Goldberg (1976)) that $f_n \to 0$ uniformly on E if and only if $\lim_{n\to\infty} \sup_{x\in E} |f_n(x)| = 0$. Our definition follows from this fact, since we impose the constraint that $f_n(x)$ be non-negative valued for all n, for all $x \in E$.

PROPOSITION 1. *Let $(d_i^*)^\infty$ be the stationary infinite horizon optimal policy on I_i, where I_i is an interval in the canonical partition. For any $x \in \mathbb{X}$ and $[\alpha_0, \tilde{\alpha}] \subset I_i$, each of the following families of functions*

(a) $\left\{v_\alpha(x) - v_{n,\alpha}^{d_i^*}(x), n \geq 0\right\}$, *and*
(b) $\left\{G_{n,\alpha}^{d_i^*}(x), n \geq 0\right\} := \left\{v_{n,\alpha}(x) - v_{n,\alpha}^{d_i^*}(x), n \geq 0\right\}$
converge to the zero on the interval $[\alpha_0, \tilde{\alpha}]$ *uniformly in* α, *as* $n \to \infty$.

Proof. Recall that $v_\alpha(x)$ is a power series in α (see Sennott (1999), page 70). In our case, the power series converges for $\alpha \in [0,1)$. It is a standard result that a power series converges uniformly on every closed subinterval inside its radius of convergence. Since $(d_i^*)^\infty$ is optimal for all $\alpha \in [\alpha_0, \tilde{\alpha}]$, we have that $v_{n,\alpha}^{d_i^*}(x)$ is the n-th partial sum of $v_\alpha(x)$. Thus, the family of functions

$$\left\{v_\alpha(x) - v_{n,\alpha}^{d_i^*}(x), n \geq 0\right\}$$

converges uniformly to the zero function over $[\alpha_0, \tilde{\alpha}]$, establishing (a). Note that

$$v_\alpha(x) - v_{n,\alpha}^{d_i^*}(x) \geq G_{n,\alpha}^{d_i^*}(x) \geq 0. \tag{5}$$

Part (b) follows from part (a) and the inequality above. ∎

We are now in a position to prove our first major result.

THEOREM 1. *For any* $[a,b] \subset I_i$ *where* I_i *is an interval in the canonical partition not containing any degenerate points, the set of turnpike integers* $\{N(\alpha) : \alpha \in [a,b]\}$ *is finite.*

Proof. We prove the result for a compact subset of the Blackwell interval $[\alpha_\infty, 1)$ with corresponding Blackwell optimal policy $(d^*)^\infty$. Let α_0 and $\tilde{\alpha}$ be as in the proof of Proposition 1. Define

$$F^n([\alpha_0, \tilde{\alpha}]) := \{f : \ v_{n,\alpha}(x) = r(x, f(x)) + \alpha \sum_y p(y|x, f(x)) v_{n-1,\alpha}(y), x \in \mathbb{X}, \text{ for some } \alpha \in [\alpha_0, \tilde{\alpha}]\}$$

Note that for each integer n each decision rule in $F^n([\alpha_0, \tilde{\alpha}])$ is an optimal first period decision rule for the n-horizon problem for some $\alpha \in [\alpha_0, \tilde{\alpha}]$. The finite state and action space assumptions guarantee that $F^n([\alpha_0, \tilde{\alpha}])$ is non-empty for each n. Similarly, define

$$F^\infty([\alpha_0, \tilde{\alpha}]) := \{f : \ v_\alpha(x) = r(x, f(x)) + \alpha \sum_y p(y|x, f(x)) v_\alpha(y), x \in \mathbb{X}, \text{ for some } \alpha \in [\alpha_0, \tilde{\alpha}]\}.$$

The elements of $F^\infty([\alpha_0, \tilde{\alpha}])$ define all the stationary infinite horizon policies that are optimal at some discount rate α, as α ranges over the interval $[\alpha_0, \tilde{\alpha}]$. To prove the result we first show that there exists N^* such that for all $n \geq N^*$ we have

$$F^n([\alpha_0, \tilde{\alpha}]) \subseteq F^\infty([\alpha_0, \tilde{\alpha}]). \tag{6}$$

Suppose no such N^* exists. Then we claim that it follows that there exists a decision rule $g \notin F^\infty([\alpha_0, \tilde{\alpha}])$, an infinite sequence $\{n_i, i \geq 1\}$, and an infinite sequence of discount rates $\{\alpha_i, i \geq 1\}$ such that $\alpha_i \in [\alpha_0, \tilde{\alpha}]$ such that

$$v_{n_i,\alpha_i}(x) = r(x, g(x)) + \alpha_i \sum_y p(y|x, g(x)) v_{n-1,\alpha_i}(x). \tag{7}$$

To see this, note that if N^* does not exist there exists infinitely many horizon lengths $n_1, n_2, \ldots$ such that for all n_i, $F^{n_i}([\alpha_0, \tilde{\alpha}])$ is not contained in $F^\infty([\alpha_0, \tilde{\alpha}])$. That is to say that for each n_i we can pick an initial decision rule g_i that is n_i-horizon optimal for some $\alpha_i \in [\alpha_0, \tilde{\alpha}]$ such that $(g_i)^\infty$ is not infinite horizon optimal for any $\alpha \in [\alpha_0, \tilde{\alpha}]$. Since there are only finitely many initial decision rules, this is equivalent to the statement that for each n_i we can pick an initial decision

rule g and a discount rate $\alpha_i \in [\alpha_0, \tilde{\alpha}]$ such that g is not an element of $F^\infty([\alpha_0, \tilde{\alpha}])$. This is precisely the content of the claim associated with (7).

