Easy Affine Markov Decision Processes: Theory

Jie Ning

Weatherhead School of Management, Case Western Reserve University, jie.ning@case.edu

Matthew J. Sobel

Weatherhead School of Management, Case Western Reserve University, matthew.sobel@case.edu

July 11, 2017

Abstract

This paper characterizes the class of *decomposable affine Markov decision processes* (MDPs) whose optimal policy and value function can be specified analytically. These MDPs have continuous multidimensional endogenous states and actions, and an exogenous state that follows a finite Markov chain. Their three defining features are affine dynamics and single-period rewards, decomposable sets of feasible actions, and polyhedral properties of the decomposed sets of feasible actions. It is shown that decomposable affine MDPs with discounted criteria have an affine value function and an extremal optimal policy, both of which are determined by the solution of a small system of auxiliary equations. Also, the paper characterizes the solutions of MDPs that are partially decomposable and affine, and shows that they are composites of two smaller MDPs, one of which is a decomposable affine MDP.

1 Introduction and related literature

Many potential applications of Markov decision processes (MDPs) invite models with states and actions that are real continuous vectors and a criterion of maximizing the EPV (expected value of the present value) of the sequence of rewards. However, unless the dimensions of the state and action vectors are low, it is usually extremely challenging to solve such an MDP. This "curse of dimensionality" significantly limits the applicability of MDPs in professional practice and research. For example, the curse of dimensionality usually inhibits the inclusion of important exogenous information as state variables, such as market and economic conditions in business operations, and environmental conditions in fisheries and resource management.

A large literature focuses on narrowing the gap between the need for large-scale MDPs that incorporate relevant exogenous information and the limited ability to solve them. One stream of this literature takes an analytical perspective and identifies classes of MDPs with structured optimal policies that are relatively easy to analyze and compute; another stream takes a computational perspective and develops generic algorithms for approximate solutions. This paper is in the former camp, which is reviewed in detail below. See Powell (2007) and Zéphyr et al. (2017) for portals to the approximation literature.

LQG models comprise a well-known class of MDPs with structured optimal policies that are relatively easy to analyze and compute. They have **L**inear dynamics, concave **Q**uadratic immediate rewards, and stochastic elements with normal (**G**aussian) distributions. Much of the LQG literature consists of continuous-time models, but it was initiated with discrete-time models by Simon (1956) and Theil (1957). LQG models have a closed-form solution, the optimal policy is an affine function of the state, and the value function is a quadratic function of the state (Aström 1970, Bertsekas 1995).

This paper is similar in spirit to the LQG literature in that it identifies a class of MDPs, called *decomposable affine MDPs* (defined in §2), that can be analyzed and solved exactly and easily, regardless of the dimensions of the state and action vectors. Specifically, it has a value function and an *extremal* optimal policy that are affine functions of the endogenous state (i.e., the controlled state). The affine coefficients of the optimal policy and value function depend on the solution of *auxiliary equations*, which can be solved easily numerically.

Several classes of MDPs have the property that a myopic policy is optimal. Denardo and Rothblum (1979, 1983), Sobel (1990a,b), and others show that, with proper reformulation, the temporal dependence disappears in such MDPs. A solution can be obtained by optimizing a static (i.e., one-period) problem, which is much easier to accomplish than to optimize the original MDP directly.

Decomposable affine MDPs are similar to MDPs that admit myopic optimal policies in the following sense. In the latter, the temporal dependence can be suppressed and solving the original MDP reduces to solving a static optimization. In the former, the dependence on the endogenous state can be suppressed, and solving the original MDP reduces to solving auxiliary equations, which depend only on the exogenous state. Thus, the exogenous states can be viewed as the core of such an MDP; together with the auxiliary equations, they determine the computational burden and the slopes in the affine solution. The endogenous state, on the other hand, is more peripheral. As discussed in this paper and in Ning and Sobel (2017), hereafter designated [NS], this insight about the central role of the exogenous state and auxiliary equations can be exploited to simplify significantly the computational and qualitative analysis of decomposable affine MDPs.

This paper derives results concerning the value function and optimal policy of decomposable affine MDPs with discounted criteria. It also characterizes solutions of a class of MDPs that are partially affine and decomposable, and shows that a *partially decomposable affine MDP* is a composite of two related sub-MDPs. Its value function and optimal policy are composites of the solutions of these two sub-MDPs, one of which is a decomposable affine MDP. Thus, solving a partially decomposable affine MDP reduces to solving a set of auxiliary equations and a smaller sub-MDP.

The companion paper [NS] exploits the results in this paper. It develops algorithms to solve decomposable affine MDPs exactly, and presents applications of decomposable affine MDPs and partially decomposable affine MDPs in fishery management, vehicle fleet management, dynamic capacity portfolio management, and commodity procurement and trading.

In the remainder of this paper, §2 defines a decomposable affine MDP and §3 shows that the value function is an affine function, and that an extremal policy is optimal for finite-horizon and infinite-horizon discounted criteria. Section 3 also presented the auxiliary equations for each criterion. Section 4 extends the results in §3 to *partially decomposable affine MDPs*. Section 5 summarizes the paper.

2 Decomposable affine MDPs

The model is a time-homogeneous Markov decision process (MDP) with affinity and decomposability assumptions. The beginning of this section presents generic notation, §2.1 states affinity and decomposability assumptions, and §2.2 illustrates the notation and assumptions with a model of fishery management.

The sequence of states is $(s_1, e_1), (s_2, e_2), \ldots$ with $s_1, s_2, \ldots$ being endogenous, in the sense that these state variables are partially controlled (as is standard in an MDP), and $e_1, e_2, \ldots$ being exogenous (not controlled). The sequence of actions is a_1, a_2, $\ldots$ which are subject to constraints $a_t \in \mathcal{B}_{(s_t,e_t)} \subset \Re^{m\times 1}$ ($m \times 1$ real matrices) and $t \in \mathbb{N} = \{1, 2 \ldots\}$. Structural assumptions for $\mathcal{B}_{(s,e)}$, which is assumed nonempty for all $(s, e) \in \Re^{n\times 1} \times \Omega$, are given later in this section. The sequence of single-period rewards is $R_1, R_2, \ldots$, and $H_t = (s_1, e_1, a_1, R_1, ..., s_{t-1}, e_{t-1}, a_{t-1}, R_{t-1}, s_t, e_t)$ is the elapsed history up to the beginning of period t. Let $r(s_t, e_t, a_t) = \mathbb{E}(R_t|H_t, a_t)$ denote the expected single-period reward, where $\mathbb{E}$ is the expectation operator.

The exogenous process $e_1, e_2, \ldots$ (mnemonic for **e**xogenous) is a homogeneous Markov chain with state space Ω containing $\omega < \infty$ elements and generic one-step transition probability p_{ez} $(e, z \in \Omega)$. Exogeneity and homogeneity imply, for each $e \in \Omega$, that there is a random variable (r.v.) $\xi(e)$ on Ω with a probability distribution that depends only on $e \in \Omega$ such that $e_{t+1}|(H_t, a_t) \sim \xi(e_t)$. That is, the conditional distribution of e_{t+1} given (H_t, a_t) is the distribution of $\xi(e_t)$.

The endogenous state process $s_1, s_2, \ldots$ has state space $\Re_+^{n\times 1}$ and is homogeneous in the following sense. For each $(s, e, a) \in \mathcal{C} = \{(s, e, a) : a \in \mathcal{B}_{(s,e)}, e \in \Omega, s \in \Re_+^{n\times 1}\}$, there is an r.v. $T(s, e, a)$ that takes values in $\Re_+^{n\times 1}$ such that $s_{t+1}|(H_t, a_t) \sim T(s_t, e_t, a_t)$. That is, the conditional distribution of s_{t+1} given (H_t, a_t) is the distribution of $T(s_t, e_t, a_t)$.

Henceforth, given a vector-valued function $w(\cdot)$ and a matrix-valued function $W(\cdot)$, let $w_i(\cdot)$ denote the i-th element of $w(\cdot)$, and $W_{ij}(\cdot)$, $W_{\cdot j}(\cdot)$, and $W_{i\cdot}(\cdot)$ denote the respective (i, j)-th element, j-th column, and i-th row of $W(\cdot)$. The cardinality of a finite set $\mathcal{X}$ is denoted $|\mathcal{X}|$.

2.1 Decomposability and affinity assumptions

The MDP is called a *decomposable affine MDP* if it satisfies the assumptions in this subsection.

Decomposability. Let s_i and a_k denote the i-th and k-th components of s and a, respectively. The decomposability assumption is

$$\mathcal{B}_{(s,e)} = \times_{i=1}^{n} \mathcal{B}^i_{(s_i,e)}, \qquad (s, e) \in \Re_+^{n\times 1} \times \Omega. \tag{1}$$

That is, the set of feasible actions $\mathcal{B}_{(s,e)}$ is a product of n sets, the i-th of which depends on s only via s_i. Thus, given (s, e), the m components of $a \in \mathcal{B}_{(s,e)} \subset \Re^{m\times 1}$ can be partitioned into n subsets with the following property. Let $a^i(e)$ denote the subvector of a whose components are contained in subset i, and let $K^i(e)$ denote the dimension of $a^i(e)$. Then $a^i(e) \in \mathcal{B}^i_{(s_i,e)} \subset \Re^{K^i(e)\times 1}$ for $i = 1, \ldots, n$, and $\sum_{i=1}^{n} K^i(e) = m$. As is implied by the notation, the partition for $\{a^i(e) : i = 1, \ldots, n\}$ may vary with e but not with s. Henceforth, superscript i labels entities that are related to $a^i(e)$.

