MATHEMATICS OF OPERATIONS RESEARCH
Vol. 36, No. 1, February 2011, pp. 147–164
ISSN 0364-765X | EISSN 1526-5471 | 11 | 3601 | 0147

DOI 10.1287/moor.1100.0478

Average Optimality in Nonhomogeneous Infinite Horizon Markov Decision Processes

Allise O. Wachs
Integral Concepts, Inc., West Bloomfield, Michigan 48325, allise@integral-concepts.com

Irwin E. Schochetman
Mathematics and Statistics, Oakland University, Rochester, Michigan 48309, schochet@oakland.edu

Robert L. Smith
Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan 48109, rlsmith@umich.edu

We consider a nonhomogeneous stochastic infinite horizon optimization problem whose objective is to minimize the overall average cost per period of an infinite sequence of actions (average optimality). Optimal solutions to such problems will in general be nonstationary. Moreover, a solution that initially makes poor decisions, and then selects wisely thereafter, can be average optimal. However, we seek average optimal solutions with optimal short-term, as well as long-term, behavior. Our approach is to first transform our stochastic problem into one that is deterministic, using the standard device of formulating the problem as one of choosing a sequence of policies, as opposed to actions. Within this deterministic framework, states become probability distributions over the original stochastic states. Then, by weakening the notion of state reachability, and strengthening the notion of efficiency traditionally used in the deterministic framework, we prove that such efficient solutions exist and are average optimal, thus simultaneously exhibiting both optimal long- and short-run behavior. This deterministic view of the property of stochastic ergodicity offers the potential to relax the traditional conditions for average optimality that use coefficients of ergodicity, as well as the opportunity to strengthen the criterion of average optimality through the property of efficiency.

Key words: infinite horizon optimization; average optimal; strongly efficient; near reachability; Markov decision problem; coefficient of ergodicity
MSC2000 subject classification: Primary: 90C40; secondary: 90C39
OR/MS subject classification: Primary: dynamic programming/optimal control: Markov infinite state; secondary: programming: infinite dimensional
History: Received July 7, 2006; revised January 17, 2009, August 10, 2009, November 29, 2010, and December 2, 2010. Published online in *Articles in Advance* February 2, 2011.

1. Introduction. The problem of optimally selecting a sequence of decisions over an infinite horizon is complicated by the need to select criteria for imposing preferences over the collection of associated cost streams. Even in the case in which the infinite stream of cost flows is discounted, the resulting discounted total costs will all be infinite when the costs grow sufficiently fast. In a previous paper, Schochetman and Smith [20] considered the criterion of optimality termed *efficiency* (see Ryan et al. [18]) or sometimes finite optimality (Halkin [11]). A solution is termed efficient if, roughly speaking, it is optimal to each of the states through which it passes. Efficiency avoids being overselective in that the existence of efficient solutions is ensured by mild topological conditions. Nor is it underselective, because the requirement that efficient solutions be optimal to each state constrains prior attained states to be along optimal paths to those states.

For deterministic problems, it was shown in Schochetman and Smith [20] that efficient solutions are average optimal under a state reachability condition. The reachability condition roughly required the existence of decision sequences that eventually reach any feasible state sequence from any given feasible state. Of course, in the stochastic setting, state reachability fails. As we shall see, however, by transforming the problem to a deterministic setting through replacement of actions by policies (see, for example, Bertsekas and Shreve [3] for an early use of this device), one can, under appropriate ergodicity conditions, achieve a type of reachability we term *near reachability*. Within this deterministic framework, the stochastic states are replaced by deterministic states corresponding to probability distributions over the original stochastic states. Near reachability holds when there exist policy sequences that can eventually get arbitrarily close to any feasible state sequence from any given feasible state. Because near reachability is a weakening of the traditional hypothesis in the deterministic setting, one needs to correspondingly strengthen the notion of efficiency. We call this notion *strong efficiency*. It requires that a policy sequence be optimal among all policy sequences "close" to states along its path. We show in §3 that strongly efficient strategies exist and are average optimal under near reachability. The development is extremely general to this point, but the intended principal application area is nonhomogeneous infinite horizon Markov decision process (MDP) problems, which are addressed in §4. Average cost optimality in the homogenous case has been extensively studied (see, for example, Puterman [16], Tijms [22], Federgruen and Tijms [6],

Ross [17], Derman [4]). The traditional approach to establishing existence of an average optimal policy is through an optimality equation that is satisfied by the relative value function under certain ergodicity conditions (see, for example, Puterman [16], Dynkin and Yushkevich [5], Sennott in Feinberg and Shwartz [8]). Although the nonhomogeneous case is formally included within the homogeneous case by using the device of augmenting the state variable with time (see, for example, Bertsekas and Shreve [3]), the resulting homogeneous MDP has a countably infinite state space that can pose severe analytical and algorithmic challenges. We specifically require a uniform bound on the number of states within each period for the nonhomogeneous problem we address in this paper so that this device would yield an MDP problem that would not satisfy our assumptions. Our use of efficiency and reachability properties for such stochastic decision problems affords the opportunity to potentially relax traditional ergodicity conditions through their expression within a purely deterministic framework. We should also note that our approach is restricted to finding optimal average cost policies among the class of all deterministic policies. This restriction can be important because it has been shown that nonrandomized strategies may be outperformed by randomized strategies in the case of the upper limit of average costs (see Dynkin and Yushkevich [5]), whereas in the case of the lower limit of average costs for a fixed initial state, it is sufficient to consider nonrandomized policies (Feinberg [7]). We will return to this point later in the Discussion.

The paper is organized as follows. The general deterministic average cost optimization problem we consider is formally introduced in §2. Section 3 introduces the notions of near reachability and strong efficiency for these problems and shows that every strongly efficient strategy is average optimal in the presence of near reachability. In §4, we illustrate the general theory with our principal application of average cost optimality in nonhomogeneous MDP problems by transforming these into deterministic equivalent problems. Here we provide sufficient conditions for MDP problems to exhibit near reachability. Appendix A gives a formal proof of a folklore result related to coefficients of ergodicity while in Appendix B we provide a motivating numerical illustration of these results for a problem in equipment replacement in the presence of machine failures.

2. The general deterministic problem. The problem involves choosing a decision or action y_j at the beginning of each period $j = 1, 2, \ldots$. Let $Y_j = \{1, 2, \ldots, a_j\}$ represent the *finite* discrete set of all possible decisions (or controls) available in period j, where we assume that the cardinalities of the Y_j are uniformly bounded, i.e., there exists $a > 0$ such that $1 \le a_j \le a$, $\forall\, j = 1, 2, \ldots$. Let S denote the metric space of all possible (deterministic) states of the system at any time. Let $s_0 \in S$ denote the initial state of the system (beginning period 1), and s_{j-1} the state ending period $j-1$ (beginning period j). Let $S_j \subseteq S$ denote the (finite) set of *feasible* states ending period j (with $S_0 = \{s_0\}$), so that $s_j \in S_j$, for all $j = 1, 2, \ldots$. Define $Y_j(s_{j-1}) \subseteq Y_j$ to be the (finite) nonempty set of decisions available in period j, given that the system is in state $s_{j-1} \in S_{j-1}$ at the start of period j. Then, selecting decision $y_j \in Y_j(s_{j-1})$ causes the system to transition to state $s_j \in S_j$ at the end of period j by means of the state transition equation $s_j = f_j(s_{j-1}, y_j)$, where $f_j\colon F_j \to S_j$ is the (given) state transition function in period j, with domain

$$F_j = \big\{(s_{j-1}, y_j) \in S_{j-1} \times Y_j\colon\ y_j \in Y_j(s_{j-1})\big\},$$

and range

$$S_j = \big\{f_j(s_{j-1}, y_j)\colon\ s_{j-1} \in S_{j-1},\ \ y_j \in Y_j(s_{j-1})\big\}, \quad \forall\, j = 1, 2, \ldots,$$

so that each f_j is an onto mapping. In particular, we have

$$S_1 = \big\{f_1(s_0, y_1)\colon\ y_1 \in Y_1(s_0)\big\}.$$

Let Y denote the product space $\prod_{j=1}^{\infty} Y_j$ of all possible decision sequences over the infinite horizon (includes both feasible and infeasible sequences). An infinite decision sequence $y = \{y_j\}_{j=1}^{\infty}$ in Y will be called a *strategy*. The topological space Y is compact by the Tychonoff theorem and Hausdorff relative to the topology of componentwise convergence (see Munkres [15]). Because of the discreteness of the Y_j, componentwise convergence of a sequence yields eventual agreement in each component of the sequence, i.e., if $y^n \to y$ in Y, then for each k, there exists a positive integer m_k such that $n \ge m_k$ implies $y_j^n = y_j$, for each $j = 1, 2, \ldots, k$. Moreover, the product topology on Y is metrizable (Schochetman and Smith [19]). For each N, define $y \in Y$ to be *feasible through period* N if $y_j \in Y_j(s_{j-1})$, where $s_j = f_j(s_{j-1}, y_j)$, for all $j = 1, 2, \ldots, N$. Denote by X_N the subset of Y consisting of all such y, and by X, those y which are feasible through each $N = 1, 2, \ldots$. By our assumptions, the infinite horizon feasible set X is nonempty and closed in Y, that is, X is compact, and

$$X \subseteq X_{N+1} \subseteq X_N, \quad \forall\, N,$$

i.e., the X_N are nested downward. Moreover,

$$X = \bigcap_{N=1}^{\infty} X_N = \lim_{N\to\infty} X_N,$$

in the sense of Kuratowski (see Aubin [2], Kuratowski [14]). Now let $y = (y_1, y_2, \dots)$ be a *feasible* strategy, i.e., $y \in X$. For each $j \geq 1$, define $s_j(y)$ to be the state that y passes through at the end of period j. Hence, $s_j(y) = f_j(s_{j-1}(y), y_j)$, for all $j \geq 2$, with $s_1(y) = f_1(s_0, y_1)$. If $y \in X_N$, then the previous holds for $j = 1, \dots, N$ but not necessarily for $j > N$. Moreover, suppose $y, z \in X_N$, with $y_j = z_j$, for all $j = 1, \dots, N$. Then, $s_j(y) = s_j(z)$, for all $j = 1, \dots, N$.

