

# Conditions for the discovery of solution horizons

James C. Bean and Robert L. Smith\*

*Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, USA*

Received 28 September 1990

Revised manuscript received 21 May 1991

We present necessary and sufficient conditions for discrete infinite horizon optimization problems with unique solutions to be solvable. These problems can be equivalently viewed as the task of finding a shortest path in an infinite directed network. We provide general forward algorithms with stopping rules for their solution. The key condition required is that of weak reachability, which roughly requires that for any sequence of nodes or states, it must be possible from optimal states to reach states close in cost to states along this sequence. Moreover the costs to reach these states must converge to zero. Applications are considered in optimal search, undiscounted Markov decision processes, and deterministic infinite horizon optimization.

*Key words:* Dynamic programming, deterministic and Markov, infinite state: infinite stage problems, programming, infinite dimensional: infinite horizon optimization, shortest paths, infinite networks.

## Introduction

Many important planning problems, such as capacity expansion, equipment replacement and production planning, involve sequences of related decisions over an indefinite time horizon. The mathematical formulation and solution of such problems is called infinite horizon optimization. A common approach to solving these problems involves truncating the infinite horizon problem to a finite horizon problem. This approximation may lead to error, although it may be possible to establish a bound on this error as a function of the horizon employed (see for example, [12] and [1]). In some cases, if the finite horizon is sufficiently distant, the error can be isolated to later decisions. In this way optimal early decisions are obtained and can be implemented without error. At the end of the time interval covered by these early decisions, the problem is solved again for the next set of decisions.

A horizon time sufficiently long to eliminate error from at least the first decision, for that horizon or longer, is called a *solution horizon* [6]. To find solution horizons,

*Correspondence to:* J.C. Bean and R.L. Smith, Department of Industrial and Operations Engineering, The University of Michigan, IOE Building, 1205 Beal Avenue, Ann Arbor, MI 48109, USA.

\* This work was supported in part by NSF Grant ECS-8700836 to The University of Michigan.

practitioners commonly employ *forward algorithms*. Such techniques solve finite horizon problems of increasing horizon length. A *stopping rule* indicates that they have gone far enough to ensure that the first decisions can be implemented without error. We typically must solve beyond the solution horizon to determine that a shorter horizon would have sufficed. The longest horizon actually solved in this search is called a *forecast horizon* since the planner must forecast data to this time.

Not all problems have solution horizons. As we increase the horizons, if two different initial decisions indefinitely interchange as finite horizon optimal, and both are optimal in the infinite horizon model, no solution horizon exists. This is not a problem theoretically since we do not care which one we choose. In practice, however, it is difficult to differentiate this case from the case where only one of the two is optimal (over an infinite horizon), but we have not looked far enough out in time to determine this.

Historically, papers considering solution horizon approaches have ignored this problem, reporting algorithms that are not guaranteed to stop in finite time. The classic example is the Wagner-Whitin algorithm with stopping rule [19]. Recently, more attention has been paid to this issue. In [3], solution horizon existence is shown for a general class of deterministic discounted problems under the condition of a unique optimum. In [7] and [10], these results are extended to discounted and weakly ergodic stochastic models, respectively.

However, existence of a solution horizon alone is not sufficient to guarantee existence of a finite algorithm. It must be shown, given that a solution horizon exists, that a stopping rule exists that is certain to recognize it. Most forward algorithms in the literature fail to satisfy this property. Exceptions include [3-5, 8, 11, 15, 17].

As is common with an emerging literature, the arguments supporting the existence results for each special problem have been as varied as the problems themselves. Reference [17] relies on algebraic arguments, [3] on discounting, [8] on regeneration sets, [11] on (strong) reachability, and [10] on weak ergodicity of Markov processes.

This paper seeks an underlying explanation for the existence of solution horizons, common to most of these approaches, where explicit or implicit discounting is present. It presents a general sequential decision problem framework and also unifies many of the stopping rules now found in the literature. Included within this framework are stochastic as well as deterministic problems. In the presence of a unique infinite horizon optimum, we state conditions that are both necessary and sufficient for forward procedures to converge to the optimal solution. We also provide a general forward algorithm with a stopping rule that, under these conditions, finds the optimal first decision in finite time.

Section 2 introduces the notion of weak reachability. Weak reachability is then shown to be a necessary and sufficient condition for solution horizon discovery. Section 2 also presents a general forward algorithm and stopping rule for the discovery of solution horizons in the presence of weak reachability. Section 3 establishes weak reachability for a number of general problems. New algorithms

are proposed for discrete optimal search and Markov decision processes that are guaranteed to finitely terminate. Section 4 contains a summary and conclusions.

## 1. Existence theory

### 1.1. Problem statement

Consider the following general infinite horizon sequential decision problem. We begin with a countably infinite directed graph,  $(\mathcal{N}, \mathcal{A})$ , with single root node of in-degree zero together with a real-valued cost function  $C: \mathcal{A} \rightarrow \mathbb{R}$ . We refer to  $(\mathcal{N}, \mathcal{A}, C)$  as the *decision network*, arcs  $(i, j) \in \mathcal{A}$  as *decisions*, nodes  $i \in \mathcal{N}$  as *states*, and arc costs  $C(i, j)$  as *decision costs*. The decisions could represent, for example, order quantities or capacity additions, and would induce states representing inventory levels and excess capacities over time, respectively.