Since $\{\alpha_i, i \geq 1\}$ is a bounded sequence, there exists a convergent subsequence $\{\alpha_{i_j}, j \geq 1\}$ converging to say $\alpha^* \in [\alpha_0, \tilde{\alpha}]$. There is no loss of generality (and a gain in notational clarity) in assuming that the sequence $\{\alpha_i, i \geq 1\}$ converges to $\alpha^* \in [\alpha_0, \tilde{\alpha}]$. Letting $i \to \infty$ in (7) we get

$$\lim_{i\to\infty} v_{n_i,\alpha_i}(x) = r(x, g(x)) + \lim_{i\to\infty} \alpha_i \sum_y p(y|x, g(x)) v_{n_i-1,\alpha_i}(y) \tag{8}$$

Since $\{v_{n_i,\alpha_i}(x), i \geq 1\}$ is a bounded set, it has at least one limit point, so the limit in (8) can be taken along a suitable subsequence. Again we take the limit along the entire sequence for simplicity.

We claim that $\lim_{n_i\to\infty, \alpha_i\to\alpha^*} v_{n_i,\alpha_i}(x)$ exists and is equal to the repeated limit $\lim_{n_i\to\infty} \lim_{\alpha_i\to\alpha^*} v_{n_i,\alpha_i}(x)$. To see this, note from Proposition 1 that

$$\lim_{i\to\infty} \sup_{\alpha\in[\alpha_0,\tilde{\alpha}]} [v_\alpha(x) - v_{n_i,\alpha}(x)] = 0.$$

That is, the sequence $\{v_\alpha(x) - v_{n_i,\alpha}(x), i \geq 1\}$ converges uniformly (to zero) for all $\alpha \in [\alpha_0, \tilde{\alpha}]$ as $n_i \to \infty$. It is a classic result in the theory of double sequences and double series that this uniform convergence implies that the double limit exists and is equal to each of the repeated limits. Therefore we have that

$$\lim_{i\to\infty} [v_{\alpha_i}(x) - v_{n_i,\alpha_i}(x)] = \lim_{n_i\to\infty} \lim_{\alpha_i\to\alpha^*} [v_{\alpha_i}(x) - v_{n_i,\alpha_i}(x)] = \lim_{n_i\to\infty} [v_{\alpha^*}(x) - v_{n_i,\alpha^*}(x)] = 0.$$

This yields

$$\lim_{i\to\infty} v_{n_i,\alpha_i}(x) = v_{\alpha^*}(x).$$

Therefore, for each $x \in \mathbb{X}$

$$v_{\alpha^*}(x) = r(x, g(x)) + \alpha^* \sum_y p(y|x, g(x)) v_{\alpha^*}(y).$$

This implies that g^∞ is infinite horizon optimal at discount rate α^*; a contradiction.

Consider again fixed $\alpha \in [\alpha_0, \tilde{\alpha}]$. It follows from (6) that for some N^* we can let f be a first period optimal decision rule for the n-horizon α-discounted problem for all $n \geq N^*$ and that f^∞ is an optimal stationary policy for the infinite horizon problem *for some* (non-empty) subset of discount rates in $[\alpha_0, \tilde{\alpha}]$. Define this subset by $S(\alpha)$ which a priori may or may not contain α. To complete the proof of the theorem we establish that $\alpha \in S(\alpha)$. Since α is an arbitrary element of $[\alpha_0, \tilde{\alpha}]$, the theorem then follows from this fact and (6) since N^* is a turnpike integer for each $\alpha \in [\alpha_0, \tilde{\alpha}]$ and upper bounds $N(\alpha)$ for all $\alpha \in [\alpha_0, \tilde{\alpha}]$.

To establish that $\alpha \in S(\alpha)$, suppose f^∞ is the unique stationary infinite horizon optimal policy for all discount rates in $[\alpha_0, \tilde{\alpha}]$. Thus, (6) implies that f is a first period optimal decision rule for the n-step α-discounted problem for all $n \geq N^*$ and all $\alpha \in [\alpha_0, \tilde{\alpha}]$. Therefore $[\alpha_0, \tilde{\alpha}] = S(\alpha)$ in this case. Next, suppose f^∞ and g^∞ are two distinct infinite horizon optimal policies for *all* discount rates in $[\alpha_0, \tilde{\alpha}]$. Note by the definition of a canonical partition interval that it is not possible for there to be two distinct infinite horizon optimal policies for a proper subset of a canonical partition interval. Fix $\alpha' \in [\alpha_0, \tilde{\alpha}]$. Suppose f is *not* a first period optimal decision rule for the n-step α'-discounted problem for all $n \geq N^*$. This implies that g is a first period optimal decision rule for the n-step α'-discounted problem for all $n \geq N^*$ (otherwise there is a third distinct infinite-horizon optimal policy at α'). Therefore $\alpha' \in S(\alpha')$. The same reasoning applies if there are more than two distinct infinite horizon optimal policies for *all* discount rates in $[\alpha_0, \tilde{\alpha}]$. The theorem follows. ∎

REMARK 4. We note in Theorem 1 that we have explicitly chosen an interval that does not contain any break or degenerate points. This implies that there exists a policy that is optimal throughout this interval. Violations of this assumption require special treatment; see Theorems 4 and 5

We conclude this section by showing that any proper subset of an element of the canonical partition not containing a degenerate point can be partitioned into a finite number of turnpike intervals. To ease the notation, let $w_{n,\alpha}(h,x)$ be the optimal value of the n-horizon MDP with discount factor α that fixes the initial decision rule h and follows an optimal policy thereafter.

THEOREM 2. *Suppose $[a,b] \subset I_i$ is a compact (strict) subset of a set in the canonical partition not containing any degenerate points. Then $[a,b]$ can be partitioned into finitely many turnpike intervals.*

Proof. First, we show that $[a,b]$ can be partitioned into intervals, each associated with a fixed finite turnpike integer. Fix $\alpha_0 \in [a,b]$ and $x \in \mathbb{X}$. Let h^∞ be the unique (up to equivalence) stationary infinite horizon optimal policy for all $\alpha \in [a,b]$ (see Remark 2). Recall that the turnpike integer $N(\alpha_0)$ is finite, and that the value function $v_{n,\alpha}(x)$ is a continuous function of α. We claim that there exists $\epsilon > 0$ and n^* (dependent on ϵ) such that for all $n \geq n^*$, h is an optimal first period decision rule for the n-period problem for all discount rates in the open interval $Q_{n^*} := (\alpha_0 - \epsilon, \alpha_0 + \epsilon)$.