Affinity. The affinity assumption is

$$T(s,e,a) = \boldsymbol{Y}^{\mathbb{S}}(e)s + \boldsymbol{Y}^{\mathbb{A}}(e)a + \boldsymbol{Y}^{o}(e), \qquad (s,e,a) \in \mathcal{C}, \tag{2a}$$

$$r(s,e,a) = y^{\mathbb{S}}(e)s + y^{\mathbb{A}}(e)a + y^{o}(e), \qquad (s,e,a) \in \mathcal{C}. \tag{2b}$$

That is, the dynamics and expected single-period reward are affine functions of s and a with coefficients that may depend on e. The r.v.s $\boldsymbol{Y}^{\mathbb{S}}$(e), $\boldsymbol{Y}^{\mathbb{A}}(e)$, and $\boldsymbol{Y}^{o}(e)$ take values in $\Re^{n\times n}, \Re^{n\times m}$, and $\Re^{n\times 1}$, respectively, and $y^{\mathbb{S}}(e) \in \Re^{1\times n}$, $y^{\mathbb{A}}(e) \in \Re^{1\times m}$, and $y^{o}(e) \in \Re$. Boldface is a visual reminder that $\boldsymbol{Y}^{\mathbb{S}}(e), \boldsymbol{Y}^{\mathbb{A}}(e)$, and $\boldsymbol{Y}^{o}(e)$ are random (unlike $y^{\mathbb{S}}(e), y^{\mathbb{A}}(e)$, and $y^{o}(e)$ which are deterministic). Superscripts $\mathbb{S}$, $\mathbb{A}$, and o signify whether or not $\boldsymbol{Y}$ and y are coefficients of s and a.

Compound decomposability and affinity. This assumption consists of three parts. First, it requires that $\mathcal{B}^{i}_{(s_i,e)}$ is a bounded polyhedron for all $(s_i, e) \in \Re_+ \times \Omega$ and $i = 1, \ldots, n$. Let $\mathcal{X}^{i}_{(s_i,e)}$ denote the set of extreme points of $\mathcal{B}^{i}_{(s_i,e)}$. The second part of the assumption is that $|\mathcal{X}^{i}_{(s_i,e)}|$ is invariant with respect to s_i, i.e., the number of extreme points in $\mathcal{X}^{i}_{(s_i,e)}$ does not depend on s_i. For simplicity, henceforth $|\mathcal{X}^{i}_{e}|$ denotes the cardinality of $\mathcal{X}^{i}_{(s_i,e)}$. The last part of the assumption is that there exist $M^{i}(e) \in \Re^{K^{i}(e)\times|\mathcal{X}^{i}_{e}|}$ and $c^{i}(e) \in \Re$ such that

$$\mathcal{X}^{i}_{(s_i,e)} = \left\{M^{i}_{\cdot k}(e)s_i + c^{i}(e)1^{i}(e),\ k = 1, \ldots, |\mathcal{X}^{i}_{e}|\right\}, \tag{3}$$

where $1^{i}(e)$ is a column vector of all ones with the same dimension as $a^{i}(e)$. Thus, each extreme point of $\mathcal{B}^{i}_{(s_i,e)}$ is an affine function of $s_i \in \Re_+$; they share the same intercept $c^{i}(e)1^{i}(e)$ and the intercept vector consists of identical elements $c^{i}(e)$. Henceforth, $M^{i}_{\cdot k}(e)s_i + c^{i}(e)1^{i}(e)$ is termed the *k-th extreme point in* $\mathcal{X}^{i}_{(s_i,e)}$.

2.2 Fishery example

The following fishery management model exemplifies a decomposable affine MDP. The elements of the endogenous state vector $s_t \in \Re^{n\times 1}_{+}$ are the numbers of fish in n age classes in year t before harvesting begins. The action vector a_t specifies for each age class the number of fish that remain after harvesting is completed, so $n = m$ and the components of $s_t - a_t$ are the numbers of harvested fish in the various age classes. Exogenous state e_t captures the relevant aquatic and economic environment in year t, and is assumed not to be affected by a_t and s_t.

Let $w(e_t) \in \Re^{1\times n}$ denote a vector of the expected net profit per harvested fish in various age classes. Thus, the expected single-period reward in year t is $w(e_t)(s_t - a_t)$, which satisfies the affinity assumption in (2b) with $y^{\mathbb{S}}(e) = w(e)$, $y^{\mathbb{A}}(e) = -w(e)$, and $y^o(\cdot) \equiv 0$. Let $\theta_i(e)$ and $\lambda_i(e)$ denote the random survival and fecundity rates of age-i fish when the exogenous state is e. Assume that harvesting occurs before natural mortality, fertilization, and birth. Then the population size of age-i $(2 \le i \le n-1)$ fish before harvesting begins in year $t+1$ is distributed as $s_{i,t+1} \sim \theta_{i-1}(e_t)a_{i-1,t}$. The population size of the highest age class n is distributed as $s_{n,t+1} \sim \theta_{n-1}(e_t)a_{n-1,t} + \theta_n(e_t)a_{nt}$. The population size of age-1 fish is distributed as $s_{1,t+1} \sim \sum_{i=1}^{n} \lambda_i(e_t)a_{i-1,t}$ In the case of $n = 3$, the dynamical equations are

$$\begin{pmatrix} s_{1,t+1} \\ s_{2,t+1} \\ s_{3,t+1} \end{pmatrix} \sim \begin{pmatrix} \lambda_1(e_t) & \lambda_2(e_t) & \lambda_3(e_t) \\ \theta_1(e_t) & 0 & 0 \\ 0 & \theta_2(e_t) & \theta_3(e_t) \end{pmatrix} \begin{pmatrix} a_{1t} \\ a_{2t} \\ a_{3t} \end{pmatrix}, \tag{4}$$

which satisfies the affinity assumption in (2a).

The number of harvested fish in each age-class is nonnegative and at most the number before harvesting begins: $0 \le s_t - a_t \le s_t$. Thus, given $(s_t = s, e_t = e, a_t = a)$, the decomposability assumption is satisfied with $a^i(e) = a_i$ and $\mathcal{B}^i_{(s_i,e)} = [0, s_i]$ $(i = 1, \ldots, n)$. In this case, the rearrangement of a into subvectors $a^1(e), ..., a^n(e)$ does not vary with the exogenous state e and $K^i(e) \equiv 1$. Polyhedron $\mathcal{B}^i_{(s_i,e)}$ has two extreme points, $\mathcal{X}^i_{(s_i,e)} = \{0, s_i\}$, $|\mathcal{X}^i_e| \equiv 2$, and both extreme points are linear in s_i. Thus, one can construct a $K^i(e) \times |\mathcal{X}^i_e| = 1 \times 2$ matrix $M^i(e) = (0, 1)$, and define constant $c^i(\cdot) \equiv 0$ to satisfy (3) in the compound decomposability assumption.

3 Value function and optimal policy

This section shows that a decomposable affine MDP has a value function that is affine in the endogenous state, and an optimal policy that is *extremal*: i.e., given (s, e), it assigns $a^i(e)$ to the *optimal extreme point* of $\mathcal{B}^i_{(s_i,e)}$ for $i = 1, \ldots, n$. Subsection 3.1 defines optimality and the value functions for finite- and infinite-horizon criteria. Subsection 3.2 characterizes the value function and optimal policy for a finite-horizon criterion, and specifies recursive auxiliary equations. Subsection 3.3

does the same thing for an infinite-horizon criterion and shows that the auxiliary equations become functional equations.

Under the decomposability assumption, it is convenient to rewrite (2):

$$T(s,e,a) = \boldsymbol{Y}^{\mathbb{S}}(e)s + \sum_{i=1}^{n} \boldsymbol{Y}^{\mathbb{A}i}(e)a^i(e) + \boldsymbol{Y}^o(e), \qquad (s,e,a) \in \mathcal{C}, \tag{5a}$$

$$r(s,e,a) = y^{\mathbb{S}}(e)s + \sum_{i=1}^{n} y^{\mathbb{A}i}(e)a^i(e) + y^o(e), \qquad (s,e,a) \in \mathcal{C}, \tag{5b}$$

where $\boldsymbol{Y}^{\mathbb{A}i}(e)$ is the submatrix of $\boldsymbol{Y}^{\mathbb{A}}(e)$ that multiplies subvector $a^i(e)$ and takes values in $\Re^{n\times K^i(e)}$, and $y^{\mathbb{A}i}(e) \in \Re^{1\times K^i(e)}$ is the subvector of $y^{\mathbb{A}}(e)$ that multiplies $a^i(e)$.

3.1 Definitions of value function and optimal policy

This subsection heuristically defines optimality in finite- and infinite-horizons. See chapter 6 in Bertsekas and Shreve (1978) and section 4-2 in Heyman and Sobel (2004) for more precise developments. The definitions are important because they correspond to the dynamic programs that are the basis for subsequent analyses. Recall that H_t denotes an elapsed history up to the beginning of period t. Let $\mathcal{H}_t$ denote the set of possible H_t and let $\pi = (\delta_1, \delta_2, ...)$ be a policy, i.e., a (Borel-measurable) mapping of histories to feasible actions. Thus, $\delta_t(H_t) \in \mathcal{B}_{(s_t,e_t)}$ for all $H_t \in \mathcal{H}_t$ and $t \in \mathbb{N} = \{1, 2, ...\}$. Let $0 \le \beta < 1$ be a single-period discount factor, and let $\mathbb{E}_\pi$ denote the expectation operator applied to an r.v. with a distribution that depends on policy π.