Next, we introduce a cost structure. The cost in period j depends on the state s_{j-1} of the system and the chosen decision y_j, given that state. Thus, let $c_j(s_{j-1}, y_j)$ denote this real-valued cost, so that $c_j\colon F_j \to \mathbb{R}$. We assume that all costs are uniformly bounded, i.e., that there exists $0 < b < \infty$ such that

$$|c_j(s_{j-1}, y_j)| \leq b, \quad \forall\, (s_{j-1}, y_j) \in F_j, \ \ \forall\, j = 1, 2, \dots .$$

Let $C(x : j, k)$ denote the total cost of a strategy $x \in X_k$ from period j through period k inclusive, i.e.,

$$C(x : j, k) = \sum_{i=j}^{k} c_i(s_{i-1}(x), x_i), \quad \forall\, 1 \leq j \leq k, \ \ \forall k = 1, 2, \dots .$$

In particular, the total cost of reaching state $s_N(x)$ at the end of horizon N following strategy $x \in X_N$ from period 1 is given by

$$C(x : 1, N) = \sum_{i=1}^{N} c_i(s_{i-1}(x), x_i).$$

Also, the corresponding average cost-per-period is

$$A(x : 1, N) = \frac{1}{N} \sum_{i=1}^{N} c_i(s_{i-1}(x), x_i) = C(x : 1, N)/N.$$

In particular, if $x \in X \subseteq X_N$, then (conservatively) the average cost-per-period of x over the infinite horizon is given by

$$A(x) = \limsup_N A(x : 1, N) = \limsup_N [C(x : 1, N)/N].$$

Note that $|A(x)| \leq b, \ \forall\, x \in X$.

Our goal is to study the existence of *average optimal* solutions for our problem, i.e., optimal solutions for the mathematical programming problem $(\mathscr{D})$ given by

$$\inf_{x\in X} A(x). \tag{$\mathscr{D}$}$$

The set of such optimal solutions will be denoted by X^a, i.e.,

$$X^a \equiv \{x \in X\colon A(x) \leq A(y), \ \forall\, y \in X\}.$$

Although X is closed in Y, X^a need not be closed in Y (Schochetman and Smith [20]). Moreover, it is well-known that average optimal strategies can be far from optimal over the short-term, i.e., over finite horizons.

Next, we consider "finite horizon" truncations of $(\mathscr{D})$. Define

$$K_N \equiv \{(x_1, \dots, x_N) \in Y_1 \times \cdots \times Y_N\colon x_j \in Y_j(s_{j-1}), \ \forall\, 1 \leq j \leq N, \ s_j = f_j(s_{j-1}, x_j), \ \forall\, 1 \leq j \leq N-1\},$$

so that

$$X_N = K_N \times Y_{N+1} \times Y_{N+2} \times \cdots .$$

Hence, each X_N is the closed set of all arbitrary infinite extensions of elements of the finite set K_N, and X_N is compact, for all N. Note that the first N decisions of every member of X belong to K_N. For each N, consider the following problem $(\mathscr{D}_N)$:

$$\min_{x\in X_N} A(x : 1, N), \qquad \text{equivalently,} \qquad \min_{x\in X_N} C(x : 1, N). \tag{$\mathscr{D}_N$}$$

The real-valued functions $x \to A(x:1,N)$ defined on X_N are continuous, because they depend only on K_N, which is finite. (These functions attain finitely many distinct values on X_N.) Let X_N^a denote the set of average optimal strategies to $(\mathscr{D}_N)$, i.e.,

$$X_N^a \equiv \{x \in X_N\colon\ C(x:1,N) \le C(y:1,N),\ \forall\, y \in X_N\} = \{x \in X_N\colon\ A(x:1,N) \le A(y:1,N),\ \forall\, y \in X_N\},$$

which is not empty, for all N. At each stage, there exists a finite number of decisions, and hence, a finite number of possible strategies to each horizon. However, there exist infinitely many infinite horizon extensions of these strategies.

In Schochetman and Smith [20], an infinite horizon feasible strategy is defined to be *efficient* if it is optimal to each of its attained states. Accordingly, for each $N = 1, 2, \ldots$, let X_N^e denote the set of N-horizon feasible strategies that are efficient through period N, i.e.,

$$X_N^e \equiv \{x \in X_N\colon\ C(x:1,N) \le C(y:1,N),\ \forall\, y \in X_N \text{ such that } s_N(y) = s_N(x)\}.$$

These sets are nested downward, i.e.,

$$X_{N+1}^e \subseteq X_N^e, \quad \forall\, N,$$

by the principle of optimality. Also, let X^e denote the set of infinite horizon efficient strategies, i.e.,

$$X^e \equiv \bigcap_{N=1}^{\infty} X_N^e = \lim_N X_N^e = \limsup_N X_N^e = \liminf_N X_N^e,$$

where the limits are in the sense of Kuratowski. From Schochetman and Smith [20], it follows that $X^e \neq \varnothing$. (Note that in this reference, X^e and X_N^e are denoted by $\mathscr{X}$ and $\mathscr{X}_N$, respectively.) In Schochetman and Smith [20], it is shown that under a bounded reachability condition, we have $X^e \subseteq X^a$ and, in particular, there exists an average optimal solution.

In the next section we introduce a weakening of bounded reachability that we call near reachability. We show in §4 that a deterministic equivalent formulation of MDP problems satisfies this property under a mild ergodicity condition. In §3, we introduce a strengthening of efficiency that we term strong efficiency and establish that such solutions always exist and are, moreover, average optimal under near reachability.

3. Near reachability, strong efficiency, and average optimality. Recall that for those problems in Schochetman and Smith [20] that have the following bounded *time* reachability property, efficient solutions (which exist) are average optimal, i.e., $\varnothing \neq X^e \subseteq X^a$.

Definition 3.1 (Bounded Reachability (BR)). For problem $(\mathscr{D})$, there exists a positive integer r such that, for each $1 \le k < \infty$, each $s \in S_k$, and each finite sequence of states $(t_k, \ldots, t_{k+r})$ in $S_k \times \cdots \times S_{k+r}$, there exists $k \le l \le k + r$, and $w \in X_l$ for which $s_k(w) = s$ and $s_l(w) = t_l$.

Bounded reachability requires that it be possible to feasibly reach from any feasible state to any sequence of feasible states within a uniformly bounded time r.

As we shall see in Appendix B, problem $(\mathscr{D})$ need not have property (BR). Consequently, to obtain further results of the form $\varnothing \neq X^e \subseteq X^a$ for $(\mathscr{D})$, we require a weaker reachability property. In particular, this will be the case in §4, where we consider a natural deterministic problem corresponding to an infinite horizon, nonhomogeneous MDP. Accordingly, we introduce the following near (*state*) reachability property. Let ρ denote a metric on S.

Definition 3.2 (Near Reachability (NR)). For problem $(\mathscr{D})$:

(i) there exists a sequence $\{b_k\}_{k=1}^{\infty}$ of positive real numbers with $\lim_k (b_k/k) = 0$,

(ii) for each $\epsilon > 0$, there exists a sequence $\{l_{k,\epsilon}\}_{k=1}^{\infty}$ of positive integers, and

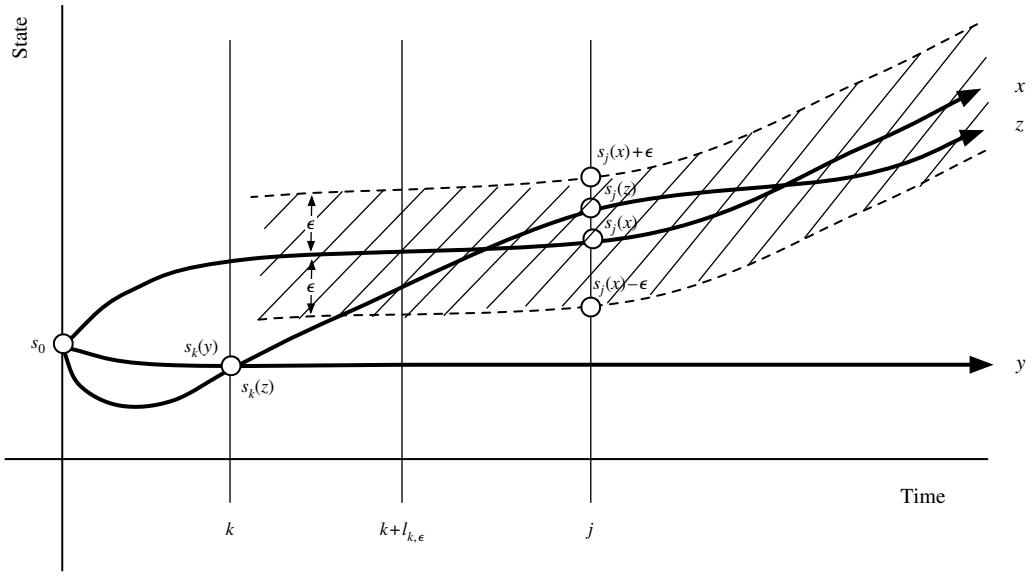
(iii) for each $x, y \in X$, and positive integer k, there exists $z \in X$ (depending on k, ϵ, x, y), for which

(iiia) $$s_k(z) = s_k(y),$$

(iiib) $$\rho(s_j(x), s_j(z)) < \epsilon, \quad \forall\, j \ge k + l_{k,\epsilon}, \quad \text{and}$$

(iiic) $$|C(x:k+1,j) - C(z:k+1,j)| \le b_k, \quad \forall\, j \ge k + l_{k,\epsilon}.$$

(See Figure 1.)



$$|C(x{:}k+1,j)-C(z{:}k+1,j)|\le b_k$$

FIGURE 1. Illustration of near reachability.

Near reachability roughly requires that we can reach from any state of a feasible decision sequence to a state close to a state of any other feasible decision sequence at an average cost that goes to zero as the period of that state goes to infinity.

LEMMA 3.1. *For problem* $(\mathscr{D})$, *property* (BR) *implies property* (NR).