We establish this claim by contradiction. Fix $\epsilon > 0$. Consider any strictly decreasing sequence $\{\epsilon_i, i \geq 0\}$ such that $\epsilon_i > 0$ and $\epsilon_i \downarrow 0$. Suppose there exists i^* is such that $i \geq i^*$ implies h is not an optimal first period decision rule for all discount rates in $(\alpha_0 - \epsilon_i, \alpha_0)$. Defining $\alpha_i = \alpha_0 - \epsilon_i$ yields that $\{\alpha_i, i \geq i^*\}$ is a strictly increasing sequence such that $\alpha_i \uparrow \alpha_0$. Fix arbitrary n. For $i \geq i^*$ there exists and $g_i \neq h$ such that $\{g_i, i \geq i^*\}$ is a sequence of optimal initial decision rules for the n-horizon problem at discount rate α_i. Note that there are only finitely many distinct g_i; therefore there exists $g \neq h$ such that g is an optimal initial decision for the n-horizon problem at each discount rate in a subsequence of $\{\alpha_i, i \geq i^*\}$ converging to α_0. Thus, $w_{n,\alpha}(g,x)$ and $w_{n,\alpha}(h,x)$ are both continuous functions of α such that $w_{n,\alpha}(h,x) < w_{n,\alpha}(g,x)$ at infinitely many points α in $(\alpha_0 - \epsilon, \alpha_0)$. In addition, at α_0 we have $w_{n,\alpha_0}(h,x) > w_{n,\alpha_0}(g,x)$. This implies that the graphs of $w_{n,\alpha}(h,x)$ and $w_{n,\alpha}(g,x)$ (regarded as functions of α) intersect infinitely many times in a left-neighborhood of α_0. This contradicts the fact that they are both rational functions. Since ϵ and n are arbitrary, this proves our claim for a left-neighborhood of α_0. The proof for a right-neighborhood follows in the same manner.

Repeating this procedure for each $\alpha \in [a,b]$, we obtain a family of open intervals covering $[a,b]$. By the Heine-Borel theorem, we can extract a finite subcover. Note that each interval I in the subcover is associated with a fixed integer $n^*(I)$. This completes the proof of the theorem in case there is a unique stationary infinite horizon optimal policy for all $\alpha \in [a,b]$. If there are K distinct stationary infinite policies, we simply repeat the foregoing argument K times, and the same conclusion holds. This completes the proof of the theorem. ∎

Fix $[A,B] \subset [0,1)$. Consider the set of turnpike integers

$$C_{AB} := \{N(\alpha) : \alpha \in [A,B] \subset [0,1)\}.$$

The proof of the following theorem follows in the same manner as that of Theorem 2.

THEOREM 3. *Fix $[A,B] \subset [0,1)$. Suppose the set of turnpike integers C_{AB} is bounded. There exists a finite multi-set of integers $\{N_0, ..., N_r\}$ such that N_i is a turnpike integer for all discount factors in some interval $(\alpha_i^L, \alpha_i^U]$ and $[A,B] \subset \bigcup_{i=0}^r (\alpha_i^L, \alpha_i^U]$.*

We remark that $\{N_0, ..., N_r\}$ is a multi-set because its elements are not necessarily distinct; the same turnpike integer may characterize non-contiguous intervals. We characterize when C_{AB} is bounded in the next section.

3.2. Turnpike integers on the boundary of the canonical partition. Again, consider Example 1. Recall, Section 3.1 explains the finiteness of turnpike integers and intervals on the interior of each interval in the canonical partition. Action b is infinite horizon optimal at discount rate $\alpha = 0.7 + \epsilon$ (where $.3 > \epsilon > 0$). According to (4) action b becomes n-horizon optimal at $\alpha = 0.7 + \epsilon$ when

$$\frac{(0.7+\epsilon)(1-(0.7+\epsilon)^n)}{1-0.7-\epsilon} > \frac{7}{3}.$$

This inequality simplifies to

$$(0.7+\epsilon)^n < 1 - \left(\frac{7}{3}\right)\left(\frac{0.3-\epsilon}{0.7+\epsilon}\right). \tag{9}$$

Define $N_1(\epsilon)$ as the smallest integer such that

$$(0.7)^n < 1 - \left(\frac{7}{3}\right)\left(\frac{0.3-\epsilon}{0.7+\epsilon}\right) \tag{10}$$

holds, and define $N_2(\epsilon)$ as the smallest integer such that

$$(0.7+\epsilon)^n < 1 - \left(\frac{7}{3}\right)\left(\frac{0.3-\epsilon}{0.7+\epsilon}\right). \tag{11}$$

holds. It is clear that $N_2(\epsilon) \geq N_1(\epsilon)$ for every $\epsilon > 0$. Note that $1 - \left(\frac{7}{3}\right)\left(\frac{0.3-\epsilon}{0.7+\epsilon}\right) \downarrow 0$ as $\epsilon \downarrow 0$, and therefore $N_1(\epsilon) \to \infty$ as $\epsilon \downarrow 0$. Since $N_2(\epsilon) \geq N_1(\epsilon)$, we have that $N_2(\epsilon) \to \infty$ as $\epsilon \downarrow 0$. We have shown that the set of turnpike integers in the neighborhood of 0.7 is unbounded.

The next theorem characterizes the conditions under which the set of turnpike integers is unbounded. This is perhaps the most novel result in this paper; the existence of discount rates bounded away from unity that are associated with arbitrarily large turnpike integers is not hinted at in past research. The theorem shows that the turnpike integer of an interval containing the endpoint of a sub-interval in the canonical partition is infinity, for all but some exceptional finite Markov decision processes. Before stating the theorem we require a lemma. Recall the definition of $w_{n,\alpha}(f,x)$ immediately preceding Theorem 2 and let I be an interval bounded away from $\alpha = 1$, containing exactly one break point a_i of the canonical partition, and no degenerate points. Consider **Condition C** below.