We impose two structural assumptions on the decision network,  $(\mathcal{N}, \mathcal{A}, C)$ . First, we assume that the set of decisions available at any node is non-empty and *finite*, i.e., all node out-degrees are non-zero and finite. Second, we assume that there are at most a finite number of preceding decisions associated with every state, i.e., that the nodes have finite predecessor sets. More precisely, we assume that all nodes have *finite* cumulative in-degrees where the cumulative in-degree associated with a node is the sum of the in-degrees associated with that node and all nodes from which there is a directed path to that node. From the second assumption, we can number the nodes  $\mathcal{N} = \{0, 1, 2, \dots\}$  such that  $(i, j) \in \mathcal{A}$  only if  $i < j$  [18, p. 230]. Hence, the node numbers can serve as a surrogate for time. Also since there is a unique root node by assumption, it follows that there is a directed path from the root node to each node in  $\mathcal{N}$ . Moreover since all nodes have non-zero out-degree, each path can be feasibly continued to form a path covering the infinite horizon.

An (infinite) *path* is an infinite sequence of states  $(s_0, s_1, s_2, \dots)$  where  $s_0$  is the root node and  $(s_i, s_{i+1}) \in \mathcal{A}$  for all  $i = 0, 1, 2, \dots$ . We shall also refer to  $(s_i, s_{i+1}, \dots, s_j)$  as a path from  $s_i$  to  $s_j$ . The cost  $C(s_i, s_{i+1})$  associated with the decision  $(s_i, s_{i+1})$  will also be referred to as the *length* of arc  $(s_i, s_{i+1})$ . The length of the path  $(s_0, s_1, s_2, \dots)$  is then given by  $\sum_{i=0}^{\infty} C(s_i, s_{i+1})$ . Following the notation in [3], define a strategy  $\pi = (\pi_1, \pi_2, \dots)$  as the infinite sequence of decisions  $\pi_i$  along a path  $\{s_i\}_{i=0}^{\infty}$  where  $\pi_i = (s_{i-1}, s_i)$ . The set of all feasible strategies is denoted  $\Pi$ . The cost,  $f_\pi$ , of strategy  $\pi$  is then the length of the path corresponding to  $\pi$ , i.e.,  $f_\pi = \sum_{i=0}^{\infty} C(\pi_i)$ . Our problem is to find an optimal strategy,  $\pi^*$ , which minimizes  $f_\pi$  over  $\pi \in \Pi$ .

Drawing on the analogy relating node numbers to time,  $\pi^*$  will be referred to as the optimal infinite horizon strategy. Let  $f$  be the length of a shortest path covering the infinite horizon so that  $f = f_{\pi^*}$ . Define the optimal value functions  $f(s)$  and  $f'(s)$  as the shortest distance from the root node to node  $s$  and the shortest distance from node  $s$  through the infinite horizon, respectively. Let  $f(|s)$  be the length of a shortest path from the root node, covering the infinite horizon, which passes through node  $s$ .

By the principle of optimality, we have  $f(|s) = f(s) + f'(s)$  for all  $s \in \mathcal{N}$ . Note that some of these values will not be well defined in general. Sufficient conditions for their existence will now be established.

In [3], a metric topology on  $\Pi$  is constructed which metrizes the product topology of componentwise convergence. Specifically, define the metric,  $\rho$ , as

$$\rho(\pi, \pi') = \sum_{i=1}^{\infty} \phi_i(\pi, \pi')2^{-i},$$

where

$$\phi_i(\pi, \pi') = \begin{cases} 0, & \text{if } \pi_i = \pi'_i \\ 1, & \text{otherwise} \end{cases}.$$

Under this topology, a sequence,  $\{\pi^n\}$  in  $\Pi$ , converges to  $\pi$  in  $\Pi$  if and only if its components,  $\{\pi_i^n\}$ , converge to  $\pi_i$  for all  $i$ . That is,  $\pi^n \rightarrow \pi$  as  $n \rightarrow \infty$  if and only if, for all  $i$ , there is an  $N_i$ , such that  $\pi_i^n = \pi_i$  for  $n \geq N_i$ . The metric is motivated by the fact that under this metric the closeness of two strategies is measured by the number of initial decisions over which they agree. This topology is extended to a general class of stochastic problems in [7]. For further details the reader is referred to these papers.

**Lemma 1.**  *$\Pi$  is complete and, hence, compact.*

**Proof.**  $\Pi$  is complete if every Cauchy sequence of paths converges to a feasible path. Consider any Cauchy sequence of paths  $\{\pi^n\}$ . Take  $\varepsilon = 2^{-i}$ ,  $i \geq 1$ . Then for each  $i \geq 1$ , there exists an integer  $N(i)$ , such that for all  $n, m \geq N(i)$ ,  $\rho(\pi^n, \pi^m) < 2^{-i}$  implying that  $\pi_l = \pi_l^n = \pi_l^m$  for all  $l = 1, 2, \dots, i$ . Hence  $\pi^n \rightarrow \pi$  as  $n \rightarrow \infty$ . Moreover one inductively infers that  $\pi$  is a feasible path in the network, i.e.,  $\pi \in \Pi$ . Hence  $\Pi$  is complete. Compactness then follows from completeness and the fact that  $\Pi$  is totally bounded since all out-degrees are finite.  $\square$

We make the following assumption about  $C$ .

**Assumption 1.**  $f_\pi$  is uniformly convergent over  $\pi \in \Pi$ .

Assumption 1 states that for all  $\varepsilon > 0$ , and all  $\pi \in \Pi$ , there exists  $N_\varepsilon$  that is independent of  $\pi$ , such that  $\sum_{i=n}^{\infty} |C(s_i, s_{i+1})| < \varepsilon$ , for all  $n \geq N_\varepsilon$ , where  $\{s_i\}$  is the state path corresponding to  $\pi$ . It is equivalent to say that  $f(s_n)$  converges uniformly over all feasible paths,  $\{s_n\}_{n=0}^{\infty}$ . A sufficient condition is for example that decision costs  $C(s_i, s_{i+1})$  are discounted by a factor  $\alpha^i$  where  $0 < \alpha < 1$  and the undiscounted costs are uniformly bounded.