PROOF. Suppose Property (BR) holds with $r>0$, as in Definition 3.1. Let $b_k=2br>0$, $\forall k$. Given $\epsilon>0$, let $l_{k,\epsilon}=r$, $\forall k$. Then $\lim_{k\to\infty} b_k=\lim_{k\to\infty}(2br/k)=0$.

Next, let x, y be elements of X, with k a fixed positive integer. By Property (BR) (for $s=s_k(y)$ and $t_l=s_l(x)$), there exists $k\le l\le k+r$ and $w\in X$ such that $s_k(w)=s_k(y)$ and $s_l(w)=s_l(x)$. Define

$$z\equiv(w_1,w_2,\ldots,w_l,x_{l+1},x_{l+2},\ldots).$$

Then $z\in X$ because $s_l(z)=s_l(w)=s_l(x)$. Also, $s_k(z)=s_k(w)=s_k(y)$. If $j\ge l$, then $s_j(x)=s_j(z)$, i.e.,

$$\rho(s_j(x),s_j(z))=0.$$

In particular, this is true for $j\ge k+l_{k,\epsilon}=k+r\ge l$. For such j, we have

$$C(x:k+1,j)=C(x:k+1,k+r)+C(x:k+r+1,j)$$

and

$$C(z:k+1,j)=C(z:k+1,k+r)+C(z:k+r+1,j).$$

However, $z_i=x_i$, for $i\ge l+1$, so that $s_i(z)=s_i(x)$ for $i\ge l$, and, in particular, for $i\ge k+r+1$. Thus,

$$C(z:k+r+1,j)=C(x:k+r+1,j),\quad \forall\, j\ge k+r+1,$$

so that

$$\begin{aligned}
|C(x:k+1,j)-C(z:k+1,j)| &= |C(x:k+1,k+r)-C(z:k+1,k+r)|\\
&= \left|\sum_{i=k+1}^{k+r} c_i(s_{i-1}(x),x_i)-\sum_{i=k+1}^{k+r} c_i(s_{i-1}(z),z_i)\right|\\
&= \left|\sum_{i=k+1}^{k+r}(c_i(s_{i-1}(x),x_i)-c_i(s_{i-1}(z),z_i))\right|\\
&\le [k+r-(k+1)+1]2b\\
&\le 2br,\quad \forall\, j\ge k+r=k+l_{k,\epsilon}.\quad \square
\end{aligned}$$

We turn now to strengthening the notion of efficiency. For fixed $\epsilon > 0$, $N = 1, 2, \ldots,$ and $s \in S_N$, let

$$B_\epsilon(s:N) \equiv \{t \in S_N\colon \rho(s,t) < \epsilon\},$$

which denotes the (open) ball in S_N consisting of all (finitely many) states t that are within ϵ of s. Also let

$$C_N^*(s) \equiv \min\{C(x:1,N)\colon x \in X_N,\ s_N(x) = s\}$$

and $A_N^*(s) = C_N^*(s)/N$, $\forall\, s \in S_N$. Then $C_N^*(s)$ (resp. $A_N^*(s)$), which is attained, is the smallest total (resp. average) cost of feasibly transitioning from the initial state s_0 to state s at the end of period N. Also define

$$S_N^*(\epsilon) \equiv \left\{s \in S_N\colon C_N^*(s) \le C_N^*(t),\ \forall\, t \in B_\epsilon(s:N)\right\}, \quad \forall\, N = 1, 2, \ldots,$$

so that $S_N^*(\epsilon)$ is the collection of feasible states s at time N having the smallest associated optimal cost $C_N^*(s)$ of any state t within a distance ϵ of s. Observe that

- if $\epsilon_1 < \epsilon_2$, then $S_N^*(\epsilon_2) \subseteq S_N^*(\epsilon_1) \subseteq S_N$;
- $S_N^*(\epsilon)$ is not empty, because at stage N, there is a finite number of feasible states; and
- if ϵ is sufficiently small, then $B_\epsilon(s:N) = \{s\}$ and $S_N^*(\epsilon) = S_N$, because S_N is a finite subset of S.

Definition 3.3 (N-Horizon ϵ-Efficient Strategies). Let $\epsilon > 0$ and $N = 1, 2, \ldots$. A strategy $x \in X_N$ is *N-horizon ϵ-efficient* if it has least total cost of all strategies y whose states $s_N(y)$ are within ϵ of $s_N(x)$ at time N, i.e.,

$$C(x:1,N) \le C(y:1,N), \quad \forall\, y \quad \text{such that} \quad s_N(y) \in B_\epsilon(s_N(x):N).$$

Hence, if $x \in X_N$ is N-horizon ϵ-efficient, then $s_N(x) \in S_N^*(\epsilon)$.

Let $X_N^e(\epsilon)$ denote the set of such strategies. Observe that

- if $\epsilon_1 < \epsilon_2$, then $X_N^e(\epsilon_2) \subseteq X_N^e(\epsilon_1) \subseteq X_N^e$;
- $X_N^e(\epsilon)$ is not empty;
- if ϵ is sufficiently small, then $X_N^e(\epsilon) = X_N^e$.

This notion was motivated by the fact that in the probabilistic framework, it may not be possible to reach a particular state exactly at some future horizon (as is possible in the deterministic case); so, instead of optimality to a single state, we allow optimality to a group of states in close proximity to the desired state. For each N, and each $s \in S_N$, let

$$X_N^*(s) \equiv \left\{x \in X_N\colon C_N^*(s) = C(x:1,N),\ s_N(x) = s\right\} = \left\{x \in X_N\colon A_N^*(s) = A(x:1,N),\ s_N(x) = s\right\}.$$

Thus, $X_N^*(s)$ is the set of all strategies in X_N which attain the given state s at time N at the lowest total (or average) cost. Observe that $X_N^*(s)$ is nonempty and closed. Also, for $\epsilon > 0$, we have

$$X_N^e(\epsilon) = \bigcup_{s \in S_N^*(\epsilon)} X_N^*(s) \subseteq X_N.$$

Then $X_N^e(\epsilon)$ is the set of all strategies in X_N that ϵ-efficiently pass through states in $S_N^*(\epsilon)$ at time N; it is closed, compact, and nonempty, since it is a finite union of closed, nonempty sets in compact Y. Observe that these strategies do *not* necessarily pass through an ϵ-efficient state (i.e., a state in $S_j^*(\epsilon)$) at any period j, before or after N. Hence, in particular we lack a "principle of optimality" for N-horizon ϵ-efficient solutions, i.e., in general, $X_{N+1}^e(\epsilon) \not\subseteq X_N^e(\epsilon)$, for $\epsilon > 0$.

Lemma 3.2. *For all $\epsilon > 0$, and all N, $X_N^a \subseteq X_N^e(\epsilon) \subseteq X_N$.*

Proof. Fix a positive integer N, $\epsilon > 0$, and suppose $x \in X_N^a$. Then x has the lowest cost (total or average) of all strategies in X_N. Hence, in particular, x has the lowest cost of all strategies to all s in $B_\epsilon(s_N(x):N)$. Thus, $x \in X_N^e(\epsilon)$ by definition. □

For the next definition, recall the following. If $V_n \subseteq Y$, $\forall\, n$, then $\limsup_n V_n$ is the subset of Y consisting of those y for which there exist a subsequence $\{V_{n_k}\}_{k=1}^\infty$ of $\{V_n\}_{n=1}^\infty$ and a corresponding sequence $\{y_k\}_{k=1}^\infty$ such that $y_k \in V_{n_k}$, $\forall\, k$, and $\lim_{k\to\infty} y_k = y$.

Defintion 3.4 (Strong Efficiency). Define

$$X^{se} \equiv \bigcup_{\epsilon > 0} \left(\limsup_N X_N^e(\epsilon) \right).$$

Because for all $\epsilon > 0$, $X_N^e(\epsilon) \neq \varnothing$, there is a sequence $x_N^e(\epsilon)$, $N = 1, 2, \ldots$ in compact Y with $x_N^e(\epsilon) \in X_N^e(\epsilon)$, for all N and hence a convergent subsequence $x_{N_k}^e(\epsilon)$, $k = 1, 2, \ldots$ with limit point $x^e(\epsilon) = \lim_{k=1}^{\infty} x_{N_k}^e(\epsilon) \in \limsup_N X_N^e(\epsilon)$, so that

$$X^{se} \neq \varnothing.$$

We refer to the elements of X^{se} as *strongly efficient* strategies. By contrast, note that

$$\limsup_N \left(\bigcup_{\epsilon>0} X_N^e(\epsilon) \right) = \limsup_N X_N^e = X^e.$$

The previous definition is justified by the following.

Lemma 3.3. *In general, $X^{se} \subseteq X^e$, i.e., $\bigcup_{\epsilon>0}(\limsup_N X_N^e(\epsilon)) \subseteq \limsup_N(\bigcup_{\epsilon>0} X_N^e(\epsilon))$.*

Proof. We have $X_N^e \supseteq X_N^e(\epsilon)$, $\forall\, \epsilon > 0$. Then, because the X_N^e are nested downward to X^e (Schochetman and Smith [19]),

$$\limsup_N X_N^e(\epsilon) \subseteq \limsup_N X_N^e = X^e, \quad \forall\, \epsilon > 0.$$

Hence,

$$X^{se} = \bigcup_{\epsilon>0} \left(\limsup_N X_N^e(\epsilon) \right) \subseteq X^e. \quad \square$$

The following result is our extension of Theorem 4.2 of Schochetman and Smith [20] to the case of problems $(\mathscr{D})$ for which property (BR) may fail. It is the main result of this section. We show that strongly effficient strategies are average optimal under property (NR). (In §4, we will apply this result to nonhomogeneous MDPs).