Define

$$V_\pi^\tau(s,e) = \mathbb{E}_\pi\Big[\sum_{t=1}^{\tau} \beta^{t-1} R_t | H_1 = (s,e)\Big], \qquad \tau \in \mathbb{N}. \tag{6}$$

The *finite-horizon value function* is

$$v^\tau(s,e) = \sup_\pi V_\pi^\tau(s,e), \qquad \tau \in \mathbb{N}, \tag{7}$$

and policy π^* is *finite-horizon optimal* if

$$V_{\pi^*}^\tau(s,e) = v^\tau(s,e), \qquad \text{for all } (s,e) \in \Re_+^{n\times 1} \times \Omega \text{ and } \tau \in \mathbb{N}. \tag{8}$$

Let Π denote the set of policies for which the following *infinite-horizon* value function of π exists for all $(s, e) \in \Re_+^{n\times 1} \times \Omega$:

$$V_\pi(s,e) = \mathbb{E}_\pi\Big[\sum_{t=1}^{\infty} \beta^{t-1} R_t | (s_1, e_1) = (s,e)\Big]. \tag{9}$$

The *infinite-horizon value function* $v(\cdot,\cdot)$ is

$$v(s,e) = \sup_{\pi\in\Pi}\{V_\pi(s,e)\}. \tag{10}$$

Policy π^* is *infinite-horizon optimal* if

$$V_{\pi^*}(s,e) = v(s,e), \qquad \text{for all } (s,e) \in \Re_+^{n\times 1}\times\Omega. \tag{11}$$

3.2 Finite-horizon criterion

Using (5), the finite-horizon value function v^τ ($\tau \in \mathbb{N}$) satisfies the dynamic program $v^0(\cdot,\cdot) \equiv 0$ and, for $(s,e) \in \Re^{n\times 1}\times\Omega, \tau\in\mathbb{N}$,

$$\begin{aligned} v^\tau(s,e) &= \max_{a\in\mathcal{B}_{(s,e)}} \big\{ r(s,e,a) + \beta\mathbb{E}\big[v^{\tau-1}\big(T(s,e,a),\ \xi(e)\big)\big]\big\} \\ &= y^{\mathbb{S}}(e)s + y^o(e) \\ &+ \max_{a\in\mathcal{B}_{(s,e)}} \Bigg\{ \sum_{i=1}^{n} y^{\mathbb{A}i}(e)a^i(e) + \beta\mathbb{E}\Bigg[v^{\tau-1}\Bigg(\boldsymbol{Y}^{\mathbb{S}}(e)s + \sum_{i=1}^{n}\boldsymbol{Y}^{\mathbb{A}i}(e)a^i(e) + \boldsymbol{Y}^o(e),\ \xi(e)\Bigg)\Bigg]\Bigg\}. \end{aligned} \tag{12}$$

Define $Y^{\mathbb{S}}(e,z) = \mathbb{E}[\boldsymbol{Y}^{\mathbb{S}}(e)|\xi(e) = z] \in \Re^{n\times n}$, $Y^{\mathbb{A}i}(e,z) = \mathbb{E}[\boldsymbol{Y}^{\mathbb{A}i}(e)|\xi(e) = z] \in \Re^{n\times K^i(e)}$ for $i = 1,\ldots,n$, and $Y^o(e,z) = \mathbb{E}[\boldsymbol{Y}^o(e)|\xi(e) = z] \in \Re^{n\times 1}$. Theorem 1 states that the value function is affine in the endogenous state and specifies recursive auxiliary equations. The proof of the theorem reveals the roles of the assumptions in yielding the stated results.

Theorem 1. *The finite-horizon value function has the affine representation,*

$$v^\tau(s,e) = f^\tau(e)s + g^\tau(e), \qquad \tau\in\mathbb{N}, \tag{13}$$

where $f^\tau(e) = \big(f_1^\tau(e), ..., f_n^\tau(e)\big) \in \Re^{1\times n}$ *and* $g^\tau(e) \in \Re$ *satisfy recursive auxiliary equations* $f^0(\cdot) \equiv 0 \in$

$\Re^{1\times n}$, $g^0(\cdot) \equiv 0 \in \Re$, *and for* $\tau \in \mathbb{N}$,

$$f_i^\tau(e) = y_i^{\mathbb{S}}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y_{\cdot i}^{\mathbb{S}}(e,z) + \max_{k=1,\dots,|\mathcal{X}_e^i|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \right] M_{\cdot k}^i(e) \right\}, \tag{14a}$$

$$g^\tau(e) = y^o(e) + \beta \sum_{z\in\Omega} p_{ez}[f^{\tau-1}(z) Y^o(e,z) + g^{\tau-1}(z)] + \sum_{i=1}^n \left[c^i(e) \left(y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \right) 1^i(e) \right], \tag{14b}$$

Proof. 1. Initiate a proof by induction on τ by confirming (13) at $\tau = 0$ with $v^0(\cdot,\cdot) \equiv 0, f^0(\cdot) \equiv 0 \in \Re^{1\times n}$, and $g^0(\cdot) \equiv 0$. For any $\tau \in \mathbb{N}$, if $v^{\tau-1}(s,e) = f^{\tau-1}(e)s + g^{\tau-1}(e)$ for all $(s,e) \in \Re^{n\times 1} \times \Omega$, then (12) yields

$$\begin{aligned} & v^\tau(s,e) - y^{\mathbb{S}}(e)s - y^o(e) \\ = & \max_{a\in\mathcal{B}_{(s,e)}} \left\{ \sum_{i=1}^n y^{\mathbb{A}i}(e) a^i(e) + \beta \mathbb{E}\left[v^{\tau-1}\left(\boldsymbol{Y}^{\mathbb{S}}(e)s + \sum_{i=1}^n \boldsymbol{Y}^{\mathbb{A}i}(e) a^i(e) + \boldsymbol{Y}^o(e),\ \xi(e) \right) \right] \right\} \\ = & \max_{a\in\mathcal{B}_{(s,e)}} \left\{ \sum_{i=1}^n y^{\mathbb{A}i}(e) a^i(e) + \beta \mathbb{E}\left[f^{\tau-1}\big(\xi(e)\big) \left(\boldsymbol{Y}^{\mathbb{S}}(e)s + \sum_{i=1}^n \boldsymbol{Y}^{\mathbb{A}i}(e) a^i(e) + \boldsymbol{Y}^o(e) \right) + g^{\tau-1}\big(\xi(e)\big) \right] \right\}. \end{aligned} \tag{15}$$

Recall definition $Y^{\mathbb{S}}(e,z) = \mathbb{E}[\boldsymbol{Y}^{\mathbb{S}}(e)|\xi(e) = z]$ to write

$$\mathbb{E}\left[f^{\tau-1}\big(\xi(e)\big) \boldsymbol{Y}^{\mathbb{S}}(e) \right] = \sum_{z\in\Omega} \mathbb{E}\big[\, f^{\tau-1}\big(\xi(e)\big) \boldsymbol{Y}^{\mathbb{S}}(e) \big|\, \xi(e) = z \big] p_{ez} = \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{S}}(e,z). \tag{16}$$

Similarly, $\mathbb{E}\left[f^{\tau-1}\big(\xi(e)\big) \boldsymbol{Y}^{\mathbb{A}i}(e) \right] = \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \in \Re^{1\times K^i(e)}$,
$\mathbb{E}\big[f^{\tau-1}\big(\xi(e)\big) \boldsymbol{Y}^o(e) \big] = \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^o(e,z) \in \Re$, and $\mathbb{E}\big[g^{\tau-1}\big(\xi(e)\big) \big] = \sum_{z\in\Omega} p_{ez} g^{\tau-1}(z) \in \Re$.

Therefore, (15) is

$$
\begin{aligned}
&v^\tau(s,e) - y^{\mathbb{S}}(e)s - y^o(e) \\
&= \max_{a\in\mathcal{B}_{(s,e)}} \Bigg\{ \sum_{i=1}^n y^{\mathbb{A}i}(e)a^i(e) \\
&\qquad + \beta\sum_{z\in\Omega} p_{ez}\left[f^{\tau-1}(z)Y^{\mathbb{S}}(e,z)s + \sum_{i=1}^n f^{\tau-1}(z)Y^{\mathbb{A}i}(e,z)a^i(e) + f^{\tau-1}(z)Y^o(e,z) + g^{\tau-1}(z)\right]\Bigg\} \\
&= \beta\sum_{z\in\Omega} p_{ez}\left[f^{\tau-1}(z)Y^{\mathbb{S}}(e,z)s + f^{\tau-1}(z)Y^o(e,z) + g^{\tau-1}(z)\right] \qquad \text{(17a)} \\
&\qquad\quad + \max_{a\in\mathcal{B}_{(s,e)}} \left\{ \sum_{i=1}^n y^{\mathbb{A}i}(x)a^i(e) + \beta\sum_{z\in\Omega} p_{xz}f^{\tau-1}(z)\sum_{i=1}^n Y^{\mathbb{A}i}(e,z)a^i(e)\right\}. \qquad \text{(17b)}
\end{aligned}
$$

The remainder of the proof shows that (17b) is an affine function of s.

The linearity of the optimand of (17b) and the decomposability of $\mathcal{B}_{(s,e)}$ imply that (17b) is the sum of n sub-optimizations, the i-th of which has the components of $a^i(e) \in \mathcal{B}^i_{(s_i,e)}$ as variables and has $y^{\mathbb{A}i}(x)a^i(e) + \beta\sum_{z\in\Omega} p_{xz}f^{\tau-1}(z)Y^{\mathbb{A}i}(e,z)a^i(e)$ as the maximand. Let Z_i denote the optimal value of the i-th optimization. Then (17b) equals $\sum_{i=1}^n Z_i$ where

$$
Z_i = \max_{a^i(e)\in\mathcal{B}^i_{(s_i,e)}} \left\{ \left[y^{\mathbb{A}i}(e) + \beta\sum_{z\in\Omega} p_{ez}f^{\tau-1}(z)Y^{\mathbb{A}i}(e,z)\right] a^i(e)\right\}. \tag{18}
$$

This is a linear program with a bounded polyhedral feasibility set $\mathcal{B}^i_{(s_i,e)}$, so its optimum is achieved at an extreme point. Since the set of extreme points of $\mathcal{B}^i_{(s_i,e)}$ satisfies (3),

$$
\begin{aligned}
Z_i &= \max_{k=1,\dots,|\mathcal{X}^i_e|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta\sum_{z\in\Omega} p_{ez}f^{\tau-1}(z)Y^{\mathbb{A}i}(e,z)\right]\left(M^i_{\cdot k(e)}s_i + c^i(e)1^i(e)\right)\right\} \qquad \text{(19a)} \\
&= \left[\max_{k=1,\dots,|\mathcal{X}^i_e|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta\sum_{z\in\Omega} p_{ez}f^{\tau-1}(z)Y^{\mathbb{A}i}(e,z)\right] M^i_{\cdot k}(e)\right\}\right] s_i \qquad \text{(19b)} \\
&\qquad + \left[y^{\mathbb{A}i}(e) + \beta\sum_{z\in\Omega} p_{ez}f^{\tau-1}(z)Y^{\mathbb{A}i}(e,z)\right] 1^i(e)c^i(e), \qquad \text{(19c)}
\end{aligned}
$$

where the second equality is implied by $s_i \geq 0$ and (19c) being constant with respect to k.