**Lemma 2.**  $f_\pi$  is a continuous function of  $\pi \in \Pi$ .

**Proof.** We have by definition and Assumption 1 that  $f_\pi = \lim_{n \rightarrow \infty} \sum_{i=1}^n C(\pi_i)$  is the limit of a uniformly convergent sequence of continuous functions and is therefore continuous.  $\square$

**Theorem 3.** Under Assumption 1,  $f$  exists and is attained by some strategy  $\pi^*$ .

**Proof.** A continuous function on a compact set attains its minimum.  $\square$

## 2. Weak reachability

Sufficient conditions for the existence of solution horizons have only recently appeared in the literature. In [3], a general deterministic infinite horizon problem is formulated that relies on discounting for existence. It is assumed that at each time the optimal strategy through that time is found, and proven that, in the absence of multiple infinite horizon optima, these finite horizon optima converge to the infinite horizon optimum. In [7], this same technique is used in a stochastic problem. In [11], a concave cost inventory problem is considered. It is assumed that any state is (strongly) reachable from any other state (in the sense of [13], see below), and it is shown that  $f(s_n) \rightarrow f$  for any sequence of states  $s_0 < s_1 < s_2 \cdots <$  which diverges to infinity. Nonhomogeneous Markov decision processes are considered in [10] where solution horizons are shown to exist in the absence of discounting, provided the transition matrix of every policy is weakly ergodic (see [16]) and provided a unique infinite horizon optimum exists. In [20], an infinite search problem is proposed and existence of solution horizons is established. In this case the result follows from finiteness of the expected time to capture.

In these papers, different problem characteristics are exploited to show existence of solution horizons. We seek, in this paper, to unify these results by showing that in each case a single necessary and sufficient condition for solution horizon existence must hold. We call this key condition weak reachability.

To understand weak reachability, we begin by contrasting it to the traditional notion of strong reachability. If a sequence of states is strongly reachable, then, eventually, there is a connecting path from some state on an optimal path, to each state in the sequence, at a cost converging to zero. Qualitatively, we are requiring that our sequence of states not be so far off that it is isolated from all optimal solutions.

This notion is too restrictive for general infinite horizon optimization. For example, if the underlying graph is a tree, once you leave the optimal path, there is no longer any chance of reconnecting. Weak reachability relaxes the requirement so that we need not get from a state on the optimal path to each state in our sequence, but only to some other state that is close to the state in our sequence, where close means close in optimal costs through the states.

**Definition** (weak reachability). The sequence of states  $\{s_n\}_{n=1}^\infty$  with  $s_n \rightarrow \infty$  is called *weakly reachable* if there are associated sequences of states  $\{u_n\}_{n=1}^\infty$  and  $\{t_n\}_{n=1}^\infty$  such that for every  $\varepsilon > 0$ , the following three conditions are satisfied for all sufficiently large  $n$ :

$$f(s_n) < f(t_n) + \varepsilon, \tag{1}$$

$$c(u_n, t_n) < \varepsilon, \tag{2}$$

$$f(u_n) < f + \varepsilon, \tag{3}$$

where  $c(u_n, t_n)$  is the cost of a path from  $u_n$  to  $t_n$ ,  $n = 1, 2, \dots$ . (For maximization problems, the inequalities are reversed in (1) and (3) with  $\varepsilon$  replaced by  $-\varepsilon$ .)

Figure 1 depicts the intent of this condition. Roughly speaking, the first condition requires that the cost of the shortest path to  $s_n$  be no more than  $\varepsilon$  greater than the cost of the shortest path to  $t_n$ . The second two conditions require the state  $t_n$  be reachable at cost  $\varepsilon$  from a state  $u_n$  that itself is within  $\varepsilon$  of being infinite horizon optimal. Perhaps the most natural choice for  $\{u_n\}$  is a sequence of infinite horizon optimal states  $\{s_n^*\}$  with  $s_n^* \rightarrow \infty$  since condition (3) then follows automatically from  $f(s_n^*) \rightarrow f$  as  $n \rightarrow \infty$  (see below).

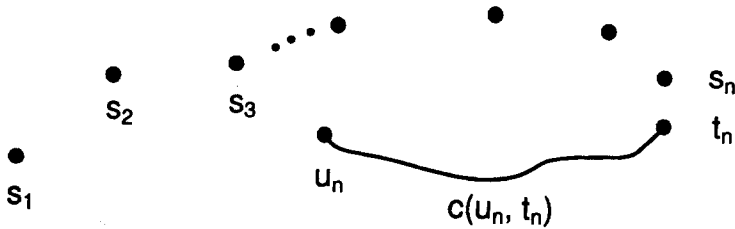


Fig. 1. Illustration of weak reachability.

**2.1. Value and policy convergence**

Consider the sequence  $\{s_n^*\}$  of states passed through by an optimal strategy  $\pi^*$ . Since  $f(s_n^*) + f'(s_n^*) = f$  by the principle of optimality and  $f'(s_n^*) \rightarrow 0$  by Assumption 1, we have

$$f(s_n^*) \rightarrow f \text{ as } n \rightarrow \infty.$$

We extend this value convergence property to arbitrary state sequences by showing that weak reachability is necessary as well as sufficient for value convergence to hold.