Condition C: There exists $M < \infty$ such that $w_{n,a_i}(f,x) = w_{n,a_i}(h,x)$ for all $n \geq M$ and for all $x \in \mathbb{X}$.

LEMMA 2. *Satisfying **Condition C** is equivalent to the condition that $v_{n,a_i} = v_{a_i}$ for all $n \geq M'$, for some positive integer M'.*

Proof. Writing the Bellman recursion, we see that **Condition C** holds if and only if there exists M such that for all $n \geq M$ and for all $x \in \mathbb{X}$ we have

$$\begin{aligned} w_{n+1,a_i}(f,x) &= r(x,f) + a_i \sum_{y\in\mathbb{X}} p(y|x,f(x)) v_{n,a_i}(y) \\ w_{n+1,a_i}(h,x) &= r(x,h) + a_i \sum_{y\in\mathbb{X}} p(y|x,h(x)) v_{n,a_i}(y) \\ w_{n+1,a_i}(f,x) &= w_{n+1,a_i}(h,x). \end{aligned}$$

We can rewrite this system of equations as

$$r(x,f) - r(x,h) = a_i \sum_{y \in \mathbb{X}} [p(y|x,f(x)) - p(y|x,h(x))] v_{n,a_i}(y) \tag{12}$$

for all $n \geq M$, for each $x \in \mathbb{X}$.

Notice that (12) is, for each n, a system of $|\mathbb{X}|$ linear equations in $|\mathbb{X}|$ variables (the variables being $\{v_{n,a_i}(y), y \in \mathbb{X}\}$). Now for each y we have that $\{v_{n,a_i}(y), n \geq 0\}$ is an infinite sequence of numbers converging to the optimal value function $v_{a_i}(y)$. Consider the (typical) case when for each y the sequence $\{v_{n,a_i}(y), n \geq 0\}$ contains a strictly increasing subsequence. Under **Condition C**, (12) requires a finite number of linear equations (the finite number being $|\mathbb{X}|$) to have infinitely many solutions, which is impossible. The only exception is when there $v_{n,a_i}(y)$ remains at a fixed value for all $n \geq M' \geq M$ for each $y \in \mathbb{X}$ for some finite M'. Since $\lim_{n\to\infty} v_{n,a_i}(y) = v_{a_i}(y)$, this fixed value has to be $v_{a_i}(y)$. That is to say, satisfying **Condition C** is equivalent to the condition that $v_{n,a_i} = v_{a_i}$ for all $n \geq M'$, for some positive integer M' as desired. ∎

THEOREM 4. *Let I be an interval bounded away from $\alpha = 1$, containing exactly one break point a_i of the canonical partition, and no degenerate points. The turnpike integer on I is finite if and only if value iteration has finite convergence at a_i. That is, $N(I) < \infty$ if and only if there exists $M < \infty$ such that for all $x \in \mathbb{X}$, $v_{n,a_i}(x) = v_{a_i}(x)$ for all $n \geq M$.*

Proof. Suppose f^∞ and h^∞ are both infinite horizon optimal policies at a_i (one can check that the ensuing argument remains valid even if there are more than two infinite horizon optimal policies at a_i). Without loss of generality assume f is optimal for the canonical partition interval immediately to the left of a_i and h is optimal immediately to the right. We have that $a_i \in I$.

We claim that **Condition C** is necessary and sufficient for $N(I)$ to be finite. First we show necessity. Suppose **Condition C** does not hold. This means that for all M there is at least one $x \in \mathbb{X}$ where $w_{n,a_i}(f,x) \neq w_{n,a_i}(h,x)$ holds for some $n \geq M$. Thus one of the following cases hold:

(a) the inequality $w_{n,a_i}(f,x) > w_{n,a_i}(h,x)$ holds for infinitely many n, and the reverse inequality $w_{n,a_i}(f,x) < w_{n,a_i}(h,x)$ holds for infinitely many n as well;
(b) $w_{n,a_i}(f,x) > w_{n,a_i}(h,x)$ for all sufficiently large n;
(c) $w_{n,a_i}(f,x) < w_{n,a_i}(h,x)$ for all sufficiently large n.

Consider (a). Suppose $w_{n,a_i}(f,x) > w_{n,a_i}(h,x)$ for infinitely many n. For any such n, continuity implies there exists $\epsilon_n > 0$ such that $w_{n,\alpha}(f,x) > w_{n,\alpha}(h,x)$ for all $\alpha \in (a_i, a_i + \epsilon_n) \subset I$. Noting that h^∞ is infinite-horizon optimal for all discount rates in $[a_i, a_i + \epsilon_n)$, and that f^∞ is not infinite-horizon optimal in this interval of discount rates, implies that $N([a_i, a_i + \epsilon_n)) > n$. Since this holds for infinitely many n, it follows that there exists $\epsilon > 0$ such that $N([a_i, a_i + \epsilon)) = \infty$. Similarly, if $w_{n,a_i}(h,x) > w_{n,a_i}(f,x)$ then there exists $\epsilon_n > 0$ where the analogous inequality holds for $\alpha \in [a_i, a_i - \epsilon_n)$. Using similar logic it follows that for some $\epsilon > 0$, $N([a_i, a_i - \epsilon)) = \infty$. So we have that in Case (a) $N(I) = \infty$. The same argument shows that in Case (b) we have $N([a_i, a_i + \epsilon)) = \infty$ and in Case (c) we have $N([a_i, a_i - \epsilon)) = \infty$ for sufficiently all small $\epsilon > 0$. Thus **Condition C** is necessary. We note that Example 1 (at the point $\alpha = a_i = 0.7$) exemplifies Case (b).