Therefore, (15) is:

$$
v^\tau(s,e) = y^{\mathbb{S}}(e)s + y^o(e) + \beta\sum_{z\in\Omega} p_{ez}\left[f^{\tau-1}(z)Y^{\mathbb{S}}(e,z)s + f^{\tau-1}(z)Y^o(e,z) + g^{\tau-1}(z)\right] + \sum_{i=1}^n Z_i. \tag{20}
$$

Use (19) and collect terms in (20) to complete the induction and the proof by confirming that $v^\tau(s,e) = f^\tau(e)s + g^\tau(e)$ with f^τ and g^τ satisfying (14). $\square$

The inductive proof illustrates the roles of the affinity and decomposability assumptions in ensuring the affinity of the value function. First, the affinity of the dynamics and single-period expected reward yields an affine objective function in the induction step. Second, the decomposability assumption allows the constraints to be decomposed. Paired with the affinity of the objective function, this allows the decomposition of optimization (17b) over $\mathcal{B}_{(s,e)}$ into n smaller optimizations over $\mathcal{B}^i_{(s_i,e)}$. Third, the assumption that $\mathcal{B}^i_{(s_i,e)}$ is a bounded polyhedron makes the optimization over $\mathcal{B}^i_{(s_i,e)}$ a linear program. Finally, the affinity in s_i of the extreme points due to (3) and $s_i \geq 0$ ensure that the optimal objective is an affine function of s_i, which establishes the induction step. The assumption that the extreme points are linear in s_i is especially noteworthy. Generally, an extreme point of a linear program is only piece-wise linear in s_i, which could not sustain the induction. This explains why a plethora of linear programming formulations of dynamic optimization models have *nonlinear* value functions (see material on sensitivity analysis in, e.g., Dantzig 1963 and Mathur and Solow 1994).

Let $J_i^{\tau-1}(e)$ denote an index $k \in \{1, \ldots, |\mathcal{X}^i_e|\}$ that achieves the maximum in (14a):

$$J_i^{\tau-1}(e) \in \underset{k=1,\ldots,|\mathcal{X}^i_e|}{\arg\max} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \right] M^i_{\cdot k}(e) \right\}, \tag{21}$$

and let $A^{\tau i}(\cdot, e)$ denote an optimal single-period decision rule for $a^i(e)$ with τ periods remaining in the horizon. Thus, $A^{\tau i}(s,e) \in \mathcal{B}^i_{(s_i,e)} \subset \Re^{K^i(e)\times 1}$ and $\left(A^{\tau i} : i = 1, \ldots, n\right)$ is an optimal single-period decision rule for a.

Theorem 2. *With $\tau \in \mathbb{N}$ periods remaining in the horizon, $(A^{\tau i} : i = 1, \ldots, n)$ is an optimal single-period decision rule where $A^{\tau i}(s,e)$ is the $J_i^{\tau-1}(e)$-th extreme point in $\mathcal{X}^i_{(s_i,e)}$:*

$$A^{\tau i}(s,e) = M^i_{\cdot, J_i^{\tau-1}(e)}(e) \times s_i + c^i(e) 1^i(e). \tag{22}$$

Proof. Substitute $v^\tau(s,e) = f^\tau(e)s + g^\tau(e)$ (Theorem 1) in (12) to obtain (17b). Thus, the solution of (18) implies that (22) is optimal. $\square$

It follows from Theorem 2 that assigning $a^i(e)$ to be the k-th extreme point of $\mathcal{B}^i_{(s_i,e)}$ is optimal only if $\left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z)\right] M^i_{\cdot k}(e)$ is maximal with respect to $k = 1, \ldots, |\mathcal{X}^i_e|$. Thus, this optimal policy is provided by the solution of the recursive auxiliary equations. Furthermore, if multiple maximands on the right side of (21) achieve the maximum, then each of the associated extreme points is optimal, implying that such a tie is broken arbitrarily. That is, there can be multiple optimal policies that give rise to the unique value function.

From Theorems 1 and 2, the specification of the value function and optimal policy requires knowledge of f^τ and g^τ, which solve the recursive auxiliary equations (14). Thus, the computational task of solving dynamic program (12) reduces to solving the recursive auxiliary equations, which is easy for two reasons. First, the domain of the auxiliary equations is only $\{1, ..., n\} \times \Omega$ rather than the entire state space $\Re^n_+ \times \Omega$ as in (12). Second, the optimization in the auxiliary equations is over a finite set of extreme points of $\mathcal{B}^i(s_i, e)$, rather than over the continous set $\mathcal{B}_{(s,e)}$ as in (12). Therefore, the obstacle erected by the continuity of the endogenous state and action vectors does not arise in decomposable affine MDPs. An exact solution free of discretization errors can be obtained easily.

The extremal property of an optimal policy provides an intuitive basis for the significant computational advantages of decomposable affine MDPs. It implies that, for all $i = 1, \ldots, n$, no interior point of $\mathcal{B}^i(s_i, e)$ can strictly dominate a best extreme point; thus, discretization of the set of feasible actions is unnecessary.

The absence of the necessity to discretize action and endogenous state vectors and, therefore, the alleviation of the curse of dimensionality through auxiliary equations, comes at a mild price. Unlike dynamic program (12), which has a scalar-valued value function v^τ but an intimidatingly complex domain (s and e), the auxiliary equations have vector-valued functions (f^τ, g^τ) but almost trivially small domains ($\{1, ..., n\} \times \Omega$ for f^τ and Ω for g^τ). Since $f^\tau(e) \in \Re^{n\times 1}$, an increase in the dimension of the endogenous state vector increases the size of the problem one needs to solve.

3.3 Infinite-horizon criterion

This subsection shows that an infinite-horizon value function and optimal policy inherit key properties of their finite-horizon counterparts. The value function is affine in the endogenous state and a stationary extremal policy is optimal. As in the finite-horizon case, the value function and optimal policy are completely characterized by auxiliary equations which are a stationary version of (14).

The following notation is used to specify a sufficient condition in the next theorem. For $i = 1, \dots, n$ and $e, z \in \Omega$, define $U^i(e,z) \in \Re^{n \times |\mathcal{X}_e^i|}$ with k-th column

$$U^i_{\cdot k}(e,z) = Y^{\mathbb{S}}_{\cdot i}(e,z) + Y^{\mathbb{A}i}(e,z) M^i_{\cdot k}(e), \qquad \big(k = 1, \dots, |\mathcal{X}_e^i|\big) \tag{23}$$

and define

$$\theta = \beta \max\left\{ \sum_{z \in \Omega} p_{ez} \sum_{j=1}^{n} |U^i_{jk}(e,z)| : k = 1, \dots, |\mathcal{X}_e^i|;\ e \in \Omega;\ i = 1, ..., n \right\}. \tag{24}$$

Theorem 3. *If $\theta < 1$ and $V^\tau_\pi(s,e) \to V_\pi(s,e)$ as $\tau \to \infty$ (for all (s,e) and $\pi \in \Pi$), then the infinite-horizon value function is*

$$v(s,e) = f(e)s + g(e), \tag{25}$$

where $f(e) = \big(f_1(e), ..., f_n(e)\big) \in \Re^{1 \times n}$ and $g(e) \in \Re$ uniquely solve auxiliary equations

$$f_i(e) = y^{\mathbb{S}}_i(e) + \beta \sum_{z \in \Omega} p_{ez} f(z) Y^{\mathbb{S}}_{\cdot i}(e,z) + \max_{k=1,\dots,|\mathcal{X}_e^i|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z \in \Omega} p_{ez} f(z) Y^{\mathbb{A}i}(e,z) \right] M^i_{\cdot k}(e) \right\}, \tag{26a}$$

$$g(e) = y^o(e) + \beta \sum_{z \in \Omega} p_{ez} \big[f(z) Y^o(e,z) + g(z) \big] + \sum_{i=1}^{n} \left[c^i(e) \left(y^{\mathbb{A}i}(e) + \beta \sum_{z \in \Omega} p_{ez} f(z) Y^{\mathbb{A}i}(e,z) \right) 1^i(e) \right]. \tag{26b}$$

Proof. The proof uses the following lemma (which is proved in the appendix).

Lemma 1. *If $\theta < 1$, for each $e \in \Omega$ there exist $f(e) = \lim_{\tau \to \infty} f^\tau(e)$ and $g(e) = \lim_{\tau \to \infty} g^\tau(e)$ which are the unique solutions of* (26a) *and* (26b)*, respectively.*

It follows from Lemma 1 that taking limits on both sides of $v^\tau(s,e) = f^\tau(e)s + g^\tau(e)$ yields $\bar{v}(s,e) = f(e)s + g(e)$, where $\bar{v}(s,e) = \lim_{\tau \to \infty} v^\tau(s,e)$. The remainder of the proof shows that $\bar{v} = v$. The definition of $v^\tau(s,e)$ implies $v^\tau(s,e) \geq V^\tau_\pi(s,e)$ for every $\pi \in \Pi$ and $\tau \in \mathbb{N}$; thus, $V^\tau_\pi(s,e) \to V_\pi(s,e)$ as $\tau \to \infty$ implies $\bar{v}(s,e) \geq V_\pi(s,e)$. Therefore, definition $v(s,e) = \sup_{\pi \in \Pi} V_\pi(s,e)$ implies $V_\pi(s,e) \leq v(s,e) \leq \bar{v}(s,e)$. If $v(s,e) < \bar{v}(s,e)$, then there exists $\epsilon > 0$ and $\tau_\epsilon < \infty$ such that $V^\tau_\pi(s,e) + \epsilon \leq v^\tau(s,e)$ for all $\tau \geq \tau_\epsilon$ and $\pi \in \Pi$. Consequently, $v^\tau(s,e) > \sup_{\pi \in \Pi} V^\tau_\pi(s,e)$ which contradicts the definition of $v^\tau(s,e)$. Therefore, $\bar{v} = v$. □

Although it may be intuitive that $V^\tau_\pi(s,e) \to V_\pi(s,e)$ as $\tau \to \infty$, this assumption is not superfluous as shown by the following counterexample with $m = n = \omega = 1$. Let $\mathcal{B}_{(s,e)} = [0, s]$, $r(s, e, a) = a$, $T(s, e, a) = a/\beta$, and let π be the policy $a_t = s_t$ for all t. This yields $V_\pi^\tau(s, e) = \tau$ which diverges as $\tau \to \infty$. A more profound counterexample is Example 1 on page 215 of Bertsekas and Shreve (1978). The following proposition provides a sufficient condition for the validity of this assumption.