**Theorem 4** (value convergence). *Under Assumption 1, for any sequence of states  $\{s_n\}_{n=0}^\infty$  such that  $s_n \rightarrow \infty$ ,*

$$f(s_n) \rightarrow f \text{ if and only if } \{s_n\}_{n=0}^\infty \text{ is weakly reachable.}$$

**Proof.** *Sufficiency:* Assume  $\{s_n\}$  is weakly reachable and that  $s_n \rightarrow \infty$ . Since  $f(|s_n) = f(s_n) + f'(s_n)$  and  $\lim_n f'(s_n) = 0$  by Assumption 1 and the fact that  $s_n \rightarrow \infty$ , we have  $\lim_n f(s_n) = f$  if and only if  $\lim_n f(|s_n) = f$ . Pick any  $\varepsilon > 0$ . Since  $s_n \rightarrow \infty$ , Assumption 1 implies that  $f'(s_n) < \varepsilon$  for all sufficiently large  $n$ . This and weak reachability guarantee that, for all sufficiently large  $n$ ,

$$\begin{aligned} f(|s_n) &= f(s_n) + f'(s_n) \\ &< f(s_n) + \varepsilon \\ &< f(t_n) + 2\varepsilon \\ &\leq f(u_n) + c(u_n, t_n) + 2\varepsilon \\ &< f + 4\varepsilon. \end{aligned}$$

Since the above inequality holds for all  $\varepsilon > 0$ , we conclude that  $\limsup_{n \rightarrow \infty} f(|s_n) \leq f$ . It remains to show that  $\lim_{n \rightarrow \infty} f(|s_n) = f$ . If not, then there exists a subsequence  $\{s_{n_k}\}$  for which  $\limsup_{k \rightarrow \infty} f(|s_{n_k}) < f$ . By the compactness of  $\Pi$ , we may pass to a subsequence if necessary to get  $\lim_{k \rightarrow \infty} \pi_{n_k} = \pi$  for some strategy  $\pi \in \Pi$ . By Lemma 2,  $\lim_{k \rightarrow \infty} f(\pi_{n_k}) = f(\pi) < f$  contradicting the optimality of  $f$ .

*Necessity:* Suppose  $\{s_n\}$  satisfies  $f(s_n) \rightarrow f$  and  $s_n \rightarrow \infty$ . For each  $n$ , let  $t_n = s_n$  so that (1) is trivially satisfied. Let  $(u_n, s_n)$  be the final arc on a path to node  $s_n$  whose length equals  $f(s_n)$ . By the finite out-degree assumption, only finitely many arcs emanate from each node so it must be that  $u_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Pick any  $\varepsilon > 0$ . Assumption 1 then guarantees that  $c(u_n, t_n) = C(u_n, s_n) < \varepsilon$  for all sufficiently large  $n$ , so that (2) is satisfied. Finally, by hypothesis,  $f(u_n) + C(u_n, s_n) = f(s_n) \rightarrow f$  and  $C(u_n, s_n) \rightarrow 0$  so that (3) holds. Hence, weak reachability is satisfied.  $\square$

**Remark.** From the proof above, we also have that weak reachability is a necessary and sufficient condition for  $f(|s_n) \rightarrow f$  as  $n \rightarrow \infty$ .

With an additional uniqueness assumption we obtain the far stronger result that the optimal strategies to a sequence of weakly reachable states converge to the infinite horizon optimal strategy.

**Assumption 2.** The optimal infinite horizon strategy,  $\pi^*$ , is unique.

Although it is hard to construct (discounted) problems for which Assumption 2 fails to hold, sufficient conditions for it to hold are difficult to come by. It is shown in [3] that if discounting is present and there are at most countably many potentially optimal strategies, then Assumption 2 holds for almost all interest rates. More generally, a tie-breaking rule is provided in [15] that forces policy convergence when Assumption 2 fails to hold.

**Corollary 5** (policy convergence). *Under Assumptions 1 and 2,  $\pi^*(s_n) \rightarrow \pi^*$  as  $n \rightarrow \infty$  if and only if  $\{s_n\}$  is weakly reachable, where  $\pi^*(s_n)$  is any strategy that attains  $f(s_n)$ .*

**Proof.** If  $\pi^*(s_n)$  does not converge to  $\pi^*$ , then there is an infinite subsequence bounded away from  $\pi^*$ . This, in turn, has a convergent infinite subsequence. By Theorem 4 the optimal values of these strategies to these states converge to the infinite horizon optimal value. Hence, by Assumption 1 and the continuity of  $f_\pi$ , the limit of this subsequence is also infinite horizon optimal. This contradicts Assumption 2. The only if direction is omitted.  $\square$

Under the  $\rho$ -metric,  $\pi^*(s_n) \rightarrow \pi^*$  as  $n \rightarrow \infty$  implies that there exists an  $N^*$  such that for all  $n \geq N^*$ ,  $\pi_1^*(s_n) = \pi_1^*$ . The value  $N^*$  plays the role of a solution horizon here, although  $s_n$  is *not* restricted here to be an optimal finite horizon state. Hence, a first decision on an optimal path to  $s_n$ , for sufficiently large  $n$ , will also be optimal for the infinite horizon problem. The next section gives a stopping rule to numerically establish how large  $n$  must be. By fixing the first decision to be  $\pi_1^*$ , we can repeat this procedure to establish  $\pi_2^*$ , and in this fashion recursively recover  $\pi^*$ . It is interesting to note that policy convergence takes place here in the absence of terminal values to “steer” the finite horizon solutions.