Next we show that **Condition C** is sufficient for $N(I)$ to be finite. Suppose **Condition C C** holds. Then there exists $M < \infty$ such that $w_{n,a_i}(f,x) = w_{n,a_i}(h,x)$ for all $n \geq M$ and for all $x \in \mathbb{X}$. Without loss of generality, we take $M \geq N(a_i)$, the turnpike integer at a_i. Following the decision rule f at each of the first m steps of a $(M+m)$-horizon problem is optimal for all positive integers

m. Using Lemma 2, **Condition C** implies that there exists a positive integer $M' \geq M$ such that $v_{n,a_i} = v_{a_i}$ for all $n \geq M'$. Therefore a policy that is optimal for the n-horizon problem at discount rate a_i (say π^*_n) is n-horizon optimal at discount rate a_i for all $n \geq M'$, as well as infinite horizon optimal. In fact note that we need only ensure that we have followed $\pi^*_{M'}$ for the first M' periods to reach the optimal expected value of the infinite horizon problem.

Now consider any n-horizon problem such that $n \geq M' + M$ with the discount rate is fixed at a_i. From the preceding facts, the n-horizon policy of following $\pi^*_{M'}$ in the first M' periods and any sequence of decision rules f from period $M'+1$ onwards is optimal for any finite horizon problem with $M'+1$ or more periods. At the same time, following decision rule f in the first M' periods of the n-period horizon is optimal for the n-horizon problem all $n \geq M' + M$ (noting that we have ensured that M, and hence $M' + M$, is greater than the turnpike integer at a_i and that $w_{n,a_i}(f,x) = w_{n,a_i}(h,x)$ for all $n \geq M$). Therefore, using f $n-$times is an optimal policy for the n-horizon problem for all $n \geq M + M'$. Using f^∞ is optimal for the infinite horizon problem as well. Likewise, using h, $n-$times is an optimal policy for the n-horizon problem for all $n \geq M + M'$, and h^∞ is optimal for the infinite horizon problem as well.

Now we have by the continuity of value functions that h^∞ is infinite horizon optimal in $[a_i, a_i + \epsilon)$ for all sufficiently small $\epsilon > 0$, and that f^∞ is infinite-horizon optimal in $(a_i - \epsilon, a_i]$ for all sufficiently small $\epsilon > 0$. Moreover, since for all $n \geq M + M'$ we have already have $v_{n,a_i} = v_{a_i}$, we have that using h, $n-$times is $n-$horizon optimal in $[a_i, a_i + \epsilon)$ for all sufficiently small $\epsilon > 0$, while using f $n-$times is not $n-$horizon optimal in this neighborhood. Similarly, for all $n \geq M + M'$ we have that using f $n-$times is n-horizon optimal in $(a_i - \epsilon, a_i]$ for all sufficiently small $\epsilon > 0$ while using h $n-$times is not $n-$horizon optimal in this neighborhood. We claim that this implies that the turnpike integer associated with $(a_i - \epsilon, a_i + \epsilon)$ for all sufficiently small $\epsilon > 0$ is finite. We prove this by contradiction.

First we show that the turnpike integer $[a_i, a_i + \epsilon)$ is finite, for any sufficiently small $\epsilon > 0$. The proof for $(a_i - \epsilon, a_i]$ is similar. Suppose not. This implies there exists a decreasing sequence of discount rates $r_k \downarrow a_i$ such that $N(r_k) \to \infty$ as $k \to \infty$. In this case, there exists an initial decision rule d (where $d \neq h, d \neq f$) and a sequence of integers $n_k \to \infty$ such that d is an optimal initial decision rule for the n_k-horizon problem at discount rate r_k. Now repeating the argument of Theorem 1 (in particular, the steps from (7) onwards), we see that this implies that d^∞ is infinite-horizon optimal for some $\alpha \in [a_i, a_i + \epsilon]$. This is impossible, and we have a contradiction. This completes the proof of the sufficiency clause of our theorem is complete. ∎

An analogous result holds for turnpike integers in the neighborhood of a degenerate point; the proof is almost identical to that of Theorem 4.

THEOREM 5. *Let I be an open sub-interval of a partition interval bounded away from break points and let f^∞ be the unique stationary infinite-horizon policy associated with this partition interval. Suppose I contains a degenerate point a^* such that h^∞ is infinite-horizon optimal at a^*. Then $N(I) < \infty$ if and only if value iteration has finite convergence at a^*.*

Proof. The proof of the statement that if value iteration does not have finite convergence at a^*, then $N(I) = \infty$ is identical to that in the proof of Theorem 4.

To show sufficiency, suppose value iteration has finite convergence at a^*. By repeating the steps in the proof of Theorem 4, we arrive at the finding that using f, $n-$times is an optimal policy for the n-horizon problem for all $n \geq M + M'$, and it is optimal for the infinite horizon problem as well. Likewise, using h $n-$times is an optimal policy for the n-horizon problem for all $n \geq M + M'$, and it is optimal for the infinite horizon problem as well. By the continuity of value functions and the definition of a degenerate point, f^∞ is infinite horizon optimal in an ϵ-neighborhood of a_i for all sufficiently small $\epsilon > 0$, and therefore using f $n-$times is n-horizon optimal for all

$n \geq M + M'$ in an ϵ-neighborhood of a^* for all sufficiently small $\epsilon > 0$. At the same time note that h^∞ is not infinite horizon optimal in this neighborhood, which implies that for all $n \geq M + M'$ using h $n-$times is optimal at discount rate a^* and sub-optimal in an ϵ-neighborhood of a^* for all sufficiently small $\epsilon > 0$. Repeating the last part of the proof of Theorem 4, we see that this implies that the turnpike integer associated with an ϵ-neighborhood of a^* for all sufficiently small $\epsilon > 0$ is finite. This completes the proof of sufficiency. ∎

3.2.1. Discussion The results of Theorems 1-5 require some comments. Consider the discount factor as it varies from 0 to 1 and recall the canonical decomposition partitions this interval for the infinite horizon problem in such a way that the infinite horizon optimal policies change along this partition. First, Theorem 1 explains that any compact set that does not contain an endpoint of the canonical decomposition or a degenerate point is associated with a finite turnpike integer. Second, Theorems 2 and 3 explain that any compact set whose turnpike integers are bounded can be decomposed into a finite number of turnpike intervals, where a turnpike interval is an interval of discount rates all of which have exactly the same turnpike integer. Lastly, we explain in Theorem 4 and Theorem 5 that the only places where the turnpike integer can approach infinity is at the endpoints of the canonical partition intervals and at a degenerate point. Regarding this last point, note that it is quite exceptional to **not** have a discount with a neighborhood that has an infinite turnpike integer.