Theorem 3.1 (Average Optimality of Strongly Efficient Strategies). *Suppose problem $(\mathscr{D})$ has property (NR). Then, $\varnothing \neq X^{se} \subseteq X^a$.*

Proof. We showed above that $X^{se} \neq \varnothing$. Now suppose $x \in X^{se}$, so that

$$x \in \bigcup_{\epsilon>0} \left(\limsup_N X_N^e(\epsilon) \right).$$

This implies that there exists $\epsilon > 0$ such that $x \in \limsup_N X_N^e(\epsilon)$. We show that $x \in X^a$, i.e., $A(x) \leq A(y)$, $\forall\, y \in X$. Let $y \in X$. Also let $\{b_k\}_{k=1}^{\infty}$ and, for the given $\epsilon > 0$, let $\{l_{k,\epsilon}\}_{k=1}^{\infty}$ be as in the definition of (NR). Because $x \in \limsup_N X_N^e(\epsilon)$, there exist a subsequence $\{N_n\}_{n=1}^{\infty}$ and a corresponding sequence $\{x^n\}_{n=1}^{\infty}$ with $x^n \in X_{N_n}^e(\epsilon)$, $\forall\, n$, such that $x^n \to x$ in Y, as $n \to \infty$. Fix k. From §2, the x^n eventually agree with x in the first k components, i.e., there exists m_k large enough so that $n \geq m_k$ implies $x_j^n = x_j$, $\forall\, j = 1, 2, \ldots, k$. Choose m such that $m \geq m_k$ and $N_m > k + l_{k,\epsilon}$. Observe that $x_j^m = x_j$ for at least the first k components. Note also that $x^m \in X_{N_m}^e(\epsilon)$ implies that $x^m \in X_{N_m}^*(s)$, for some $s \in S_{N_m}^*(\epsilon)$. Hence, $s_{N_m}(x^m) = s$,

$$A(x^m : 1, N_m) = A_{N_m}^*(s_{N_m}(x^m)) \leq A_{N_m}^*(t), \quad \forall\, t \in B_\epsilon(s_{N_m}(x^m) \colon N_m).$$

By Property (NR) applied to k, y, and x^m, there exists $z \in X$ such that

(iiia) $s_k(z) = s_k(y)$;
(iiib) $\rho(s_j(x^m), s_j(z)) < \epsilon$, $\forall\, j \geq k + l_{k,\epsilon}$; and
(iiic) $|C(x^m : k+1, j) - C(z : k+1, j)| \leq b_k$, $\forall\, j \geq k + l_{k,\epsilon}$.

Let w denote the strategy

$$w = (y_1, \ldots, y_k, z_{k+1}, z_{k+2}, \ldots).$$

Then w is feasible since $s_k(z) = s_k(y)$. Note also that $s_j(w) = s_j(z)$, $\forall\, j \geq k+1$. First, consider the cost of following strategy x^m through period N_m. Because $N_m > k + l_{k,\epsilon}$, by property (iiib) we have that

$$\rho(s_{N_m}(x^m),\, s_{N_m}(z)) < \epsilon, \quad \text{so that} \quad s_{N_m}(w) = s_{N_m}(z) \in B_\epsilon(s_{N_m}(x^m) \colon N_m),$$

i.e.,

$$\rho(s_{N_m}(x^m), s_{N_m}(w)) < \epsilon.$$

Recall that $X_{N_m}^e(\epsilon)$ is the set of all $v \in X_{N_m}$ for which $C(v : 1, N_m) \leq C(u : 1, N_m)$, for all $u \in X_{N_m}$ such that $s_{N_m}(u) \in B_\epsilon(s_{N_m}(v) \colon N_m)$. Hence, because $x^m \in X_{N_m}^e(\epsilon)$ and $s_{N_m}(w) \in B_\epsilon(s_{N_m}(x^m) \colon N_m)$, we have that

$$C(x^m : 1, N_m) \leq C(w : 1, N_m) = C(w : 1, k) + C(w : k+1, N_m) = C(y : 1, k) + C(z : k+1, N_m).$$

Because $x_j = x_j^m$, for all $j = 1, \ldots, k$, we have $s_k(x) = s_k(x^m)$ and

$$\begin{aligned} C(x^m : 1, N_m) &= C(x^m : 1, k) + C(x^m : k+1, N_m) \\ &= C(x : 1, k) + C(x^m : k+1, N_m), \end{aligned}$$

which implies that

$$C(x : 1, k) + C(x^m : k+1, N_m) \le C(y : 1, k) + C(z : k+1, N_m),$$

i.e,

$$\begin{aligned} 0 &\le C(y : 1, k) - C(x : 1, k) + C(z : k+1, N_m) - C(x^m : k+1, N_m) \\ &\le C(y : 1, k) - C(x : 1, k) + \left| C(z : k+1, N_m) - C(x^m : k+1, N_m) \right|. \end{aligned}$$

By (iiic) on the previous page, because $N_m > k + l_{k,\epsilon}$, we have that

$$\left| C(z : k+1, N_m) - C(x^m : k+1, N_m) \right| \le b_k.$$

Thus,

$$0 \le C(y : 1, k) - C(x : 1, k) + b_k, \quad \text{i.e.,} \quad C(x : 1, k) \le C(y : 1, k) + b_k.$$

Because k is arbitrary,

$$\frac{C(x : 1, k)}{k} \le \frac{C(y : 1, k) + b_k}{k}, \ \forall\, k, \quad \text{so that} \quad A(x) = \limsup_k \frac{C(x : 1, k)}{k} \le \limsup_k \frac{C(y : 1, k) + b_k}{k}.$$

For bounded sequences, we have from Goldberg [9] that

$$\limsup_k \frac{C(y : 1, k) + b_k}{k} \le \limsup_k \frac{C(y : 1, k)}{k} + \limsup_k \frac{b_k}{k}, \quad \text{i.e.,} \quad A(x) \le A(y) + \limsup_k \frac{b_k}{k},$$

where $\limsup_k (b_k/k) = 0$. Hence, $A(x) \le A(y)$. Because y is arbitrary in X, $x \in X^a$. □

4. Application to nonhomogeneous MDPs. Our goal in this section is to apply the results of §3 to a stochastic problem recast as a deterministic optimization problem ($\mathscr{D}(\mathscr{S})$), where ($\mathscr{S}$) is the stochastic optimization problem corresponding to a general nonhomogeneous MDP. In particular, we give sufficient conditions, in terms of coefficients of ergodicity, for the MDP to have property (NR), i.e., for ($\mathscr{D}(\mathscr{S})$) to have property (NR).

Consider a system in which

- $I = \{1, 2, \ldots, m\}$ is the (finite, discrete) set of *MDP states* i of the system in any period j;
- $\sigma_0(i)$ is the probability that the initial MDP state of the system is i, so that

$$0 \le \sigma_0(i) \le 1, \quad \forall\, 1 \le i \le m,$$

$$\sum_{i=1}^{m} \sigma_0(i) = 1, \quad \text{and}$$

$$\sigma_0 = [\sigma_0(1) \ \ldots \ \sigma_0(m-tuple)] \in \mathbb{R}^m$$

is the associated probability mass function (pmf).

- $D_j(i)$ is the set of decisions that are admissible in period j, given that the system is currently in MDP state $i \in I$. We assume that the cardinality $|D_j(i)|$ of $D_j(i)$ is at most c, $\forall\, i$, and $\forall\, j$. Also,

$$D_j \equiv D_j(1) \times \cdots \times D_j(m), \quad \forall\, j = 1, 2, \ldots,$$

is the set of all admissible policies or *decision rules* δ_j in period j, so that the cardinalities $|D_j|$ of the D_j are uniformly bounded by c^m.

- $p_j(i, k : d)$ is the probability, in period j, that the system transitions to MDP state $k \in I$, given that it was in MDP state $i \in I$ ending period $j-1$, and admissible decision $d \in D_j(i)$ was selected. Necessarily, $\sum_{k=1}^{m} p_j(i, k : d) = 1$. For each $\delta_j \in D_j$, define the stochastic $m \times m$ matrix $P_j(\delta_j)$ as follows:

$$[P_j(\delta_j)]_{ik} \equiv p_j(i, k\colon \delta_j(i)), \quad \forall\, i, k \in I,$$

so that

$$\sum_{k=1}^{m} [P_j(\delta_j)]_{ik} = 1, \quad \forall\, i = 1, \ldots, m, \ \ \forall\, j = 1, 2, \ldots .$$

• $q_j(i,k:d)$ is the cost in period j of choosing decision $d \in D_j(i)$, given that the system is in MDP state $i \in I$ ending period $j-1$ and transitions to MDP state k at the end of period j. We assume that the $q_j(i,k:d)$ are uniformly bounded, i.e., we assume that there exists $b > 0$ sufficiently large so that

$$|q_j(i,k:d)| \le b, \quad \forall\, d \in D_j(i), \ \ \forall\, i,k \in I, \ \ \forall\, j = 1,2,\ldots .$$

If we let $\gamma_j(i:d)$ denote the expected cost, in period j, of choosing decision $d \in D_j(i)$, given that the system is in MDP state $i \in I$ ending period $j-1$, then

$$\gamma_j(i:d) = \sum_{k=1}^{m} q_j(i,k:d) \cdot p_j(i,k:d), \quad \text{and}$$

$$|\gamma_j(i:d)| \le b, \quad \forall\, d \in D_j(i), \ \ \forall\, i \in I, \ \ \forall\, j = 1,2,\ldots .$$

Thus, at the beginning of decision epoch j, the system is in some MDP state $i \in I$, and the decision maker chooses a decision $d \in D_j(i)$, generating an expected cost $\gamma_j(i:d)$. The evolution from the current MDP state i to the new MDP state k depends on the transition probabilities $p_j(i,k:d)$ which, in turn, depend on the current state i, the new state k, the decision d, and the period j.