Proposition 1. *If*

$$\mathbb{E}_\pi\left[\sum_{t=1}^{\infty} \beta^{t-1}|R_t|\big|(s_1, e_1) = (s, e)\right] < \infty, \tag{27}$$

then $V_\pi^\tau(s, e) \to V_\pi(s, e)$ *as* $\tau \to \infty$.

Proof. Condition (27) and Fubini's Theorem permit the interchange of expectation and summation:

$$V_\pi(s, e) = \mathbb{E}_\pi\left[\sum_{t=1}^{\infty} \beta^{t-1} R_t|H_1 = (s, e)\right] = \sum_{t=1}^{\infty} \beta^{t-1}\mathbb{E}_\pi[R_t|H_1 = (s, e)] \tag{28a}$$

$$= \lim_{\tau\to\infty} \sum_{t=1}^{\tau} \beta^{t-1}\mathbb{E}_\pi[R_t|H_1 = (s, e)] = \lim_{\tau\to\infty} V_\pi^\tau(s, e). \tag{28b}$$

□

A by-product of Theorem 3 is that v solves the standard functional equation of discounted MDPs.

Corollary 1. *If* $\theta < 1$ *and* $V_\pi^\tau(s, e) \to V_\pi(s, e)$ *as* $\tau \to \infty$ *(for all* (s, e) *and* $\pi \in \Pi$*), then*

$$v(s, e) = \sup_{a\in\mathcal{B}_{(s,e)}} \left\{r(s, e, a) + \beta\mathbb{E}\left[v\big(T(s, e, a),\ \xi(e)\big)\right]\right\}, \qquad (s, e) \in \Re^{n\times 1} \times \Omega. \tag{29}$$

Proof. Substitute (25) into (29) to confirm that v solves (29). □

Infinite-horizon discounted MDPs with continuous states and actions do not necessarily satisfy (29) without additional restrictions and, when the additional restrictions are imposed, the proofs are technical. For example, the results for infinite-horizon Borel models in chapter 9 in Bertsekas and Shreve (1978) are parsed into three cases that do not include decomposable affine MDPs: nonnegative single-period rewards, nonpositive single-period rewards, and uniformly absolutely bounded single-period rewards. The relatively straightforward proof of Corollary 1 indirectly relies on the finite "state space" of the auxiliary equations.

Let $J_i(e)$ achieve the maximum in (26a):

$$J_i(e) \in \underset{k=1,\ldots,|\mathcal{X}_e^i|}{\arg\max} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f(z) Y^{\mathbb{A}i}(e,z) \right] M^i_{\cdot k}(e) \right\} \tag{30}$$

The next theorem specifies a single-period decision rule A such that stationary policy $(A, A, ...)$ is an optimal infinite-horizon policy.

Theorem 4. *If $\theta < 1$ and $V_\pi^\tau(s,e) \to V_\pi(s,e)$ as $\tau \to \infty$ (for all (s,e) and $\pi \in \Pi$), then $(A, A, \ldots)$ is an optimal infinite-horizon policy, where $A(s,e) = \big(A^1(s,e), A^2(s,e), ..., A^n(s,e)\big)$ and $A^i(s,e)$ is the $J_i(e)$-th extreme point in $\mathcal{X}^i_{(s_i,e)}$ $(i = 1, \ldots, n)$:*

$$A^i(s,e) = M^i_{\cdot,J_i(e)}(e) s_i + c^i(e) 1^i(e). \tag{31}$$

Proof. The proof shows that $\tilde{\pi} = (A, A, \ldots)$ achieves v in (25), i.e., $V_{\tilde{\pi}}(s,e) = v(s,e)$ for all $(s,e) \in \Re_+^{n\times 1} \times \Omega$. The first step of the proof derives an explicit expression for $A(s,e)$ from (31). From $\big\{M^i_{\cdot,J_i(e)}(e), i = 1, \ldots, n\big\}$ and $\big\{c^i(e)1^i(e), i = 1, \ldots, n\big\}$, one can construct a matrix $M^*(e) \in \Re^{m\times n}$ and a vector $c^*(e) \in \Re^{m\times 1}$ such that

$$A(s,e) = M^*(e)s + c^*(e). \tag{32}$$

The second step of the proof uses (32) and specifies the finite-horizon value function $V^\tau_{\tilde{\pi}}(s,e)$ for $\tilde{\pi} = (A, A, \ldots)$ (see Lemma 2). Then Lemma 3 states that if $\theta < 1$, then $V^\tau_{\tilde{\pi}}$ converges to v as $\tau \to \infty$. The proofs of both lemmas are in the Appendix.

Lemma 2. *For $\tau \in \mathbb{N}$, $V^\tau_{\tilde{\pi}}(s,e) = \tilde{f}^\tau(e)s + \tilde{g}^\tau(e)$, where $\tilde{f}^\tau(e) \in \Re^{1\times n}$, $\tilde{g}^\tau(e) \in \Re$, $\tilde{f}^0(\cdot) = \tilde{g}^0(\cdot) \equiv 0$, and*

$$\tilde{f}^\tau(e) = \Big[y^{\mathbb{S}}(e) + y^{\mathbb{A}}(e)M^*(e)\Big] + \beta \sum_{z\in\Omega} p_{ez} \tilde{f}^{\tau-1}(z)\big[Y^{\mathbb{S}}(e,z) + Y^{\mathbb{A}}(e,z)M^*(e)\big], \tag{33a}$$

$$\tilde{g}^\tau(e) = y^o(e) + y^{\mathbb{A}}(e)c^*(e) + \beta \sum_{z\in\Omega} p_{ez} \left[\tilde{f}^{\tau-1}(z)\big[Y^o(e,z) + Y^{\mathbb{A}}(e,z)c^*(e)\big] + \tilde{g}^{\tau-1}(z)\right]. \tag{33b}$$

Lemma 3. *If $\theta < 1$, then $\lim_{\tau\to\infty} V^\tau_{\tilde{\pi}}(s,e) = v(s,e)$ for all $(s,e) \in \Re_+^{n\times} \times \Omega$.*

Lemma 3 implies $\tilde{\pi} \in \Pi$, which implies $V_{\tilde{\pi}}(s,e) = \lim_{\tau\to\infty} V_{\tilde{\pi}}^{\tau}(s,e)$. Since $\lim_{\tau\to\infty} V_{\tilde{\pi}}^{\tau}(s,e) = v(s,e)$ from Lemma 3, $V_{\tilde{\pi}}(s,e) = v(s,e)$, which implies that $(A, A, \ldots)$ achieves v and is optimal. □

As with the finite-horizon criterion, the infinite-horizon value function and optimal policy depend on auxiliary equations (26). The feasibility of employing (26) instead of (29) depends on having algorithms that are efficient and easily programmable. [NS] presents several algorithms to solve (26) that have striking computational advantages over the typical approach of solving discretized approximations of (29).

The auxiliary equations, with both finite- and infinite-horizon criteria, have useful economic interpretations. As exploited in Ning and Sobel (2016) and explained in the applications in [NS], these economic interpretations are associated with qualitative properties of optimal solutions of structured models, and they instill intuition behind complicated mathematical expressions.

4 Partially decomposable affine MDPs

In §2.1, the decomposability and affinity assumptions are imposed on the entire endogenous state and action vectors. However, in many models, such as the commodity procurement and processing example in [NS], only parts of the state and action vectors satisfy these assumptions. This section states results for *partially decomposable affine MDPs*. The proofs are in the Appendix.

An MDP with endogenous state $s \in \Re_{+}^{n\times 1}$, exogenous state $e \in \Omega$, and action $a \in \mathcal{B}_{(s,e)} \subset \Re^{n\times 1}$ is a *partially decomposable affine MDP* if it meets three conditions. First, s and a decompose into $s = (s^d, s^o)$ and $a = (a^d, a^o)$ such that

$$s_{t+1}^{d} | (s_t = s, e_t = e, a_t = a) \sim \boldsymbol{Y}^{\mathbb{S}d}(e) s^d + \boldsymbol{Y}^{\mathbb{A}d}(e) a^d + \boldsymbol{Y}^{od}(s^o, e, a^o), \tag{34}$$

$$s_{t+1}^{o} | (s_t = s, e_t = e, a_t = a) \sim T^o(s^o, e, a^o), \tag{35}$$

$$r(s,e,a) = y^{\mathbb{S}d}(e) s^d + y^{\mathbb{A}d}(e) a^d + y^{od}(s^o, e, a^o). \tag{36}$$

That is, $r(s,e,a)$ and the dynamics of s^d are linear in s^d and a^d, and the dynamics of s^o does not depend on s^d and a^d. The terms in (34)–(36) are matrices and vectors of appropriate dimensions, and $T^o(s^o, e, a^o)$, $\boldsymbol{Y}^{\mathbb{S}d}(e)$, $\boldsymbol{Y}^{\mathbb{A}d}(e)$, and $\boldsymbol{Y}^{od}(s^o, e, a^o)$ are r.v.s with distributions that depend only on the arguments.

Let n_d be the dimension of s^d and let s_i^d denote the i-th component of s^d. The second condition is that $\mathcal{B}_{(s,e)}$ is the product of two sets: $\mathcal{B}_{(s,e)} = \mathcal{B}^d_{(s^d,e,a^o)} \times \mathcal{B}^o_{(s^o,e)}$, such that $a^d \in \mathcal{B}^d_{(s^d,e,a^o)}$ and $a^o \in \mathcal{B}^o_{(s^o,e)}$. Furthermore, $\mathcal{B}^d_{(s^d,e,a^o)}$ is decomposable in the sense of (1), namely, a^d and $\mathcal{B}^d_{(s^d,e,a^o)}$ can be decomposed into n_d subvectors $a^{di}(e)$ and n_d subsets $\mathcal{B}^{di}_{(s_i^d,e,a^o)}$ $(i = 1, \ldots, n_d)$ so that $a^{di}(e) \in \mathcal{B}^{di}_{(s_i^d,e,a^o)}$ and

$$\mathcal{B}^d_{(s^d,e,a^o)} = \times_{i=1}^{n^d} \mathcal{B}^{di}_{(s_i^d,e,a^o)}. \tag{37}$$

The third condition is that each subset $\mathcal{B}^{di}_{(s_i^d,e,a^o)}$ satisfies the compound decomposability and affinity assumption in §2.1 for all $i = 1, \ldots, n_d$. Specifically, $\mathcal{B}^{di}_{(s_i^d,e,a^o)}$ is a polyhedron whose set of extreme points $\mathcal{X}^{di}_{(s_i^d,e,a^o)}$ has cardinality $|\mathcal{X}_e^{d_i}|$ that does not depend on s_i^d and a^o, and

$$\mathcal{X}^{di}_{(s_i^d,e,a^o)} = \left\{ M^{di}_{\cdot k}(e)\left[s_i^d + c_0^{di}(e,a^o)\right] + c^{di}(e,a^o)1^{di}(e),\ k = 1, \ldots, |\mathcal{X}_e^{di}| \right\}, \tag{38}$$

where $M^{di}_{\cdot k}(e)$ is column k of $M^{di}(e)$ and has the same dimension as $a^{di}(e)$, $c_0^{di}(e,a^o) \in \Re_+$, $c^{di}(e,a^o) \in \Re$, and $1^{di}(e)$ is the vector of all ones and has the same dimension as $a^{di}(e)$. Thus, $s_i^d + c_0^{di}(e,a^o) \in \Re_+$ and $M^{di}_{\cdot k}(e)\left[s_i^d + c_0^{di}(e,a^o)\right] + c^{di}(e,a^o)1^{di}(e)$ is referred to as the k-th extreme point in $\mathcal{X}^{di}_{(s_i^d,e,a^o)}$.