We may give an intuitive explanation for the explosion of the turnpike integers characterizing the neighborhood of a break-point (say α_0) as follows. At a break-point, we are at a transition between the zone of optimality of at least two distinct stationary infinite horizon policies (say h^∞ and g^∞). A stationary policy h^∞ may be optimal at α_0 while h may yet be a sub-optimal initial decision for infinitely many n. However, for every value of n at which h is not an optimal initial decision for the n-horizon problem, g is an optimal initial decision for the n-horizon problem, and hence the turnpike integer at α_0 is finite. However, in (either) an $\epsilon-$ right or left neighborhood of α_0, g^∞ is not infinite-horizon optimal at any of the discount rates in the neighborhood. To summarize, infinite turnpike integers characterize the neighborhood of a discount rate at which there is a transition from one infinite horizon policy to another, and in addition there are infinitely many distinct finite horizon problems for which a fixed infinite horizon optimal policy fails to supply the optimal initial decision. Consider the following example.

EXAMPLE 2. Suppose there are 3 states; $\mathbb{X} = \{1, 2, 3\}$ and with only one action in states 2 and 3 and two actions in state 1. The rest of the problem data follows:

$$c(1,b) = 3/4 \qquad c(1,c) = 1/2$$
$$c(2,a_2) = 1 \qquad c(3,a_3) = 0$$

$$p(2|1,b) = 1/4 \qquad p(3|1,b) = 3/4$$
$$p(2|1,c) = 1$$
$$p(3|2,a_2) = p \qquad p(2|2,a_2) = 1-p$$
$$p(3|3,a_3) = 1.$$

See Figure 2. A simple calculation yields $v_\alpha(3) = 0$. We also have

$$v_\alpha(2) = 1 + \alpha\big((1-p)v_\alpha(2)\big)$$

A little algebra yields

$$v_\alpha(2) = \frac{1}{1-\alpha(1-p)}.$$

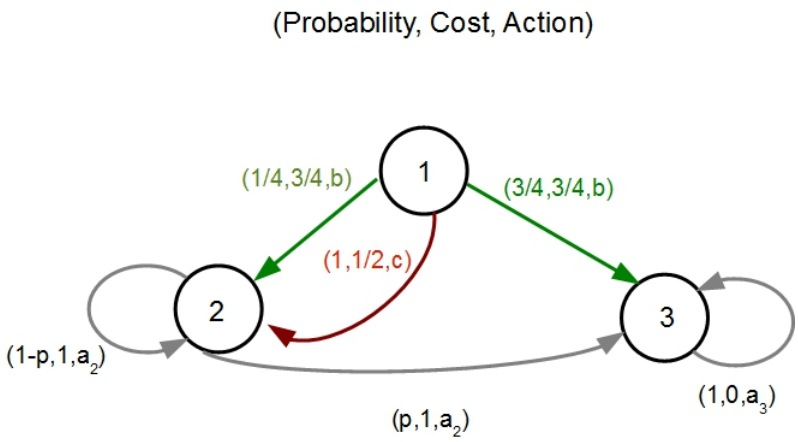

FIGURE 3. Example 2 data.

Further,

$$v_\alpha(1) = \min\left\{\overbrace{\frac{3}{4}+\frac{\alpha}{4}v_\alpha(2)}^{\text{use action b}}, \overbrace{\frac{1}{2}+\alpha v_\alpha(2)}^{\text{use action c}}\right\}$$
$$= \min\left\{\frac{1}{4}+\frac{\alpha}{4}v_\alpha(2), \alpha v_\alpha(2)\right\}+\frac{1}{2}.$$

Thus, we should choose action b to minimize expected cost over the infinite horizon if

$$\begin{aligned}\frac{1}{4}+\frac{\alpha}{4}v_\alpha(2) &\le \alpha v_\alpha(2)\\ \Leftrightarrow \frac{1}{3} &\le \alpha v_\alpha(2) \Leftrightarrow \frac{1}{3} \le \frac{\alpha}{1-\alpha(1-p)}\\ \Leftrightarrow 1 &\le (4-p)\alpha\end{aligned} \tag{13}$$

Case 1: Suppose first that $p=1$. In this case $\alpha \ge 1/3$ implies action b is optimal for the infinite horizon problem. Note that $v_{n,\alpha}(2)=1$ for $n \ge 1$. Since the process reaches state 3 either after 1 or 2 periods and since there is zero cost associated with state 3, a little algebra yields that $\alpha \ge 1/3$ implies that action b is optimal for the n-horizon problem, for all $n \ge 2$. In this case, the turnpike integer of a right-neighborhood of $\alpha = 1/3$ is equal to 2.