## 2.2. General forward algorithm

Define a sequence of *stopping sets* as a sequence of finite sets of states,  $\{S_i\}_{i=0}^\infty$ , such that if  $\{s_i\}_{i=0}^\infty \in \times_{i=0}^\infty S_i$  then  $\lim_{i \rightarrow \infty} s_i = \infty$ .

We propose the following general forward algorithm. The algorithm is an abstraction of many problem specific algorithms in the literature including [2, 3, 5, 8, 11].

### Forward Algorithm.

*Step 1.* Set  $i = 1$ .

*Step 2.* Solve for an optimal strategy,  $\pi^*(s)$ , to all states  $s \in S_i$ . If  $\pi_1^*(s)$  is invariant over all  $s \in S_i$ , stop. Otherwise, set  $i = i + 1$  and go to Step 2.

**Theorem 6.** *Under Assumptions 1 and 2, the Forward Algorithm is guaranteed to stop at some finite time  $N^*$  with  $\pi_1^*(s) = \pi_1^*$  for all  $s \in S_{N^*}$  if:*

- (1) *Every sequence of states in  $\times_{i=0}^\infty S_i$  is weakly reachable.*
- (2) *Every stopping set,  $S_i$ , contains an infinite horizon optimal state.*

**Proof.** We first show that the forward algorithm must finitely terminate. Suppose not. Then there exists  $s'_i \in S_i$  for all  $i$  such that  $\pi_1^*(s'_i) \neq \pi_1^*$  for all  $i$ . But  $\{s'_i\}_{i=0}^\infty \in \times_{i=0}^\infty S_i$  and hence is weakly reachable by (1). Then by Corollary 5, we have  $\lim_i \pi_1^*(s'_i) = \pi_1^*$ . Contradiction. When the algorithm does stop, at  $N^*$  say, then by (2) and the principle of optimality, we have that  $\pi_1^*(s) = \pi_1^*(s^*) = \pi_1^*$  for all  $s \in S_{N^*}$  where  $s^*$  is an infinite horizon optimal state in  $S_{N^*}$ .  $\square$



**Remark.** Even in the absence of a unique infinite horizon optimum, if the Forward Algorithm stops, then the invariant first decision obtained must be infinite horizon optimal.

Under conditions (1) and (2) of Theorem 6, the Forward Algorithm will finitely terminate with an invariant optimal first decision that is optimal for the infinite horizon problem. It is interesting to note that  $N^*$  here is a *forecast horizon* in that the infinite horizon optimality of the invariant first decision  $\pi_1^*$  is not dependent upon problem data beyond time  $N^*$ . The difficulty in constructing such a sequence of stopping sets lies in making each sufficiently small to satisfy condition (1) and sufficiently large to satisfy condition (2). The first paper to construct minimal stopping sets for the lot sizing problem is [8]. By doing so, they were able to find the minimum possible stopping time for this specific problem structure.

Blocking sets are a common choice of stopping sets that automatically satisfy condition (2). A stopping set is a *blocking set* if every strategy must pass through it; that is, every strategy must share at least one state with the blocking set.

**Corollary 7.** *If all stopping sets in the Forward Algorithm are blocking sets, then the conclusions of Theorem 6 hold under condition (1) of Theorem 6 and Assumptions 1 and 2.  $\square$*

Though this Corollary is an obvious implication of Theorem 6, it is presented to better show the connections between the general theory and the specific applications from the literature discussed in Section 3. Many previous results are implementations of Corollary 7 rather than the more general Theorem 6.

### 3. Applications

The existence proofs and stopping rules in the literature present a wide variety of problem specific results and algorithms. The key ideas behind these results vary as greatly as the underlying problems. We consider in this section three general classes of problems from the unifying perspective of weak reachability. Two are in the area of stochastic optimization while the third lies in deterministic optimization. One is discounted and two are formally not. The Forward Algorithm of Section 2 is applied to all three problems, resulting in new algorithms for the stochastic problems that are guaranteed to converge in the unique optimum case.

#### 3.1. Discrete optimal search

In [20], a target is considered that randomly moves among the points of a finite lattice  $L \subseteq \{a \in \mathbb{Z}^m \mid -\infty < \underline{a}_i \leq a_i \leq \bar{a}_i < \infty, i = 1, 2, \dots, m\}$  where  $\mathbb{Z}$  is the set of all integers. Let  $X_n$  be the target's position at time  $n$  where  $n = 0, 1, 2, \dots$ . Departing from [20], we assume knowledge of the probability distribution of the stochastic process  $\{X_n, n = 1, 2, \dots\}$ , where  $X_n, n = 1, 2, \dots$ , are independent but not identically

distributed random variables. A search strategy,  $\xi \in \Xi$ , is any sequence of points in  $L$  such that a point differs from its successor point by at most one in each coordinate. We assume  $\xi$  is nonadaptive and the target is nonevasive, so that the distribution of  $\{X_n, n = 1, 2, \dots\}$  does not depend upon  $\xi$ .