The set $D = \prod_{j=1}^{\infty} D_j$ of all *strategies* is then the feasible region for our optimization problem. To describe the objective function for this problem, let $x = (x_j)_{j=1}^{\infty}$ be an arbitrary element of D. Then the implementation of strategy x generates a sequence of MDP states. At the end of period $j-1$, such a state is determined by the decision rule sequence $x_1,\ldots,x_{j-1}$. Let $L_j(x_1,\ldots,x_{j-1})$ denote the (random) MDP state in I ending period j, determined by the feasible strategy x, with j-th decision $x_j(L_j(x_1,\ldots,x_{j-1}))$. Consequently, the expected cost of strategy x in period j if we end period j in state $L_j(x_1,\ldots,x_{j-1})$ is

$$\Gamma_j(x) = \gamma_j\big(L_j(x_1,\ldots,x_{j-1}):\, x_j(L_j(x_1,\ldots,x_{j-1}))\big),$$

whose expected value is given by $E[\Gamma_j(x)] = \sum_{i\in I} \gamma_j(i:\, x_j(i)) P(L_j(x_1,\ldots,x_{j-1}) = i)$. Over the first N periods, the total expected cost of strategy $x \in X$ is given by $\sum_{j=1}^{N} E[\Gamma_j(x)]$, and the average expected cost-per-period by $(1/N)\sum_{j=1}^{N} E[\Gamma_j(x)]$.

Our infinite horizon, average cost, stochastic optimization problem $(\mathcal{S})$ is then given by

$$\min_{x\in D} A(x), \tag{$\mathcal{S}$}$$

where

$$A(x) \equiv \limsup_{N} \left\{ \frac{1}{N} \sum_{j=1}^{N} E[\Gamma_j(x)] \right\}, \quad \forall\, x \in D.$$

To proceed, we recast problem $(\mathcal{S})$ in a form $(\mathcal{D}(\mathcal{S}))$, which is a particular case of problem $(\mathcal{D})$. Our goal is to give sufficient conditions for $(\mathcal{S})$, i.e., $(\mathcal{D}(\mathcal{S}))$, to admit an average optimal strategy that is also strongly efficient. Considerable effort has been devoted to solving problem $(\mathcal{S})$ for just an average optimal strategy. Note that certain standard techniques, such as policy and value iteration, fail because the problem is time-dependent. It is possible to transform the nonhomogeneous problem into one that is homogeneous, but the state space becomes infinite, and there are no general algorithms for this case. Some methods for solving the nonhomogeneous MDP include a form of value iteration designed to recursively uncover a sequence of policies by solving increasingly longer horizon problems. A more common approach involves a *rolling horizon* procedure, where a horizon N is fixed, the N-period problem is solved, and the initial policy is implemented. Then the procedure is repeated from the new state, and so on. The pitfall with this procedure is that the sequence of policies attained will not in general be optimal. In Alden and Smith [1], a bound on the error generated by the rolling horizon procedure is given.

Not surprisingly, we intend to apply the main result Theorem 3.1 of §3 to problem $(\mathcal{S})$. Define the *deterministic* states of $(\mathcal{D}(\mathcal{S}))$ to be probability mass functions, i.e., pmf-states. Accordingly, let $S = [0,1]^m$ with metric ρ given by

$$\rho(\sigma,\tau) \equiv \|\sigma - \tau\|_{\infty} = \max_{1\le i\le m} |\sigma(i) - \tau(i)|, \quad \forall\, \sigma,\tau \in \mathbb{R}^m,$$

$s_0 = \sigma_0$, $S_0 = \{\sigma_0\}$, and, for all $j = 1,2,\ldots$, let

$$Y_j \equiv D_j; \quad s_j = \sigma_j = [\sigma_j(1),\ldots,\sigma_j(m)];$$

$$S_j \equiv \bigcup_{\sigma_{j-1}\in S_{j-1}} \big\{\sigma_j(\cdot:\, \sigma_{j-1}, \delta_j):\, \delta_j \in D_j\big\} \subseteq \mathbb{R}^m,$$

where each $\sigma_j(\cdot : \sigma_{j-1}, \delta_j)$ is the pmf-state given by

$$\sigma_j(k : \sigma_{j-1}, \delta_j) = \sum_{i=1}^{m} \sigma_{j-1}(i) \cdot p_j(i, k : \delta_j(i)), \quad \forall k \in I,$$

so that

$$\sum_{k=1}^{m} \sigma_j(k : \sigma_{j-1}, \delta_j) = 1, \quad \forall \sigma_{j-1} \in S_{j-1}, \ \forall \delta_j \in D_j,$$

and, in particular,

$$S_1 = \{\sigma_1(\cdot : \sigma_0, \delta_1) \colon \delta_1 \in D_1\}.$$

Moreover, for each $j = 1, 2, \ldots$, we have

$$Y_j(\sigma_{j-1}) = Y_j = D_j, \quad \forall \sigma_{j-1} \in S_{j-1};$$
$$F_j = S_{j-1} \times D_j; \quad \text{and}$$

$$\begin{aligned} c_j(\sigma_{j-1}, \delta_j) &= \sum_{i=1}^{m} \sigma_{j-1}(i) \cdot \gamma_j(i \colon \delta_j(i)) \\ &= \sum_{i=1}^{m} \left(\sigma_{j-1}(i) \cdot \sum_{k=1}^{m} q_j(i, k \colon \delta_j(i)) \cdot p_j(i, k \colon \delta_j(i)) \right) \\ &= \sum_{i=1}^{m} \sum_{k=1}^{m} \sigma_{j-1}(i) \cdot q_j(i, k \colon \delta_j(i)) \cdot p_j(i, k \colon \delta_j(i)), \end{aligned}$$

so that

$$\begin{aligned} \left| c_j(\sigma_{j-1}, \delta_j) \right| &\le \sum_{i=1}^{m} \sum_{k=1}^{m} \left| \sigma_{j-1}(i) \cdot q_j(i, k \colon \delta_j(i)) \cdot p_j(i, k \colon \delta_j(i)) \right| \\ &= \sum_{i=1}^{m} \sum_{k=1}^{m} \sigma_{j-1}(i) \cdot \left| q_j(i, k \colon \delta_j(i)) \right| \cdot p_j(i, k \colon \delta_j(i)) \\ &\le b \sum_{i=1}^{m} \sum_{k=1}^{m} \sigma_{j-1}(i) \cdot p_j(i, k \colon \delta_j(i)) = b \sum_{i=1}^{m} \left(\sigma_{j-1}(i) \cdot \sum_{k=1}^{m} p_j(i, k \colon \delta_j(i)) \right) \\ &= b \sum_{i=1}^{m} \sigma_{j-1}(i) = b, \quad \forall \sigma_{j-1} \in S_{j-1}, \ \forall \delta_j \in D_j. \end{aligned}$$

(Note that even if $P_j(\delta_j) = P_j(\eta_j)$, it's possible that $c_j(\sigma_{j-1}, \delta_j) \neq c_j(\sigma_{j-1}, \eta_j)$, for some $\sigma_{j-1} \in S_{j-1}$ and $\delta_j, \eta_j \in D_j$. Thus, we do not identify δ_j with $P_j(\delta_j)$, even if P_j is one-to-one.) Consequently, the transition functions

$$f_j \colon S_{j-1} \times D_j \to S_j$$

are given by

$$f_j(\sigma_{j-1}, \delta_j) = \sigma_{j-1} P_j(\delta_j), \quad \forall \sigma_{j-1} \in S_{j-1}, \ \forall \delta_j \in D_j.$$

For each $x \in D$,

$$\sigma_j(x) = \sigma_0 P_1(x_1) P_2(x_2) \ldots P_j(x_j)$$

is then the probability distribution of the MDP states of strategy x at the end of period j, Furthermore,

$$X = Y = D = \prod_{j=1}^{\infty} D_j,$$

so that all strategies are feasible,

$$c_j(\sigma_{j-1}(x), x_j) = E[\Gamma_j(x)] = \sum_{i=1}^{m} \gamma_j(i \colon x_j(i)) \cdot \sigma_{j-1}(x)(i), \quad \forall j = 1, 2, \ldots, \quad \text{and}$$

$$C(x : 1, N) = \sum_{j=1}^{N} c_j(\sigma_{j-1}(x), x_j) = \sum_{j=1}^{N} \sum_{i=1}^{m} \gamma_j(i \colon x_j(i)) \cdot \sigma_{j-1}(x)(i)$$

is the total cost of strategy $x \in D$ through period N. Finally,

$$A(x) = \limsup_N \frac{1}{N} \sum_{j=1}^{N} C(x:1,j) = \limsup_N \left\{ \frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{m} \gamma_j(i:\, x_j(i)) \cdot \sigma_{j-1}(x)(i) \right\}$$

is the average cost-per-period of any strategy x in D over the infinite horizon. Note that

$$|C(x:1,N)| \le bN \quad \text{and} \quad |A(x)| \le b, \qquad \forall\, x \in D.$$

We leave it to the reader to verify that these ingredients satisfy all of the hypotheses of §2. The resulting optimization problem $(\mathscr{D}(\mathscr{S}))$ has the same feasible strategies and objective function values as does the stochastic optimization problem $(\mathscr{S})$. Therefore, in particular, the average optimal strategies are the same.

Recall that a *coefficient of ergodicity* is a function defined on the $m \times m$ stochastic matrices $P = [p_{uv}]$, with values in the closed interval $[0, 1]$, and is continuous relative to the topology of coordinate-wise convergence (Seneta [21]). Two particularly well-known examples are given by

$$\phi(P) = 1 - \max_{1 \le v \le m} \left\{ \min_{1 \le u \le m} p_{uv} \right\} \qquad \text{and} \qquad \psi(P) = \frac{1}{2} \max_{1 \le u,\, v \le m} \left\{ \sum_{k=1}^{m} |p_{uk} - p_{vk}| \right\}.$$

The following is the main result of this section (where $D = X$).