Case 2: Now suppose $p=0$. Consider the n-horizon value functions. It is straightforward to check that

$$v_{n,\alpha}(2) = \frac{1-\alpha^n}{1-\alpha}$$

and hence

$$v_{n+1,\alpha}(1) = \min\left\{\overbrace{\frac{3}{4}+\frac{\alpha}{4}v_{n,\alpha}(2)}^{\text{use action b}}, \overbrace{\frac{1}{2}+\alpha v_{n,\alpha}(2)}^{\text{use action c}}\right\}$$
$$= \min\left\{\frac{1}{4}+\frac{\alpha}{4}v_{n,\alpha}(2), \alpha v_{n,\alpha}(2)\right\}+\frac{1}{2}.$$

Action c is optimal at state 1 for the $(n+1)$-horizon problem if and only if

$$\alpha\Big(\frac{1-\alpha^n}{1-\alpha}\Big) \leq \frac{1}{4} + \frac{\alpha}{4}\Big(\frac{1-\alpha^n}{1-\alpha}\Big)$$

$$\Leftrightarrow \frac{3\alpha}{4}\Big(\frac{1-\alpha^n}{1-\alpha}\Big) \leq \frac{1}{4}. \tag{14}$$

Consider the discount rate $\alpha = 1/4$. At this discount rate, it follows from the preceding inequalities that action c is optimal at state 1 for the n-horizon problem if and only if

$$[1-(1/4)^n] \leq 1$$

which holds for all integers $n \geq 1$. This implies that action c is optimal for the n-horizon problem for all $n \geq 1$, at $\alpha = 1/4$. Note from (13) that action b is optimal for the infinite horizon problem for all discount rates $\alpha \geq 1/4$. Since **Condition C** is violated, we expect the turnpike integer of any right-neighborhood of $\alpha = 1/4$ to be ∞ (as per case (b) in the proof of the necessity of **Condition C** in Theorem 4). We explicitly verify this. Let $\epsilon > 0$. It follows from (14) that action b is optimal at $\alpha = 1/4 + \epsilon$ when

$$\frac{3}{4}\Big(\frac{1}{4}+\epsilon\Big)\Big(\frac{1-(\frac{1}{4}+\epsilon)^n}{1-(\frac{1}{4}+\epsilon)}\Big) > \frac{1}{4}$$

$$\Leftrightarrow \Big(\frac{1}{4}+\epsilon\Big)\Big(1-(\frac{1}{4}+\epsilon)^n\Big) > \frac{1}{3}\Big(1-(\frac{1}{4}+\epsilon)\Big)$$

$$\Leftrightarrow \Big(\frac{1}{4}+\epsilon\Big) - \Big(\frac{1}{4}+\epsilon\Big)^{n+1} > \frac{1}{3}\Big(1-(\frac{1}{4}+\epsilon)\Big) = \frac{1}{3}\Big(\frac{3}{4}-\epsilon\Big) = \frac{1}{4} - \frac{\epsilon}{3}$$

$$\Leftrightarrow \frac{4\epsilon}{3} > \Big(\frac{1}{4}+\epsilon\Big)^{n+1}.$$

Taking logs yields,

$$(n+1)\log\Big(\frac{1}{4}+\epsilon\Big) < \log\Big(\frac{4\epsilon}{3}\Big).$$

As $\epsilon \downarrow 0$ the right hand side goes to $-\infty$, while the left hand side goes to $(n+1)\log(.25) \approx -.6(n+1)$. That is to say, when ϵ is small, we need n large to reach the optimal stationary policy's decision rule; $N\big([\frac{1}{4}, \frac{1}{4}+\epsilon]\big) = \infty$. This is due to the fact that **Condition C** is not satisfied.

Case 3: Let $p = 1/2$. In this case $\alpha \geq 2/7$ implies action b is infinite horizon optimal. More importantly, in the finite horizon version of this problem, the time spent in state 2 is the minimum of a geometric random variable (parameter p) and the remaining time in the horizon. At $\alpha = 2/(7+.000001)$ value iteration converges to the optimal policy after 1 iteration, while $\alpha = 2/(7-.000001)$ takes 8 steps (12 steps when $\alpha = 2/(7-.000000001)$). The implication is that from below $\alpha = 2/7$ the turnpike integers are bounded, but from above $\alpha = 2/7$ the set of turnpike integers is unbounded.

3.3. Unbounded turnpike integers: practical implications We have shown that the typical MDP contains singularities, discount rates in the neighborhood of which the turnpike integer tends to infinity. We discuss the implications of this result. Note that in Table 1 for the discount factors up to and including 0.7 the finite horizon problem has a myopic optimal solution; the optimal solution is simply to repeat the optimal solution to the infinite horizon problem in

every period of a finite horizon. Suppose the discount rate is in fact just greater than 0.7 and is rounded (perhaps by an algorithm) to 0.7. We claim that this apparently innocuous rounding can lead to substantial errors in following a turnpike planning horizon procedure.

We have shown that discount factors sufficiently close to and greater than 0.7 are associated with arbitrarily large turnpike integers. If we round such discount factors to 0.7, the optimal policy for an arbitrary finite horizon is to follow the infinite horizon optimal policy in every period. Suppose the initial state is 0. The optimal policy for arbitrary finite horizon lengths is to follow action b in every period. On the other hand, we know that since the set of turnpike integers in an arbitrarily small right-neighborhood of 0.7 is unbounded, a better policy is to follow action c in every period for a very large number of periods. Let us compare the performance of these two distinct policies over a horizon of 5 periods. Following the detailed calculations previously carried out for Example 1, we see that the difference in expected cost between the myopic optimal policy and the true optimal policy over n periods (with initial state 0) is

$$f(n,\alpha) := \frac{7}{8} + \frac{\alpha(1-\alpha^{n+1})}{2(1-\alpha)} - \frac{7\alpha(1-\alpha^{n+1})}{8(1-\alpha)}.$$

A little calculation shows that using the wrong turnpike integer in this case implies an escalation in expected cost of between 29 percent and 5.4 percent of the optimal cost; the smaller the finite horizon the larger the impact (please see Table 2 below).

Number of periods in the horizon	**Percentage escalation in expected cost**
1	29.17
2	18.28
3	11.93
4	7.97
5	5.4

TABLE 2. Impact of using the wrong turnpike integer.

Past research has formulated efficient procedures for computing the canonical partition and the associated optimal policies. Smallwood (1966) gives the details of an $O(n^3)$ algorithm for computing the canonical partition. Further, Hordijk et al. (1985) develop a linear programming procedure for the computation of optimal policies over the entire range of the discount factor. Theorem 3 highlights the importance of incorporating these procedures into computational stochastic dynamic programming routines.