Let  $p_n(\xi_n)$  be the probability that the searcher meets the target at time  $n$  under strategy  $\xi$ , i.e., that the target and the search positions are identical at time  $n$ . We suppose that if target and searcher meet at time  $n$ , there is a nonnegative expected reward,  $R_n$ , obtained, where  $\sum_{n=1}^{\infty} R_n < \infty$ . The problem is to find a search strategy,  $\xi$ , that maximizes the expected total reward,  $E[R(\xi)] = \sum_{n=1}^{\infty} R_n p_n(\xi_n)$ . In the special case,  $R_n = \alpha^n$ , where  $0 < \alpha < 1$ , we conjecture that as  $\alpha$  converges to zero, a revenue maximizing search strategy converges to a strategy that minimizes time to capture. Since the feasibility and return associated with decision  $\xi_{n+1}$  depend only on  $\xi_n$ , the nodes or states in the decision network correspond to the number of points visited together with the current lattice point of the searcher,  $\xi_n$ , while arcs correspond to adjacent lattice points that differ from the current lattice point by at most one in each coordinate. Moreover, there are a finite number of points adjacent to any point in the lattice so that all node out-degrees are finite. Moreover all cumulative in-degrees are finite since there are always a finite number  $n$  of initial decisions leading to any state.

**Theorem 8.** *The discrete optimal search model satisfies Assumption 1. Further, all state sequences are weakly reachable.*

**Proof.** *Assumption 1:* Since  $\sum_{i=1}^{\infty} R_i < \infty$ , for all  $\epsilon > 0$  there exists  $N_\epsilon$  such that  $\sum_{i=n}^{\infty} R_i p_i(\xi_i) \leq \sum_{i=n}^{\infty} R_i < \epsilon$  for all  $n \geq N_\epsilon$  and all  $\xi \in \Xi$ .

*Weak reachability:* We establish weak reachability for any sequence of feasible states  $\{n, a^n\}_{n=1}^{\infty}$  with  $a^n \in L$  for all  $n$ . Let  $N = \sum_{i=1}^m (\bar{a}_i - \underline{a}_i)$  where  $m$  is the dimension of the space containing  $L$ . Then for all  $n \geq N$ , there is a path from  $\xi_{n-N+1}^*$  at time  $n - N + 1$  to  $a^n$  at some time  $n' \leq n$  since we have up to  $N - 1$  steps to get there. Having arrived at the point  $a^n$  at time  $n'$ , we can remain there from  $n'$  through  $n$ . Hence, there is also a path from  $\xi_{n-N+1}^*$  at time  $n - N + 1$  to  $a^n$  at time  $n$ . Consider a path,  $\{\xi^n\}$ , that passes through the lattice points  $\xi_{n-N+1}^*$  at time  $n - N + 1$  and  $a^n$  at time  $n$ . Since for all  $n$ ,  $E[R(\xi^n)] = \sum_{i=1}^{\infty} R_i p_i(\xi_i^n) < \sum_{i=1}^{\infty} R_i < \infty$ , we have that the cost of those paths,  $\sum_{i=n-N+1}^n R_i p_i(\xi_i^n) \rightarrow 0$  as  $n \rightarrow \infty$ . Hence, by setting  $t_n = (n, a^n)$  and  $u_n = (n - N + 1, \xi_{n-N+1}^*)$  for all  $n$ , we conclude that  $\{n, a^n\}_{n=1}^{\infty}$  is weakly reachable (in fact, strongly reachable).  $\square$

Consider now the following implementation of the Forward Algorithm of Section 2 for finding the initial move of an optimal search strategy. Find  $\xi^{*n}(x)$  for all feasible  $x \in L$  that solves the problem

$$\begin{aligned} \max \quad & \sum_{i=1}^n R_i p_i(\xi_i) \\ \text{s.t.} \quad & \xi_n = x, \quad \xi \in \Xi. \end{aligned}$$

If  $\xi_1^{*n}(x)$  is invariant over all feasible  $x \in L$ , then stop. Otherwise increment  $n$  and repeat. Since the set of states in  $L$  attainable in  $n$  steps is a blocking set, under the assumption that  $\xi^*$  is unique, we have by Corollary 7 that this algorithm will finitely terminate with  $\xi_1^* = \xi_1^{*n}(x)$  for all feasible  $x \in L$ . By repeating recursively, we can recover  $\xi^*$ .

### 3.2. Deterministic infinite horizon optimization

Consider a general deterministic sequential decision problem with the following special conditions: nonnegative cost flows are sufficiently discounted to ensure that infinite horizon costs are uniformly finite, decisions are made one at a time and we solve the infinite horizon problem by solving increasing time finite horizon problems. For details the reader is referred to [3].

The implicit decision network in [3] is sufficiently general that states are defined by the sequences of decisions leading to them. That is, problem characteristics may be entirely path dependent. In this case states are equivalent to leading segments of strategies, e.g.,  $(\pi_1, \pi_2, \dots, \pi_n)$  defines a state for any  $\pi \in \Pi$  and  $n$ . Each state has an assigned time denoting the time frame covered by the preceding sequence of decisions. Finite out-degrees follows from the assumption of finiteness of each decision set and finite cumulative in-degrees from finiteness of the decision sequence preceding any state.

**Theorem 9.** *The deterministic infinite horizon optimization model satisfies Assumption 1. Further, all sequences of finite horizon optima are weakly reachable.*

**Proof.** *Assumption 1:* In [3], it is assumed that cost functions are eventually bounded by an exponential function, uniformly over  $\Pi$ . More formally, let  $C_\pi(t)$  be the cumulative cost incurred under strategy  $\pi$  through time  $t$ . Then it is assumed that  $C_\pi(t) \leq K e^{\gamma t}$  for some  $K, \gamma > 0$  and all  $t \geq 0$ . It is further assumed that the costs are continuously discounted by a factor  $e^{-rt}$  so as to make infinite horizon costs finite, i.e.,  $r > \gamma$ . From this we can show that the  $f_\pi$  are uniformly convergent and Assumption 1 is satisfied.