THEOREM 4.1 (SUFFICIENT CONDITIONS FOR PROPERTY (NR)). *Suppose there exists $0 < \alpha < 1$, such that $\psi(P_j(x_j)) \le \alpha$, $\forall\, x \in D$, $\forall\, j = 1, 2, \ldots$. Then property (NR) holds for problem $(\mathscr{D}(\mathscr{S}))$. Consequently, $\varnothing \ne D^{se} \subseteq D^a$ and problem $(\mathscr{D}(\mathscr{S}))$, equivalently problem $(\mathscr{S})$, admits an average optimal solution that is also strongly efficient.*

PROOF. For each k, and $\epsilon > 0$, define

$$l_{k,\epsilon} = \left\lceil \frac{\ln(\epsilon/2m)}{\ln(\alpha)} \right\rceil > 0 \quad \text{and} \quad b_k = \frac{2bm}{1-\alpha}, \qquad \forall\, k = 1, 2, \ldots,$$

so that

$$\lim_{k \to \infty} \frac{b_k}{k} = 0.$$

Fix k, let $x, y \in D$, and define

$$z = (y_1, y_2, \ldots, y_k, x_{k+1}, x_{k+2}, \ldots)$$

(which is in $D = \prod_{j=1}^{\infty} D_j$, i.e., z is feasible), so that $\sigma_k(z) = \sigma_k(y)$, for all $1 \le j \le k$. We next show that z has the desired properties. Given $n = 1, 2, \ldots$ and $1 \le j \le n$, we obtain the stochastic matrices $P_j(x_j), \ldots, P_n(x_n)$ as above. For convenience, define

$$T_j^n(x) \equiv \begin{cases} P_j(x_j) P_{j+1}(x_{j+1}) \ldots P_n(x_n), & \text{for } 1 \le j \le n, \\ J, & \text{for } j > n, \end{cases}$$

where J is the $m \times m$ identity matrix. Note that $\sigma_j(x) = \sigma_0 T_1^j(x)$. Next, starting at stage k, we compare, at some later time h, the distance between the states resulting from following x versus z. Observe that for $h \ge k$, $\sigma_h(x) = \sigma_k(x) T_{k+1}^h(x)$. Then

$$\begin{aligned} \rho(\sigma_h(x), \sigma_h(z)) &= \left\| \sigma_h(x) - \sigma_h(z) \right\|_\infty = \left\| \sigma_k(x) T_{k+1}^h(x) - \sigma_k(z) T_{k+1}^h(z) \right\|_\infty \\ &= \left\| \sigma_k(x) T_{k+1}^h(x) - \sigma_k(z) T_{k+1}^h(x) \right\|_\infty = \left\| (\sigma_k(x) - \sigma_k(z)) T_{k+1}^h(x) \right\|_\infty, \end{aligned}$$

because strategy z is the same as strategy x after stage k. By Seneta [21],

$$\psi(T_{k+1}^h(x)) = \psi\big(P_{k+1}(x_{k+1}) P_{k+2}(x_{k+2}) \ldots P_h(x_h)\big) \le \psi(P_{k+1}(x_{k+1})) \psi(P_{k+2}(x_{k+2})) \ldots \psi(P_h(x_h)) \le \alpha^{h-k}, \quad \forall\, h \ge k+1.$$

Thus, for any column in $T_{k+1}^h(x)$, all entries are within $2\alpha^{h-k}$ of each other. By Seneta [21],

$$\rho(\sigma_h(x), \sigma_h(z)) = \left\| (\sigma_k(x) - \sigma_k(z)) T_{k+1}^h(x) \right\|_\infty \le 2\alpha^{h-k} m.$$

Because $\alpha < 1$, i.e., $\ln(\alpha) < 0$, we may let h be sufficiently large such that $2\alpha^{h-k}m < \epsilon$, i.e.,

$$h > k + \frac{\ln(\epsilon/2m)}{\ln(\alpha)} \quad \text{implies} \quad h \geq k + \left\lceil \frac{\ln(\epsilon/2m)}{\ln(\alpha)} \right\rceil = k + l_{k,\epsilon}.$$

This establishes part (iiib) of Definition 3.2.

We next show that the cost condition (iiic) holds for the b_k. That is, we show that,

$$\left|C(x : k+1, j) - C(z : k+1, j)\right| \leq b_k, \quad \forall\, j \geq k + l_{k,\epsilon}.$$

For $j \geq k+1$, we have

$$\begin{aligned} C(x : k+1, j) &= \sum_{h=k+1}^{j} \sum_{i=1}^{m} \big(\sigma_{h-1}(x)(i) \cdot \gamma_h(i\colon x_h(i))\big) = \sum_{h=k+1}^{j} \sum_{i=1}^{m} \big((\sigma_0 T_1^{h-1}(x))(i) \cdot \gamma_h(i\colon x_h(i))\big) \\ &= \sum_{h=k+1}^{j} \sum_{i=1}^{m} \big(\sigma_0 T_1^{k}(x) T_{k+1}^{h-1}(x)\big)(i) \cdot \gamma_h(i\colon x_h(i))\big). \end{aligned}$$

Similarly,

$$\begin{aligned} C(z : k+1, j) &= \sum_{h=k+1}^{j} \sum_{i=1}^{m} \big((\sigma_0 T_1^{k}(z) T_{k+1}^{h-1}(z))(i) \cdot \gamma_h(i\colon z_h(i))\big) \\ &= \sum_{h=k+1}^{j} \sum_{i=1}^{m} \big((\sigma_0 T_1^{k}(z) T_{k+1}^{h-1}(x))(i) \cdot \gamma_h(i\colon x_h(i))\big), \end{aligned}$$

because $z_h = x_h$, for $k+1 \leq h \leq j$. Hence,

$$\begin{aligned} \left|C(x : k+1, j) - C(z : k+1, j)\right| &= \left| \sum_{h=k+1}^{j} \sum_{i=1}^{m} \big((\sigma_0 (T_1^{k}(x) - T_1^{k}(z)) T_{k+1}^{h-1}(x))(i) \cdot \gamma_h(i\colon x_h(i))\big) \right| \\ &\leq \sum_{h=k+1}^{j} \sum_{i=1}^{m} \left| \big(\sigma_0 (T_1^{k}(x) - T_1^{k}(z)) T_{k+1}^{h-1}(x)\big)(i) \cdot \gamma_h(i\colon x_h(i)) \right| \\ &= \sum_{h=k+1}^{j} \sum_{i=1}^{m} \left| \big(\sigma_0 (T_1^{k}(x) - T_1^{k}(z)) T_{k+1}^{h-1}(x)\big)(i) \right| \cdot \left| \gamma_h(i\colon x_h(i)) \right| \\ &\leq \sum_{h=k+1}^{j} \sum_{i=1}^{m} 2b \cdot \psi(T_{k+1}^{h-1}(x)) \cdot \|\sigma_0\|_1, \end{aligned}$$

by Lemma A.3 of Appendix A. Thus, because $\|\sigma_0\|_1 = 1$, we have

$$\begin{aligned} \left|C(x : k+1, j) - C(z : k+1, j)\right| &\leq \sum_{h=k+1}^{j} 2bm \cdot \psi(T_{k+1}^{h-1}(x)) = 2bm \sum_{h=k+1}^{j} \psi(T_{k+1}^{h-1}(x)) \leq 2bm \sum_{h=k+1}^{j} \alpha^{h-k-1} \\ &= 2bm \sum_{h=0}^{j-k-1} \alpha^{h} < 2bm \sum_{h=0}^{\infty} \alpha^{h} = \frac{2bm}{1-\alpha} = b_k, \quad \forall\, j \geq k+1. \end{aligned}$$

In particular,

$$\left|C(x : k+1, j) - C(z : k+1, j)\right| \leq b_k, \quad \forall\, j \geq k + l_{k,\epsilon}.$$

For the second part, recall Theorem 3.1. □

In general, $\psi(P)$ is difficult to evaluate for general P. The following result is of some help, because $\phi(P)$ is, in general, easier to calculate.

COROLLARY 4.1. *If, for $0 < \alpha < 1$, we have $\phi(P_j(x_j)) \leq \alpha$, $\forall\, x \in D$, $\forall\, j = 1, 2, \ldots$, then the conclusions of Theorem 4.1 hold.*

PROOF. In general, $\psi \leq \phi$ (Seneta [21]). □

See Appendix B for a numerical illustration of these results for a problem in equipment replacement in the presence of machine failures.

5. Discussion. Although not explored in this paper, we conjecture that for our problem, the introduction of randomized policies (when transformed to our deterministic framework) can deliver exact reachability and thereby yield convergence of randomized efficient solutions through liminf, as well as limsup inclusion, i.e., full Kuratowski convergence. The reason for this belief is that the inclusion of randomized policies serves to enlarge the set of strategies to the convex hull of purely deterministic strategies when viewed in the deterministic framework. Nearest point selections, as in Schochetman and Smith [20], could then lead to a sequence of policies and strategies that converges to an average optimal solution, so that policy convergence, as well as average value convergence, would hold. A forward algorithm would then be in hand for recursive discovery of a strongly efficient, and hence average optimal, nonstationary strategy.

Appendix A. Coefficients of ergodicity. We establish a useful property of the coefficient of ergodicity ψ, which plays an important role in the proof of Theorem 4.1. This property is used to prove Theorem 6 of Hopp et al. [13]. However, to our knowledge, there exists no rigorous proof of the result (see Hopp [12] for its original statement). Consequently, we felt it necessary to provide a detailed proof of this property.