4. Convergence and turnpike integers over a subset of the Blackwell interval Recall that for a Blackwell optimal policy $(d^*)^\infty$ we defined $G_{n,\alpha}(x) = v_{n,\alpha}(x) - v^{d^*}_{n,\alpha}(x)$ and $[\alpha_\infty, 1)$ to be the Blackwell interval. It follows from Proposition 1 that $G_{n,\alpha}(x) = v_{n,\alpha}(x) - v^{d^*}_{n,\alpha}(x)$ converges to zero uniformly in α for $\alpha \in I$, where I is strictly contained in the Blackwell interval. However, because the value functions may approach ∞ as α approaches 1, studying the behavior of $G_{n,\alpha}(x)$ over an interval $[a,1)$, where $\alpha_\infty < a < 1$, involves examining the difference between two unbounded functions over a non-compact interval. Since $G_{n,\alpha}(x)$ is for every n the sum of the first n terms of a geometric series in α, it cannot in general be expected to converge uniformly to the zero function over the Blackwell interval (note that a geometric series in α converges uniformly only over a closed subinterval within its radius of convergence). An exception to this is when $G_{n,\alpha}(x)$ happens to be identically equal to the zero function for all $n \geq K$, for some finite K, as was the case in Example 1. We see from Table 1 that for all discount rates in $[0.88, 1)$, action c is optimal for the n-horizon problem for $n = 1, 2, 3$ whereas action b is optimal for all $n \geq 4$ (in state zero).

The Blackwell interval is $[0.7, 1)$, so action b is optimal for the infinite horizon for all discount rates in $[0.88, 1)$, and this policy is also n-horizon optimal for all discount rates in $[0.88, 1)$, for $n \geq 4$. So in this case $G_{n,\alpha}(x) = 0$ for all $\alpha \in [0.88, 1)$, for $n \geq 4$.

5. Discussion and Conclusion This study has uncovered several insights but leaves some questions unanswered. First, it follows from our results that while there are only finitely many intervals in the canonical partition, there may be infinitely many turnpike intervals; this happens precisely under the condition spelled out in Theorem 4. This in and of itself is an interesting finding. In addition to this, one might have conjectured that $N(\alpha)$ is a non-decreasing function of α. After all, as the discount rate increases, we give more weight to future periods in the horizon. However, notice from Table 1 from Example 1, $N(\alpha)$ is not a monotone function of α. This leads one to conjecture that any intuition about $N(\alpha)$ may actually hold only within sets of the canonical partition.

Infinite horizon formulations are an approximation to problems with long time horizons. They allow us to restrict attention to stationary policies and value functions independent of the current time point. Suppose a firm would like to implement an inventory control policy and approximates the decision scenario with an infinite time horizon. They could write the dynamic programming formulation with all of the estimated parameters including the discount factor. One would expect that the time horizon that needs to hold for this infinite horizon approximation to be valid is the same for discount factors that are close together. Example 1 in the preceding subsection demonstrates that turnpike intervals may in fact be very narrow. The practical import of this observation is that if a firm uses N-period rolling horizon policies, then the appropriate value of N may be quite sensitive to the discount rate. Carefully pinpointing an accurate value of discount rate is well worth the effort in this context.

In practice, there is often some ambiguity about the exact discount rate to use. This is particularly true for projects with a long time horizon, when the discount rate may change over time to reflect variations in the economic climate. A discount rate fixed at a certain point in time reflects expectations about the future which may be revised in light of events as they unfold, and these events may themselves be hard to predict accurately. Faced with a rolling horizon problem in which the discount rate is uncertain, suppose we are able to characterize our uncertainty by a discrete distribution r_i with probability p_i for $i = 1, \ldots, M$.

Consider the following alternative procedures:

A) Roll our uncertainty about discount rate into a point estimate - say the expected value, as per the discrete distribution - and proceed to find an optimal rolling horizon policy based on this point estimate.
B) Develop a discrete set of feasible discount rates, and for each discount rate locate the appropriate turnpike integer. Create a rolling horizon plan based on the maximum of the turnpike integers of the turnpike intervals corresponding to the possible discrete distribution.

Example 1 suggests that Procedure B may be well worth the extra expense in computing time, compared with Procedure A which is quicker but may lead to a bad solution. The difference between these two procedures may be particularly significant if the time horizon is large, and if the cash flows are large. It may be unwise to base a rolling horizon plan on a single point estimate of discount rate when there is inherent uncertainty about the rate itself. Rather, we recommend distilling our uncertainty into a small set of possible values of discount rate, locating these discount rates within their turnpike intervals, and using the maximum turnpike integer as the basis of a rolling horizon plan.

We suggest a study of turnpike integers for Markov decision problems with infinite state and action spaces as an interesting avenue for future work.

References

Goldberg, Richard G. 1976. *Methods of Real Analysis*. 2nd ed. Wiley. 8

Hordijk, A., R. Dekker, L.C.M. Kallenberg. 1985. Sensitivity-analysis in discounted markovian decision problems. *OR Spektrum* **7** 143–151. 18

Kallenberg, Lodewijk. 2009. Markov decision processes. http://www.math.leidenuniv.nl/~kallenberg/Lecture-notes-MDP.pdf. 2, 3

Khan, M. Ali, Adriana Piazza. 2011. An overview of turnpike theory: Towards the discounted deterministic case. *Advances on Mathematical Economics*, vol. 14. Springer, 39–67. 5

Puterman, Martin L. 1994. *Markov decision processes: Discrete stochastic dynamic programming*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York. 2

Sennott, Linn I. 1999. *Stochastic Dynamic Programming and the Control of Queueing Systems*. Wiley Series in Probability and Statistics, Wiley, New York, NY. 3, 8, 9

Shapiro, Jeremy F. 1968. Turnpike planning horizons for a Markovian decision model. *Management Science* **14**(5) 292–300. 1, 2, 4, 8

Smallwood, Richard D. 1966. Optimum policy regions for markov processes with discounting. *Operations Research* **14**(4) 658–669. 18

Zaslavski, Alexander J. 2014. *Stability of the turnpike phenomenon in discrete-time optimal control problems*. Springer. 5