*Weak reachability:* Given any finite time  $T$  define  $f(T)$  to be the minimum cost that covers the time frame  $[0, T]$ , that is, the lowest cost of any strategy to any state which has a time identification at or beyond time  $T$ , only including costs incurred up to time  $T$ . Then  $\pi^*(T)$ , an optimal strategy that attains  $f(T)$ , together with  $T$ , defines a finite horizon optimal state. Any sequence of times going to infinity, together with a corresponding  $\pi^*(T)$ , define a state sequence,  $\{s_n\}$ . We will now show that  $\{s_n\}$  is weakly reachable.

Let  $\{s_j^*\}_{j=0}^\infty$  be an infinite horizon optimal path. Then let  $j(n) = \max\{j | s_j^* < s_n\}$ , the last optimal state on that infinite horizon optimal path with index lower than state  $s_n$ . Let  $t_n = s_{j(n)+1}^*$ , the subsequent optimal state, and set  $u_n = s_{j(n)}^*$ . Since  $f(s_{j(n)}^*) \rightarrow f$  as  $n \rightarrow \infty$ , condition (3) holds. Then  $C(s_{j(n)}^*, s_{j(n)+1}^*) \rightarrow 0$  as  $n \rightarrow \infty$  since

$f < \infty$  implying (2). Moreover,  $f(s_n) \leq f(s_{j(n)+1}^*)$  since  $s_n$  is optimal for horizon  $T$  implying (1). Hence weak reachability holds.  $\square$

Since  $\{\pi^*(T), T \geq 0\}$  is weakly reachable, we conclude from Theorem 4 that  $f(T) \rightarrow f$  as  $T \rightarrow \infty$ . If the optimum is unique, the Planning Horizon Theorem of [3] follows immediately from Corollary 5.

### 3.3. Undiscounted Markov decision processes

We consider the homogeneous version of an undiscounted nonhomogeneous Markov decision process (MDP) problem considered in [10]. Every limit point of finite horizon optima is shown to be average optimal. The uniqueness of such a (limit point) average optimal solution was also shown to be a sufficient condition for solution horizon existence. However, no algorithms were provided for its discovery. We will extend their result for the homogeneous case through the device of weak reachability to provide a new finitely terminating algorithm for solving Markov decision process problems.

We begin with some notation. Let  $p_{ij}^k$  be the probability that we next will be in state  $j$  if we adopt action  $k$  in state  $i$  where both the state and action sets are finite. The rewards  $R_i^k$  earned by adopting action  $k$  in state  $i$  are assumed to be uniformly bounded in absolute value by  $\bar{R}$  as in [10]. The problem is to find, if possible, a sequence of policies  $\pi = (\pi_1, \pi_2, \dots)$  that maximizes the average reward per period.

The states we consider for this problem are not the stochastic states of the underlying Markov process. This choice is unsatisfactory since the destination state is not known when given a partial strategy. Instead, we define states, as in Section 3.2, as partial strategies. That is, a state,  $s$ , associated with time  $n$ , is defined by a partial sequence of policies,  $\pi_1, \pi_2, \dots, \pi_n$ . The decision network within this formulation is a tree whose arcs are policies and whose nodes are determined by the sequence of arcs or policies leading to that node. The MDP problem has thus been recast into a problem with *deterministic* states. Finite cumulative in-degrees and out-degrees follows immediately from the tree structure.

The key to existence of solution horizons for undiscounted Markov decision processes is (weak) ergodicity. While there are several measures of ergodicity, the one used in [10] is

$$\beta = \max_k \left\{ 1 - \max_j \left( \min_i p_{ij}^k \right) \right\}.$$

For the undiscounted problem, it is assumed that  $\beta < 1$ . That is, all transition matrices have an ergodicity measure uniformly bounded below 1.

The optimal value of a state, either to or from, would naturally be taken to be the associated expected reward. This is not satisfactory since, in the absence of discounting, the infinite horizon rewards are not certain to converge as required by Assumption 1.

For a homogeneous MDP with  $\beta < 1$ , it is shown in [14] that the undiscounted, average reward problem can be replaced by a discounted problem without loss of optimality. Rewards and the decision structure remain the same. Transition probabilities are altered and future values discounted with factor  $\beta$ . In this discounted problem, the infinite horizon rewards converge and the optimal strategy is stationary. For the following lemma and theorem we work with the deterministic version of this equivalent discounted problem referred to henceforth as the *equivalent* problem.

**Lemma 10.** *For the equivalent problem, for any states  $s$  and  $s'$ , both associated with time  $n$ ,  $|f'(s) - f'(s')| \leq K\beta^n$ , where  $K$  is a constant and  $\beta < 1$ .*

**Proof.** Since rewards are not changed in the transformation, rewards remain bounded by  $\bar{R}$  in the equivalent problem. Hence for any state  $s$  at time  $n$ ,

$$f'(s) \leq \sum_{m=n}^{\infty} \bar{R}\beta^m = \bar{R}\beta^n / (1 - \beta).$$

Letting  $K = 2\bar{R}/(1 - \beta)$ , the result follows immediately.  $\square$

**Theorem 11.** *Suppose  $\beta < 1$ . Then the equivalent problem satisfies Assumption 1. Consider the stopping sets  $\{S_n, n = 0, 1, 2, \dots\}$  where  $S_n$  is the set of all states  $s$  with associated time  $n$  such that  $f(s) \geq f_n - K\beta^n$  where  $f_n$  is the optimal value over all states at time  $n$ . Then (1) every sequence of states in  $\times_{n=0}^{\infty} S_n$  is weakly reachable, and (2) every stopping set  $S_n$  contains an infinite horizon optimal state.*

**Proof.** *Assumption 1:* Since rewards are uniformly bounded, Assumption 1 holds by the same argument given in Theorem 9.