Let $\mathbb{E}[\boldsymbol{Y}^{\mathbb{S}d}(e)|\xi(e) = z] = Y^{\mathbb{S}d}(e,z)$, $\mathbb{E}[\boldsymbol{Y}^{\mathbb{A}d}(e)|\xi(e) = z] = Y^{\mathbb{A}d}(e,z)$, and $\mathbb{E}[\boldsymbol{Y}^{od}(s^o,e,a^o)|\xi(e) = z] = Y^{od}(s^o,e,a^o,z)$. As before, let $y^{\mathbb{A}di}(e)$, $\boldsymbol{Y}^{\mathbb{A}di}(e)$, and $Y^{\mathbb{A}di}(e,z)$ denote the parts of $y^{\mathbb{A}d}(e)$, $\boldsymbol{Y}^{\mathbb{A}d}(e)$, and $Y^{\mathbb{A}d}(e,z)$ that multiply $a^{di}(e)$ $(i = 1, \ldots, n_d)$. The following theorem states that the value function of a partially decomposable affine MDP is the sum of two parts, one of which depends only on (s^d, e) and is linear in s^d, and the other depends only on (s^o, e).

Theorem 5. *The finite-horizon value function is*

$$v^\tau(s,e) = f^\tau(e)s^d + g^\tau(e) + u^\tau(s^o,e), \tag{39}$$

where f^τ *and* g^τ *satisfy recursions* $f^0(\cdot) \equiv 0$ *and* $g^0(\cdot) = u^0(\cdot,\cdot) \equiv 0$*, and for* $\tau \in \mathbb{N}$*,*

$$f_i^\tau(e) = y_i^{\mathbb{S}d}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y_{\cdot i}^{\mathbb{S}d}(e,z) + h_i^{\tau-1}(e), \qquad i = 1,\dots,n_d, \tag{40a}$$

$$g^\tau(e) = \beta \sum_{z\in\Omega} p_{ez} g^{\tau-1}(z), \tag{40b}$$

$$h_i^{\tau-1}(e) = \max_{k=1,\dots,|\mathcal{X}_e^{di}|} \left\{ \left[y_i^{\mathbb{A}d}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}di}(e,z) \right] M_{\cdot k}^{di} \right\}. \tag{40c}$$

$$\begin{aligned} u^\tau(s^o,e) = \sup_{a^o \in \mathcal{B}^o_{(s^o,e)}} & \Bigg\{ y^{od}(s^o,e,a^o) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{od}(s^o,e,z,a^o) + \sum_{i=1}^{n_d} h_i^{\tau-1}(e) c_0^{di}(e,a^o) \\ & + \sum_{i=1}^{n_d} \left(c^{di}(e,a^o) \left[y^{\mathbb{A}di}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}di}(e,z) \right] 1^{di}(e) \right) \\ & + \beta \mathbb{E}\left[u^{\tau-1}\left(T^o(s^o,e,a^o), \xi(e)\right) \right] \Bigg\}. \end{aligned} \tag{40d}$$

For $i = 1,\dots,n_d$, $e \in \Omega$, and $\tau \in \mathbb{N}$, let $J^{\tau-1,di}(e)$ denote an an index $k \in \{1,\dots,|\mathcal{X}_e^{di}|\}$ that achieves the maximum in (40c). Assume that the supremum in (40d) is achieved for all $s^o \in \Re^{(n-n_d)\times 1}$, $e \in \Omega$, and $\tau \in \mathbb{N}$, and let $A^{\tau-1,o}(s^o,e)$ denote an a^o that achieves the supremum. The next theorem specifies an optimal single-period decision rule with τ remaining periods.

Theorem 6. *With* τ *periods remaining, an optimal single-period decision rule consists of* $a^o = A^{\tau-1,o}(s^o,e)$ *and* $a^{di}(e) = A^{\tau-1,di}(e)$ $(i = 1,\dots,n_d)$*, where* $A^{\tau-1,di}(e)$ *is the* $J^{\tau-1,di}(e)$*-th extreme point in* $\mathcal{X}^i_{(s_i^d,e,a^o)}$ *in which* $a^o = A^{\tau-1,o}(s^o,e)$*.*

It follows from Theorems 5 and 6 that a partially decomposable affine MDP with action vector (a^d,a^o) and state vector (s^d,s^o,e) is a composite of two related sub-MDPs. One is a decomposable affine MDP with endogenous state $s^d \in \Re_+^{n_d\times 1}$, exogenous state $e \in \Omega$, and action a^d constrained to $\mathcal{B}^d_{(s_d,e,a^o)}$ that is parameterized by a^o. The other is a generic MDP with endogenous state $s^o \in \Re^{(n-n_d)\times 1}$, exogenous state $e \in \Omega$, and action $a^o \in \mathcal{B}^o_{(s^o,e)}$. Theorem 5 states that the value function of the composite MDP is the sum of the value functions of the two sub-MDPs. Theorem 6 states that an optimal policy consists of two parts. One is an extremal optimal policy of the decomposable affine sub-MDP, and the other is an optimal policy for the generic sub-MDP. Therefore, the solution of a partially decomposable affine MDP can be obtained by solving two sub-MDPs, which boils down to solving auxiliary equations (40a)–(40c) with respect to e, and dynamic program (40d) with respect to

(s^o, e).

5 Summary

Many phenomena can be modeled as Markov decision processes (MDPs), but the curse of dimensionality inhibits solutions and realistic applications. The paper characterizes solutions of decomposable affine MDPs with discounted criteria, which can be computed easily and exactly. A decomposable affine MDP has continuous vector-valued actions and endogenous states, exogenous states that follow a finite Markov chain, and decomposable constraints on the actions.

The paper shows that a decomposable affine MDP has a value function that is an affine function of the endogenous state, and it has an extremal optimal policy. The coefficients in the value function and the optimal policy are completely determined by the solution of a small set of auxiliary equations which depend only on the exogenous state. Thus, solving a decomposable affine MDP with continuous endogenous state and action vectors is equivalent to solving the auxiliary equations; this exorcizes the curse of dimensionality.

The paper shows that a partially decomposable affine MDP reduces to a composite of two related smaller sub-MDPs, one of which is a decomposable affine MDP. Thus, solving a partially decomposable affine MDP reduces to solving a set of auxiliary equations and a sub-MDP.

The companion paper [NS] exploits the results in this paper and presents exact algorithms to solve decomposable affine MDPs with discounted criteria. It also gives examples of decomposable affine MDPs and partially decomposable affine MDPs in various contexts.

References

Aström, Karl J. 1970. *Introduction to Stochastic Control Theory*. Academic Press.

Bertsekas, Dimitri P. 1995. *Dynamic Programming and Optimal Control*, vol. I. Athena Scientific.

Bertsekas, Dimitri P., Steven E. Shreve. 1978. *Stochastic Optimal Control: The Discrete Time Case*. Academic Press.

Dantzig, George B. 1963. *Linear Programming and Extensions*. Princeton University Press and the RAND Corporation.

Denardo, Eric V., Uriel G. Rothblum. 1979. Affine dynamic programming. Martin L. Puterman, ed., *Dynamic Programming and Its Applications*. Academic Press, New York, 255–267.

Denardo, Eric V., Uriel G. Rothblum. 1983. Affine structure and invariant policies for dynamic programs. *Mathematics of Operations Research* **8**(3) 342–365.

Heyman, Daniel P., Matthew J. Sobel. 2004. *Stochastic Models in Operations Research, Vol. II: Stochastic Optimization*. Dover Publications.

Mathur, Kamlesh, Daniel Solow. 1994. *Management Science*. Prentice-Hall, Englewood Cliffs, N.J.

Ning, Jie, Matthew J. Sobel. 2016. Production and capacity management with internal financing. Tech. rep., SSRN: https://ssrn.com/abstract=2727940.

Ning, Jie, Matthew J. Sobel. 2017. Easy affine markov decision processes: Algorithms and applications. Tech. rep., SSRN: https://ssrn.com/abstract=2998786.

Powell, Warren B. 2007. *Approximate Dynamic Programming*. John Wiley & Sons, Inc.

Simon, Herbert A. 1956. Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica* **24** 74–81.

Sobel, Matthew J. 1990a. Higher-order and average reward myopic-affine dynamic models. *Mathematics of Operations Research* **15(2)** 299–310.

Sobel, Matthew J. 1990b. Myopic solutions of affine dynamic models. *Operations Research* **38**(2) 847–853.

Theil, Henri. 1957. A note on certainty equivalence in dynamic planning. *Econometrica* **25** 346–349.

Zéphyr, Luckny, Pascal Lang, Bernard F. Lamond, Pascal Côté. 2017. Approximate stochastic dynamic programming for hydroelectric production planning. *European Journal of Operational Research* **to appear**.