If v is any element of $\mathbb{R}^m$, define

$$\max(v) \equiv \max\{v_i\colon i = 1, 2, \ldots, m\} \qquad \text{and} \qquad \min(v) = \min\{v_i\colon i = 1, 2, \ldots, m\}.$$

Lemma A.1. *For any v in $\mathbb{R}^m$, we have*

$$\max(v) - \min(v) = \max_{1 \le j,\, k \le m} |v_j - v_k|.$$

Proof. Left to the reader. □

Lemma A.2. *If p, q are arbitrary probability distributions on $\{1, 2, \ldots, m\}$, and $r \in \mathbb{R}^m$, then*

$$\left| \sum_{j=1}^{m} (p(j) - q(j)) r_j \right| \le \max(r) - \min(r).$$

Proof. We have

$$\sum_{j=1}^{m} p(j) r_j \le \max(r) \cdot \sum_{j=1}^{m} p(j) = \max(r) \qquad \text{and} \qquad \sum_{j=1}^{m} q(j) r_j \ge \min(r) \cdot \sum_{j=1}^{m} q(j) = \min(r),$$

so that

$$\sum_{j=1}^{m} (p(j) - q(j)) r_j = \sum_{j=1}^{m} p(j) r_j - \sum_{j=1}^{m} q(j) r_j \le \max(r) - \min(r).$$

Because p and q are arbitrary, we may interchange them to get

$$\sum_{j=1}^{m} (q(j) - p(j)) r_j \le \max(r) - \min(r),$$

so that

$$\sum_{j=1}^{m} (p(j) - q(j)) r_j \ge \min(r) - \max(r) = -(\max(r) - \min(r)).$$

This completes the proof. □

Lemma A.3. *Let $P = [P_{ij}]$, $Q = [Q_{ij}]$, and $R = [R_{ij}]$ be arbitrary stochastic $m \times m$ matrices. Recall that*

$$\psi(R) = \tfrac{1}{2} \max_{1 \le i,\, j \le m} \left\{ \sum_{k=1}^{m} |R_{ik} - R_{jk}| \right\}.$$

If $v \in \mathbb{R}^m$, then

$$\|v(P - Q)R\|_\infty \le 2\psi(R)\|v\|_1.$$

PROOF. Note that, in particular, the rows of P and Q are probability distributions on $\{1, 2, \ldots, m\}$. Fix $i = 1, 2, \ldots, m$. Then

$$(v(P-Q)R)_i = \sum_{h=1}^{m}\sum_{j=1}^{m} v_j(P_{jh} - Q_{jh})R_{hi},$$

so that

$$\begin{aligned}
|(v(P-Q)R)_i| &= \left|\sum_{h=1}^{m}\sum_{j=1}^{m} v_j(P_{jh}-Q_{jh})R_{hi}\right| = \left|\sum_{j=1}^{m}\sum_{h=1}^{m} v_j(P_{jh}-Q_{jh})R_{hi}\right| = \left|\sum_{j=1}^{m} v_j \cdot \left(\sum_{h=1}^{m}(P_{jh}-Q_{jh})R_{hi}\right)\right| \\
&\le \sum_{j=1}^{m}\left(|v_j| \cdot \left|\sum_{h=1}^{m}(P_{jh}-Q_{jh})R_{hi}\right|\right) \\
&\le \sum_{j=1}^{m}|v_j| \cdot \left(\max_{1\le h\le m} R_{hi} - \min_{1\le h\le m} R_{hi}\right) \quad \text{(by Lemma A.2)} \\
&= \|v\|_1 \cdot \left(\max_{1\le h\le m} R_{hi} - \min_{1\le h\le m} R_{hi}\right) \\
&= \|v\|_1 \cdot \max_{1\le j,\, h\le m} |R_{ji} - R_{hi}| \quad \text{(by Lemma A.1)} \\
&\le \|v\|_1 \cdot \max_{1\le j\, h\le m} \sum_{i=1}^{m} |R_{ji} - R_{hi}| \le 2\psi(R)\|v\|_1.
\end{aligned}$$

This completes the proof, because i is arbitrary. □

Appendix B. An example in equipment replacement with unreliable machines. We illustrate the results of §4 in the particular context of equipment replacement in the presence of machine failures.

We begin by assembling the parameters of the problem to form the expressions for the costs and transition probabilities for the generic model of §4. Consider a machine that is either working (state 1) or has failed (state 2), so that $m = 2$, $I = \{1, 2\}$, $S = [0, 1]^2$, and

$$\rho(\sigma, \tau) = \|\sigma - \tau\|_\infty = \max\{|\sigma(1) - \tau(1)|, |\sigma(2) - \tau(2)|\}.$$

If, at the start of period j, the machine is working, we may either replace it (decision 1) or keep it (decision 2), so that $c = 2$,

$$D_j(1) = \{1, 2\}, \quad D_j(2) = \{1\}, \quad D_j = \{(1, 1), (2, 1)\}, \qquad \forall\, j = 1, 2, \ldots,$$

and $X = D = \{(1, 1), (2, 1)\}^\infty$. Suppose that initially, the machine is equally likely to be working or not, i.e., $\sigma_0 = [\frac{1}{2} \;\; \frac{1}{2}]$. For each $j = 1, 2, \ldots,$ let

$$[p_j(i, 1:d) \quad p_j(i, 2:d)] = \begin{cases} \left[\frac{1}{2} \;\; \frac{1}{2}\right] & \text{if } i = 1,\ d = 2, \\ \left[\frac{2}{3} \;\; \frac{1}{3}\right] & \text{if } i = 1,\ d = 1, \\ \left[\frac{1}{3} \;\; \frac{2}{3}\right] & \text{if } i = 2,\ d = 1. \end{cases}$$

For example, at the start of any period j, if the machine is working ($i = 1$), and we choose to replace it ($d = 1$), then at the start of period $j + 1$, the machine will be working ($k = 1$) with probability 2/3 and will have failed ($k = 2$) with probability 1/3. Thus, for each j, we have

$$P_j((1, 1)) = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \quad \text{and}$$

$$P_j((2, 1)) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

We assume that there are no salvage values and that the costs $q_j(i, k : d)$, for all j, $i = 1, 2$, $d \in D_j(i)$, are arbitrary—with the exception that they are uniformly bounded by $b > 0$. Consequently, the resulting nonhomogeneous MDP is a special case of that studied in the previous section.

Let $(\mathscr{M})$ denote the resulting machine failure version of $(\mathscr{S})$, with $(\mathscr{D}(\mathscr{M}))$ the corresponding deterministic version. For this model, we have

$$\big[\sigma_j(1\colon \sigma_{j-1}, \delta_j)\ \sigma_j(2\colon \sigma_{j-1}, \delta_j)\big] = \begin{cases} \sigma_{j-1}\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}, & \text{for } \delta_j = (1,1), \\ \sigma_{j-1}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}, & \text{for } \delta_j = (2,1); \end{cases}$$

and

$$c_j(\sigma_{j-1}, \delta_j) = \big[\,\sigma_{j-1}(1) \quad \sigma_{j-1}(2)\,\big]\begin{bmatrix} \gamma_j(1\colon \delta_j(1)) \\ \gamma_j(2\colon \delta_j(2)) \end{bmatrix}, \quad \forall\, \sigma_{j-1} \in S_{j-1},\ \forall\, \delta_j \in D_j,\ \forall\, j = 1, 2, \ldots .$$

Now let $x = (x_j)_{j=1}^{\infty} \in D$ be the strategy defined by $x_j = (1,1)$, $\forall\, j$, i.e., we replace the machine in each period whether it is working or not. Then

$$P_j(x_j) = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}, \qquad T_j^h(x) = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}^{h-j+1}, \ \forall\, h \geq j, \qquad \text{and}$$

$$\begin{aligned} \sigma_j(x) &= \sigma_0 P_1(x_1), \ldots, P_j(x_j) = \sigma_0 T_1^j(x) \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}^j = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}^{j-1} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}^{j-1} = \cdots = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad \forall\, j = 1, 2, \ldots . \end{aligned}$$

In fact, it is immediately obvious that x is the only strategy in D whose state in every period is $[\frac{1}{2}\ \ \frac{1}{2}]$. Hence, being the unique strategy passing through these states, it is necessarily efficient, i.e., $x \in D^e$.

If we also let $y = (y_j)_{j=1}^{\infty} \in D$ be the strategy defined by

$$y_j = \begin{cases} (2,1), & \text{for } j = 1, \\ (1,1), & \text{for } j > 1, \end{cases}$$

i.e., we initially do not replace a working machine and then replace each period thereafter, then

$$P_j(y_j) = \begin{cases} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}, & \text{for } j = 1, \\ \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}, & \text{for } j > 1; \end{cases} \qquad T_j^h(y) = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}^{h-j+1}, \ \forall\, h \geq j \geq 2;$$

$$\sigma_1(y) = \sigma_0 P_1(y_1) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{5}{12} & \frac{7}{12} \end{bmatrix};$$

and, by matrix diagonalization,

$$\begin{aligned} \sigma_{j+1}(y) &= \sigma_0 P_1(y_1) P_2(y_2), \ldots, P_{j+1}(y_{j+1}) = \sigma_0 P_1(y_1) T_2^{j+1}(y) = \sigma_1(y) T_2^{j+1}(y) \\ &= \begin{bmatrix} \frac{5}{12} & \frac{7}{12} \end{bmatrix}\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}^j = \begin{bmatrix} \frac{5}{12} & \frac{7}{12} \end{bmatrix}\left(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\right)^j \end{aligned}$$

$$= \tfrac{1}{2}\begin{bmatrix} \frac{5}{12} & \frac{7}{12} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix}^j\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \tfrac{1}{2}\begin{bmatrix} \frac{5}{12} & \frac{7}{12} \end{bmatrix}\begin{bmatrix} 3^{-j}+1 & -3^{-j}+1 \\ -3^{-j}+1 & 3^{-j}+1 \end{bmatrix}$$

$$= \tfrac{1}{2}\begin{bmatrix} \frac{5}{12} & \frac{7}{12} \end{bmatrix}\left(3^{-j}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right) = \tfrac{1}{3^j}\begin{bmatrix} -\frac{1}{12} & \frac{1}{12} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad \forall\, j \geq 1.$$

Thus,

$$\sigma_j(y) \neq \sigma_j(x), \ \ \forall\, j \geq 1, \quad \text{and} \quad \lim_{j\to\infty} \sigma_j(y) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

i.e., the states of strategy y can never equal the corresponding states of x, but they can come arbitrarily close, for a sufficiently large horizon. Consequently, property (BR) fails for problem $(\mathscr{D}(\mathscr{M}))$ so that property (BR) is too strong for applicability for all MDP problems. However, as we shall see next, problem $(\mathscr{D}(\mathscr{M}))$ does have property (NR).

Let $w = (w_j)_{j=1}^{\infty}$ be an arbitrary element of D. Then, for the coefficient of ergodicity ϕ, we have

$$\phi(P_j(w_j)) = 1 - \max_{1\leq k\leq m}\left\{\min_{1\leq i\leq m} p_j(i,k\colon w_j(i))\right\} \leq 1 - \tfrac{1}{3} = \tfrac{2}{3} < 1,$$

because

$$p_j(i,k\colon w_j(i)) \in \left\{\tfrac{1}{3}, \tfrac{2}{3}, \tfrac{1}{2}\right\}, \quad \forall\, i,k,j,w.$$

Hence, the hypothesis of Corollary 4.1 holds because $\alpha = 2/3 < 1$.