*Weak reachability:* Note that  $f(s) \leq f_n$  for all  $s$  in  $S_n$  and for all  $n$  since we are maximizing reward.

We first show that  $s_n^* \in S_n$  for all  $n$ , where  $\{s_n^*\}$  is a sequence of states along an infinite horizon optimal strategy. If not,  $f(s_n^*) < f_n - K\beta^n$  for some  $n$ . But then by Lemma 10, the path through  $s_n^*$  cannot overtake that defining  $f_n$ , which contradicts the definition of  $s_n^*$ . Hence,  $s_n^* \in S_n$ .

We now show that any sequence  $\{s_n\} \in \times_{n=0}^{\infty} S_n$  is weakly reachable. Given  $\varepsilon > 0$ , let  $N_\varepsilon = \min\{n \mid K\beta^n < \varepsilon\}$ . Now for  $n \geq N_\varepsilon$ , set  $t_n = u_n = s_n^*$ , so that  $c(u_n, t_n) = 0$ . Since  $s_n$  and  $t_n$  are both in  $S_n$  and each is less in value than  $s_n$ , we know that  $|f(s_n) - f(t_n)| < \varepsilon$ . Hence,  $f(s_n) > f(t_n) - \varepsilon$  and  $\{s_n\}$  is weakly reachable (in the maximizing sense).  $\square$

From Theorem 6, if the optimal solution to the equivalent problem is unique, the Forward Algorithm will stop in finite time with an average optimal policy. This is in contrast to conventional value iteration that can only approximate the optimal value and hence optimal policy. Also, the algorithm presented here differs in approach from that presented in [9], which deals with variations in terminal values associated with stochastic ending states.

#### 4. Summary and conclusions

Many papers derive solution horizons and algorithms for their discovery based on special problem structure, including the three applications discussed above. We present a condition that is both necessary and sufficient for the existence of solution horizons that subsumes or extends these previous results. Under uniform finiteness and additivity of the value function, the condition is weak reachability. From this we obtain convergence of the value function. To guarantee policy convergence as well, we assume uniqueness of the infinite horizon optimal solution. A general forward algorithm is provided that is guaranteed to finitely terminate with the infinite horizon optimal initial decision.

#### Acknowledgement

We would like to express our thanks to an anonymous referee for suggestions that considerably simplified the definition of weak reachability.

#### References

- [1] J. Bean, J. Birge and R. Smith, "Aggregation in dynamic programming," *Operations Research* 35 (1987) 215-220.
- [2] J. Bean, J. Lohmann and R. Smith, "A dynamic infinite horizon replacement economy decision model," *The Engineering Economist* 30 (1985) 99-120.
- [3] J. Bean and R. Smith, "Conditions for the existence of planning horizons," *Mathematics of Operations Research* 9 (1984) 391-401.
- [4] J. Bean and R. Smith, "Optimal capacity expansion over an infinite horizon," *Management Science* 31 (1985) 1523-1532.
- [5] J. Bean, R. Smith and C. Yano, "Forecast horizons for the discounted dynamic lot size model allowing speculative motive," *Naval Research Logistics* 34 (1987) 761-774.
- [6] C. Bès and S. Sethi, "Concepts of forecast and decision horizons: applications to dynamic stochastic optimization problems," *Mathematics of Operations Research* 13 (1988) 295-310.
- [7] S. Bhaskaran and S. Sethi, "Conditions for the existence of decision horizons for discounted problems in a stochastic environment: A note," *Operations Research Letters* 4 (1985) 61-65.
- [8] S. Chand and T. Morton, "Minimal forecast horizon procedures for dynamic lot size models," *Naval Research Logistics Quarterly* 33 (1986) 111-122.
- [9] W. Hopp, "Identifying forecast horizons in nonhomogeneous Markov decision processes," *Operations Research* 37 (1989) 339-343.
- [10] W. Hopp, J. Bean and R. Smith, "A new optimality criterion for nonhomogeneous Markov decision processes," *Operations Research* 35 (1987) 875-883.
- [11] J. Lasserre, C. Bès and F. Roubellat, "Detection of planning horizons for the general inventory problem with discounted concave costs," *IEEE Transactions on Automatic Control* 29 (1984) 562-564.
- [12] C. Lee and E. Denardo, "Rolling planning horizons: Error bounds for the dynamic lot size model," *Mathematics of Operations Research* 11 (1986) 423-432.
- [13] L. McKenzie, "Turnpike theory," *Econometrica* 44 (1976) 841-864.
- [14] S. Ross, "Arbitrary state Markov decision processes," *The Annals of Mathematical Statistics* 39 (1968) 2118-2122.
- [15] S. Ryan, J. Bean and R. Smith, "A tie-breaking rule for discrete infinite horizon optimization," *Operations Research* 40 (1992) S117-S126.

- [16] E. Seneta, *Non-negative Matrices and Markov Chains* (Springer, New York, 1981).
- [17] J. F. Shapiro and H.M. Wagner, "A finite renewal algorithm for the knapsack and turnpike models," *Operations Research* 15 (1967) 318-341.
- [18] D.K. Skilton, "Imbedding posets in the integers," *Order* 1 (1985) 229-233.
- [19] H.M. Wagner and T.M. Whitin, "Dynamic version of the economic lot size model," *Management Science* 5 (1958) 89-96.
- [20] J. Wilson, "Approximating an infinite stage search problem with a finite horizon model," *Mathematics of Operations Research* 14 (1989) 433-447.