A Proofs

Lemma 1. *If $\theta < 1$, for each $e \in \Omega$ there exist $f(e) = \lim_{\tau\to\infty} f^\tau(e)$ and $g(e) = \lim_{\tau\to\infty} g^\tau(e)$ which are the unique solutions of* (26a) *and* (26b)*, respectively.*

Proof. The first step of the proof defines a mapping $\mathcal{L} : \Re^{\omega\times n} \mapsto \Re^{\omega\times n}$ such that (26a) is a fixed point equation. The second step proves that $\mathcal{L}$ is a contraction mapping on $\Re^{\omega\times n}$; hence $f(e) = \lim_{\tau\to\infty} f^\tau(e)$ is the unique solution of (26a). The third step proves that $g(e) = \lim_{\tau\to\infty} g^\tau(e)$ is the unique solution of (26b).

Let $b(e,\cdot)$ denote the e-th row of $b \in \Re^{\omega \times n}$ $(e = 1, \ldots, \omega)$ and define $\mathcal{L} : \Re^{\omega \times n} \mapsto \Re^{\omega \times n}$ with the value of $\mathcal{L}b$ at (e,i) being

$$\mathcal{L}b(e,i) = y_i^{\mathbb{S}}(e) + \beta \sum_{z \in \Omega} p_{ez} b(z,\cdot) Y_{\cdot i}^{\mathbb{S}}(e,z) + \max_{k=1,\ldots,|\mathcal{X}_e^i|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z \in \Omega} p_{ez} b(z,\cdot) Y_{\cdot i}^{\mathbb{A}}(e,z) \right] M_{\cdot k}^i(e) \right\}. \tag{41}$$

Thus, (26a) corresponds to $f = \mathcal{L}f$, where $f \in \Re^{\omega \times n}$. Its value at (e,i) is expressed as $f_i(e)$, and its e-th row is expressed as $f(e) \in \Re^{1 \times n}$.

Define

$$d(b,b') = \max \ \left\{ \left| b(e,i) - b'(e,i) \right| : i = 1, \ldots, n, e = 1, \ldots, \omega \right\}, \qquad b, b' \in \Re^{\omega \times n}. \tag{42}$$

which is a metric for $\Re^{\omega \times n}$. Thus, $(\Re^{\omega \times n}, d)$ is a complete metric space.

The second step proves that $\mathcal{L}$ is a contraction mapping. Define

$$W(k; b, i, e) = y_i^{\mathbb{S}}(e) + \beta \sum_{z \in \Omega} p_{ez} b(z,\cdot) Y_{\cdot i}^{\mathbb{S}}(e,z) + \left[y^{\mathbb{A}i}(e) + \beta \sum_{z \in \Omega} p_{ez} b(z,\cdot) Y^{\mathbb{A}i}(e,z) \right] M_{\cdot k}^i(e), \tag{43}$$

so (41) is $\mathcal{L}b(e,i) = \max\left\{ W(k; b, i, e) : k = 1, \ldots, |\mathcal{X}_e^i| \right\}$. For any $b, b' \in \Re^{\omega \times n}$, $i \in \{1, \ldots, n\}$, and $e \in \Omega$, let $k, k' \in \{1, \ldots, |\mathcal{X}_e^i|\}$ achieve $\mathcal{L}b(e,i) = W(k; b, i, e)$ and $\mathcal{L}b'(e,i) = W(k'; b', i, e)$. Label b

and b' so that $\mathcal{L}b(e,i) \geq \mathcal{L}b'(e,i)$, and use $U^i(e,z)$ defined in (23):

$$\begin{aligned}
0 \leq \mathcal{L}b(e,i) - \mathcal{L}b'(e,i) &= W(k;b,i,e) - W(k';b',i,e) \\
&\leq W(k;b,i,e) - W(k;b',i,e) \\
&= \beta \sum_{z\in\Omega} p_{ez} \left[b(z,\cdot) - b'(z,\cdot)\right] \left[Y^{\mathbb{S}}_{\cdot i}(e,z) + Y^{\mathbb{A}i}(e,z) M^i_{\cdot k}(e)\right] \\
&= \beta \sum_{z\in\Omega} p_{ez} \sum_{j=1}^{n} \left[b(e,j) - b'(e,j)\right] U^i_{jk}(e,z) \\
&\leq \beta \sum_{z\in\Omega} p_{ez} \sum_{j=1}^{n} |b(e,j) - b'(e,j)||U^i_{jk}(e,z)| \\
&\leq \beta \sum_{z\in\Omega} p_{ez} \sum_{j=1}^{n} |U^i_{jk}(e,z)| d(b,b') \\
&\leq \beta \max\left\{ \sum_{z\in\Omega} p_{ez} \sum_{j=1}^{n} |U^i_{jk}(e,z)| : k = 1, \ldots, |\mathcal{X}^i_e|;\ i = 1, ..., n;\ e \in \Omega \right\} d(b,b') \\
&= \theta d(b,b')
\end{aligned}$$

Therefore, $d(\mathcal{L}b, \mathcal{L}b') \leq \theta d(b,b')$ and $\theta < 1$ imply that $\mathcal{L}$ is a contraction mapping on the complete metric space $(\Re^{\omega \times n}, d)$. Banach's fixed point theorem implies the unique existence of $f = \mathcal{L}f = \lim_{\tau\to\infty} \mathcal{L}^\tau f^0$ which is invariant with respect to $f^0 \in \Gamma$. Therefore, $f(e) = \lim_{\tau\to\infty} f^\tau(e)$ which exists for each $e \in \Omega$, and f uniquely solves (26a).

The third step shows that $g(e) = \lim_{\tau\to\infty} g^\tau(e)$ exists for each $e \in \Omega$, and g uniquely solves (26b). The first two steps of this proof and letting $\tau \to \infty$ in (14b) yield

$$\begin{aligned}
\lim_{\tau\to\infty} g^\tau(e) = y^o(e) + \beta \sum_{z\in\Omega} p_{ez} \big[f(z) Y^o(e,z) + \lim_{\tau\to\infty} g^\tau(z) \big] \\
+ \sum_{i=1}^{n} \left[c^i(e) \left(y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f(z) Y^{\mathbb{A}i}(e,z) \right) 1^i(e) \right] \qquad (44)
\end{aligned}$$

which has a unique solution because Ω is finite, $p_{ez} \geq 0$ for all $e, z \in \Omega$, $\sum_{z\in\Omega} p_{ez} = 1$ for all $e \in \Omega$, and $\beta < 1$. Since (44) and (26b) are the same equation, $g(e) = \lim_{\tau\to\infty} g^\tau(e)$ and $g(\cdot)$ uniquely solves (26b). □

Lemma 2. *For $\tau \in \mathbb{N}$, $V^\tau_{\tilde{\pi}}(s,e) = \tilde{f}^\tau(e)s + \tilde{g}^\tau(e)$, where $\tilde{f}^\tau(e) \in \Re^{1\times n}$, $\tilde{g}^\tau(e) \in \Re$, $\tilde{f}^0(\cdot) = \tilde{g}^0(\cdot) \equiv 0$,* *and*

$$\tilde{f}^{\tau}(e) = \Big[y^{\mathbb{S}}(e) + y^{\mathbb{A}}(e)M^*(e)\Big] + \beta \sum_{z\in\Omega} p_{ez}\tilde{f}^{\tau-1}(z)\big[Y^{\mathbb{S}}(e,z) + Y^{\mathbb{A}}(e,z)M^*(e)\big], \tag{45a}$$

$$\tilde{g}^{\tau}(e) = y^{o}(e) + y^{\mathbb{A}}(e)c^*(e) + \beta \sum_{z\in\Omega} p_{ez}\left[\tilde{f}^{\tau-1}(z)\big[Y^{o}(e,z) + Y^{\mathbb{A}}(e,z)c^*(e)\big] + \tilde{g}^{\tau-1}(z)\right] \tag{45b}$$

Proof. Initiate a proof by induction on τ by confirming that at $\tau = 0$, $V^0_{\tilde{\pi}}(\cdot,\cdot) \equiv 0$ satisfies $V^0_{\tilde{\pi}}(s,e) = \tilde{f}^0(e)s + \tilde{g}^0(e)$. For any $\tau \in \mathbb{N}$, if $V^{\tau-1}_{\tilde{\pi}}(s,e) = \tilde{f}^{\tau-1}(e)s + \tilde{g}^{\tau-1}(e)$, then (2a), (2b), and (32) imply

$$\begin{aligned}
V^{\tau}_{\tilde{\pi}}(s,e) &= r\left(s,e,A(s,e)\right) + \beta\mathbb{E}\left[V^{\tau-1}_{\tilde{\pi}}\left(T(s,e,A(s,e)),\xi(e)\right)\right] \\
&= \Big[y^{\mathbb{S}}(e)s + y^{\mathbb{A}}(e)\big(M^*(e)s + c^*(e)\big) + y^{o}(e)\Big] + \\
&\qquad \beta\mathbb{E}\left[\tilde{f}^{\tau-1}(\xi(e))\left(\boldsymbol{Y}^{\mathbb{S}}(e)s + \boldsymbol{Y}^{\mathbb{A}}(e)\big(M^*(e)s + c^*(e)\big) + \boldsymbol{Y}^{o}(e)\right) + \tilde{g}^{\tau-1}(\xi(e))\right] \\
&= \left(y^{o}(e) + y^{\mathbb{A}}(e)c^*(e) + \beta\sum_{z\in\Omega} p_{ez}\left[\tilde{f}^{\tau-1}(z)\big[Y^{o}(e,z) + Y^{\mathbb{A}}(e,z)c^*(e)\big] + \tilde{g}^{\tau-1}(z)\right]\right) + \\
&\qquad \left([y^{\mathbb{S}}(e) + y^{\mathbb{A}}(e)M^*(e)] + \beta\sum_{z\in\Omega} p_{ez}\tilde{f}^{\tau-1}(z)\left[Y^{\mathbb{S}}(e,z) + Y^{\mathbb{A}}(e,z)M^*(e)\right]\right)s.
\end{aligned}$$

The last equality implies $V^{\tau}_{\tilde{\pi}}(s,e) = \tilde{f}^{\tau}(e)s + \tilde{g}^{\tau}(e)$, where $\tilde{f}^{\tau}$ and $\tilde{g}^{\tau}$ satisfy (45) which completes the induction. □

Lemma 3. *If $\theta < 1$, then $\lim_{\tau\to\infty} V^{\tau}_{\tilde{\pi}}(s,e) = v(s,e)$ for all $(s,e) \in \Re^{n\times 1}_{+} \times \Omega$.*

Proof. Adapting the first two steps of the proof of Lemma 1 to (45a) establishes that, if $\theta < 1$, then $\tilde{f}(e) = \lim_{\tau\to\infty} \tilde{f}^{\tau}(e)$ exists and uniquely solves

$$\tilde{f}(e) = \Big[y^{\mathbb{S}}(e) + y^{\mathbb{A}}(e)M^*(e)\Big] + \beta\sum_{z\in\Omega} p_{ez}\tilde{f}(z)\left[Y^{\mathbb{S}}(e,z) + Y^{\mathbb{A}}(e,z)M^*(e)\right]. \tag{46}$$

However, f uniquely solves (46) (Theorem 3), so $\tilde{f} = f$.