Consequently, for this case of $(\mathscr{M})$, we have that $\varnothing \neq D^{se} \subseteq D^a$ by Theorem 3.1 so that this is an example of an MDP with a strongly efficient strategy that is also average optimal. Note that these results are valid for any cost structure for $(\mathscr{M})$, as long as cost data are uniformly bounded.

We next show that there exist cost structures for which strategy x, although efficient is not average optimal and, hence, not strongly efficient, i.e., $x \notin D^a$, so that $x \notin D^{se}$ also. For this purpose, assume that, for all $j = 1, 2, \ldots,$

$$[q_j(i,1:d) \quad q_j(i,2:d)] = \begin{cases} [rq \quad rq], & \text{for } i=1,\ d=2, \\ [q \quad q], & \text{for } i=1,\ d=1, \\ [q \quad q], & \text{for } i=2,\ d=1, \end{cases}$$

for arbitrary $q > 0$ and $0 < r < 1$. Then

$$\gamma_j(1\colon x_j(1)) = \gamma_j(1:1) = q_j(1,1:1)\cdot p_j(1,1:1) + q_j(1,2:1)\cdot p_j(1,2:1) = \tfrac{2}{3}q + \tfrac{1}{3}q = q, \qquad \text{and}$$
$$\gamma_j(2\colon x_j(2)) = \gamma_j(2:1) = q_j(2,1:1)\cdot p_j(2,1:1) + q_j(2,2:1)\cdot p_j(2,2:1) = \tfrac{1}{3}q + \tfrac{2}{3}q = q,$$

so that

$$\begin{aligned} c_j(\sigma_{j-1}(x), x_j) &= \gamma_j(1:x_j(1))\cdot\sigma_{j-1}(x)(1) + \gamma_j(2\colon x_j(2))\cdot\sigma_{j-1}(x)(2) \\ &= \tfrac{1}{2}q + \tfrac{1}{2}q = q, \quad \forall\, j = 1,2,\ldots. \end{aligned}$$

Therefore,

$$A(x) = \limsup_N \frac{1}{N} C(x:1,N) = \limsup_N \frac{1}{N}\sum_{j=1}^{N} c_j(\sigma_{j-1}(x), x_j) = q.$$

Now define $z = (z_j)_{j=1}^{\infty}$ by $z_j = (2,1)$, $\forall\, j$. Then

$$P_j(z_j) = P_j((2,1)) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix},$$

$$\gamma_j(1:z_j(1)) = \gamma_j(1:2) = q_j(1,1:2)\cdot p_j(1,1:2) + q_j(1,2:2)\cdot p_j(1,2:2) = \tfrac{1}{2}q + \tfrac{1}{2}q = q, \qquad \text{and}$$
$$\gamma_j(2\colon z_j(2)) = \gamma_j(2:1) = q_j(2,1:1)\cdot p_j(2,1:1) + q_j(2,2:1)\cdot p_j(2,2:1) = \tfrac{1}{3}q + \tfrac{2}{3}q = q,$$

so that

$$\begin{aligned} c_j(\sigma_{j-1}(z), z_j) &= \gamma_j(1\colon z_j(1))\cdot\sigma_{j-1}(z)(1) + \gamma_j(2\colon z_j(2))\cdot\sigma_{j-1}(z)(2) \\ &= q\cdot\sigma_{j-1}(z)(1) + q\cdot\sigma_{j-1}(z)(2), \quad \forall\, j = 1,2,\ldots. \end{aligned}$$

Hence, by matrix diagonalization,

$$\begin{aligned}\sigma_j(z) &= \sigma_0 P_j(z_j)^j \\ &= \left[\tfrac{1}{2} \quad \tfrac{1}{2}\right]\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{3} & \frac{2}{3}\end{bmatrix}^j = \left[\tfrac{1}{2} \quad \tfrac{1}{2}\right]\left(\begin{bmatrix}1 & 3\\ 1 & -2\end{bmatrix}\cdot\begin{bmatrix}1 & 0\\ 0 & \frac{1}{6}\end{bmatrix}\cdot\tfrac{1}{5}\begin{bmatrix}2 & 3\\ 1 & -1\end{bmatrix}\right)^j \\ &= \left[\tfrac{1}{2} \quad \tfrac{1}{2}\right]\cdot\begin{bmatrix}1 & 3\\ 1 & -2\end{bmatrix}\cdot\begin{bmatrix}1 & 0\\ 0 & \frac{1}{6}\end{bmatrix}^j\cdot\tfrac{1}{5}\begin{bmatrix}2 & 3\\ 1 & -1\end{bmatrix} = \tfrac{1}{5}\left[1 \quad 6^{-j}/2\right]\begin{bmatrix}2 & 3\\ 1 & -1\end{bmatrix} \\ &= 1/6^j\left[\tfrac{1}{10} \quad -\tfrac{1}{10}\right] + \left[\tfrac{2}{5} \quad \tfrac{3}{5}\right], \quad \forall\, j \geq 1.\end{aligned}$$

Thus, $\sigma_j(x) \neq \sigma_j(z)$, for all $j = 1, 2, \ldots$. Moreover,

$$\begin{aligned}c_j(\sigma_{j-1}(z), z_j) &= rq\left(\tfrac{2}{5} + \frac{6^{-j}}{10}\right) + q\left(\frac{3}{5} - \frac{6^{-j}}{10}\right) \\ &= \frac{q}{5}(2r+3) + \frac{q}{10\cdot 6^j}(r-1),\end{aligned}$$

which implies that

$$\begin{aligned}\frac{1}{N}C(z:1,N) &= \frac{1}{N}\sum_{j=1}^{N}\frac{q}{5}(2r+3) + \frac{1}{N}\sum_{j=1}^{N}\frac{q(r-1)}{10\cdot 6^j} = \frac{q}{5}(2r+3) + \frac{q(r-1)}{10}\cdot\frac{1}{N}\sum_{j=1}^{N}\frac{1}{6^j} \\ &= \frac{q}{5}(2r+3) + \frac{q(r-1)}{10}\cdot\frac{1}{N}\left[\frac{1-1/6^{N+1}}{1-1/6} - 1\right] = \frac{q(2r+3)}{5} + \frac{6q(r-1)}{50N} - \frac{6q(r-1)}{50N\cdot 6^{N+1}} - \frac{q(r-1)}{10N}.\end{aligned}$$

Thus,

$$A(z) = \limsup_N \frac{1}{N}C(z:1,N) = \frac{q(2r+3)}{5} < q = A(x),$$

because $0 < r < 1$. Therefore, x is not average optimal for this cost structure, i.e., $x \notin D^a$. Consequently, $x \notin D^{se}$ either (Theorem 4.1). Hence, we see that (nonempty) D^{se} is strictly contained in D^e, which is not contained in D^a, in general.

Acknowledgments. This research was supported in part by the National Science Foundation under Grants DMI–9713723 and DMI–0322114. We are grateful to anonymous referees for several suggestions that significantly improved the clarity of the exposition.

References

[1] Alden, J. M., R. L. Smith. 1992. Rolling horizon procedures in nonhomogeneous Markov decision processes. *Oper. Res.* **40** S183–S194.
[2] Aubin, J.-P. 1990. *Set-Valued Analysis*. Birkhauser, Boston.
[3] Bertsekas, D., S. Shreve. 1978. *Stochastic Optimal Control: The Discrete Time Case*. Academic Press, San Diego.
[4] Derman, C. 1966. Denumerable state Markovian decision processes average cost case. *Ann. Math. Statist.* **37** 1545–1554.
[5] Dynkin, E., A. A. Yushkevich. 1979. *Controlled Markov Processes*. Springer, Berlin.
[6] Federgruen, A., H. C. Tijms. 1978. The optimality equation in average cost denumerable state semi-Markov decision problems, recurrency conditions and algorithms. *J. Appl. Probab.* **15** 356–373.
[7] Feinberg, E. 1982. Controlled Markov processes with arbitrary numerical criteria. *SIAM Theory Probab. Appl.* **25** 486–503.
[8] Feinberg, E., A. Shwartz. 2002. *Handbook of Markov Decision Processes: Methods and Algorithms*. Kluwer, Boston.
[9] Goldberg, R. R. 1964. *Methods of Real Analysis*. Blaisdell, Waltham, MA, 50.
[10] Guo, X., J. Liu, K. Liu. 2000. Nonhomogeneous Markov decision processes with Borel state space—The average criterion with nonuniformly bounded rewards. *Math. Oper. Res.* **25** 667–678.
[11] Halkin, H. 1974. Necessary conditions for optimal control problems with infinite horizons. *Econometrica* **42** 267–272.
[12] Hopp, W. 1984. Nonhomogeneous Markov decision processes with applications to R&D planning. Ph.D. dissertation, University of Michigan, Ann Arbor.
[13] Hopp, W., J. Bean, R. L. Smith. 1987. A new optimality criterion for nonhomogeneous Markov decision processes. *Oper. Res.* **35** 875–883.
[14] Kuratowski, K. 1966. *Topologie I, II*. Academic Press, New York.
[15] Munkres, J. R. 1975. *Topology: A First Course*. Prentice-Hall, Englewood Cliffs, NJ.
[16] Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York.

[17] Ross, S. M. 1968. Non-discounted denumerable Markovian decision models. *Ann. Math. Statist.* **39** 412–2423.
[18] Ryan, S. M., J. C. Bean, R. L. Smith. 1992. A tie-breaking rule for discrete infinite horizon optimization. *Oper. Res.* **40** S117–S126.
[19] Schochetman, I. E., R. L. Smith. 1991. Convergence of selections with applications in optimization. *J. Math. Anal. Appl.* **155** 278–292.
[20] Schochetman, I. E., R. L. Smith. 1998. Existence and discovery of average optimal solutions in deterministic infinite horizon optimization. *Math. Oper. Res.* **20** 416–432.
[21] Seneta, E. 1981. *Non-Negative Matrices and Markov Chains*. Springer-Verlag, New York.
[22] Tijms, H. C. 2003. *A First Course in Stochastic Models*. Wiley, New York.