Similarly, $\tilde{g}(e) = \lim_{\tau\to\infty} \tilde{g}^{\tau}(e)$ exists and uniquely solves

$$\tilde{g}(e) = y^{o}(e) + y^{\mathbb{A}}(e)c^*(e) + \beta\sum_{z\in\Omega} p_{ez}f(z)Y^{o}(e,z) + \beta\sum_{z\in\Omega} p_{ez}f(z)Y^{\mathbb{A}}(e,z)c^*(e) + \beta\sum_{z\in\Omega} p_{ez}\tilde{g}(z) \tag{47}$$

whose unique solution is g (Theorem 3), so $\tilde{g} = g$.

Take limits on both sides of $V_{\tilde{\pi}}^{\tau}(s,e) = \tilde{f}^{\tau}(e)s + \tilde{g}^{\tau}(e)$ (Lemma 2) to obtain $\lim_{\tau\to\infty} V_{\tilde{\pi}}^{\tau}(s,e) = f(e)s + g(e) = v(s,e)$. □

Theorem 5. *The finite-horizon value function is*

$$v^{\tau}(s,e) = f^{\tau}(e)s^{d} + g^{\tau}(e) + u^{\tau}(s^{o},e), \tag{39}$$

where f^{τ} and g^{τ} satisfy recursions $f^{0}(\cdot) \equiv 0$ and $g^{0}(\cdot) = u^{0}(\cdot,\cdot) \equiv 0$, and for $\tau \in \mathbb{N}$,

$$f_i^{\tau}(e) = y_i^{\mathbb{S}d}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y_{\cdot i}^{\mathbb{S}d}(e,z) + h_i^{\tau-1}(e), \qquad i = 1,\dots,n_d, \tag{40a}$$

$$g^{\tau}(e) = \beta \sum_{z\in\Omega} p_{ez} g^{\tau-1}(z), \tag{40b}$$

$$h_i^{\tau-1}(e) = \max_{k=1,\dots,|\mathcal{X}_e^{di}|} \left\{ \left[y_i^{\mathbb{A}d}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}di}(e,z) \right] M_{\cdot k}^{di} \right\}, \tag{40c}$$

$$\begin{aligned} u^{\tau}(s^{o},e) = \sup_{a^{o}\in\mathcal{B}^{o}_{(s^{o},e)}} & \Bigg\{ y^{od}(s^{o},e,a^{o}) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{od}(s^{o},e,z,a^{o}) + \sum_{i=1}^{n_d} h_i^{\tau-1}(e) c_0^{di}(e,a^{o}) \\ & + \sum_{i=1}^{n_d} \left(c^{di}(e,a^{o}) \left[y^{\mathbb{A}di}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}di}(e,z) \right] 1^{di}(e) \right) \\ & + \beta \mathbb{E}\left[u^{\tau-1}\big(T^{o}(s^{o},e,a^{o}),\xi(e)\big)\right] \Bigg\}. \end{aligned} \tag{40d}$$

Proof. Initiate a proof of (39) and (40) by induction on τ by confirming (39) at $\tau = 0$ with $v^{0}(\cdot,\cdot) \equiv 0$, $f^{0}(\cdot) \equiv 0$, $g^{0}(\cdot) \equiv 0$, and $u^{0}(\cdot,\cdot) \equiv 0$. For any $\tau \in \mathbb{N}$, if (39) is valid for all $(s,e) \in \Re_{+}^{n\times 1} \times \Omega$, then (34)–(36) and dynamic program (12) yield

$$v^{\tau}(s,e) = y^{\mathbb{S}d}(e)s^{d} + \sup_{a\in\mathcal{B}_{(s,e)}} \left\{ \sum_{i=1}^{n_d} y^{\mathbb{A}di}(e)a^{di}(e) + y^{od}(s^{o},e,a^{o}) + W^{\tau-1}(a;s,e) \right\} \tag{49a}$$

$$\begin{aligned} \text{where} \quad W^{\tau-1}(a;s,e) = \beta \mathbb{E}\Bigg[f^{\tau-1}(\xi(e)) \Bigg[\boldsymbol{Y}^{\mathbb{S}d}(e)s^{d} + \sum_{i=1}^{n_d} \boldsymbol{Y}^{\mathbb{A}di}(e)a^{di}(e) + \boldsymbol{Y}^{od}(s^{o},e,a^{o}) \Bigg] + g^{\tau-1}(\xi(e)) \Bigg] \\ + \beta \mathbb{E}[u^{\tau-1}(T^{o}(s^{o},e,a^{o}),\xi(e))]. \end{aligned} \tag{49b}$$

Using $\mathbb{E}[\boldsymbol{Y}^{\mathbb{S}d}(e)|\xi(e) = z] = Y^{\mathbb{S}d}(e,z)$, $\mathbb{E}[\boldsymbol{Y}^{\mathbb{A}di}(e)|\xi(e) = z] = Y^{\mathbb{A}di}(e,z)$, and $\mathbb{E}[\boldsymbol{Y}^{od}(s^{o},e,a^{o})|\xi(e) =$

$z] = Y^{od}(s^o, e, a^o, z)$, (49b) becomes

$$W^{\tau-1}(a; s, e) = \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) \Big[Y^{\mathbb{S}d}(e, z)s^d + \sum_{i=1}^{n_d} Y^{\mathbb{A}di}(e, z)a^{di}(e) + Y^{od}(s^o, e, z, a^o) \Big]$$
$$+ \beta \sum_{z\in\Omega} p_{ez} g^{\tau-1}(z) + \beta \mathbb{E}[u^{\tau-1}(T^o(s^o, e, a^o), \xi(e))]. \tag{50}$$

Using (50) in (49a) yields

$$v^\tau(s, e) = \beta \sum_{z\in\Omega} p_{ez} g^{\tau-1}(z) + y^{\mathbb{S}d}(e)s^d + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{S}d}(e, z)s^d$$
$$+ \sup_{a\in\mathcal{B}_{(s,e)}} \Big\{ W^{\tau-1,d}(a^d; e) + W^{\tau-1,o}(a^o; s, e) \Big\}, \tag{51a}$$

$$\text{where } W^{\tau-1,d}(a^d; e) = \sum_{i=1}^{n_d} y^{\mathbb{A}di}(e)a^{di}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) \sum_{i=1}^{n_d} Y^{\mathbb{A}di}(e, z)a^{di}(e), \tag{51b}$$

$$W^{\tau-1,o}(a^o; s^o, e) = y^{od}(s^o, e, a^o) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{od}(s^o, e, z, a^o) + \beta \mathbb{E}\big[u^{\tau-1}(T^o(s^o, e, a^o), \xi(e))\big]. \tag{51c}$$

Since $a^d \in \mathcal{B}_{(s^d,e,a^o)}$ and $a^o \in \mathcal{B}^o_{(s^o,e)}$,

$$\sup_{a\in\mathcal{B}_{(s,e)}} \Big\{ W^{\tau-1,d}(a^d; e) + W^{\tau-1,o}(a^o; s, e) \Big\} = \sup_{a^o\in\mathcal{B}^o_{(s^o,e)}} \Big\{ LP^{\tau-1,d}(a^o, s^d, e) + W^{\tau-1,o}(a^o; s^o, e) \Big\}, \tag{52a}$$

$$\text{where} \qquad LP^{\tau-1,d}(a^o, s^d, e) = \sup_{a^d\in\mathcal{B}^d_{(s^d,e,a^o)}} W^{\tau-1,d}(a^d; e). \tag{52b}$$

The right side of (52b) is a linear program with decision variable a^d because $W^{\tau-1,d}(a^d; e)$ is linear in a^d. Also, $\mathcal{B}^d_{(s^d,e,a^o)}$ is the product of polyhedra $\mathcal{B}^{di}_{(s^d_i,e,a^o)}$ with extreme points that satisfy (38). Thus, $s^d_i + c^{di}_0(e, a^o) \geq 0$ and arguments similar to those in the proof of Theorem 1 imply

$$LP^{\tau-1,d}(a^o, s^d, e) = \sum_{i=1}^{n_d} LP^{\tau-1,d}_i(a^o, s^d, e), \tag{53a}$$

$$\text{where } LP^{\tau-1,d}_i(a^o, s^d, e) = \Bigg[\max_{k=1,\dots,|\mathcal{X}^{di}_e|} \Bigg\{ \Big[y^{\mathbb{A}di}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}di}(e, z) \Big] M^{di}_{\cdot k}(e) \Bigg\} \Bigg] \times \big[s^d_i + c^{di}_0(e, a^o) \big]$$
$$+ \Big[y^{\mathbb{A}di}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}di}(e, z) \Big] \times 1^{di}(e) \times c^{di}(e, a^o). \tag{53b}$$

For vector w, let $Diag(w)$ denote the diagonal matrix with the components of w on its diagonals. Combine (51)–(53) and use $h^{\tau-1}(e) = (h_i^{\tau-1}(e), i = 1, \ldots, n_d)$ in (40c) to obtain

$$v^\tau(s,e) = \beta \sum_{z\in\Omega} p_{ez} g^{\tau-1}(z) + y^{\mathbb{S}d}(e)s^d + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{S}d}(e,z)s^d + Diag\big(h^{\tau-1}(e)\big)s^d \tag{54a}$$

$$+ \sup_{a^o \in \mathcal{B}^o_{(s^o,e)}} \Bigg\{ W^{\tau-1,o}(a^o; s^o, e) + \sum_{i=1}^{n_d} h_i^{\tau-1}(e) c_0^{di}(e, a^o)$$

$$+ \sum_{i=1}^{n_d} \Bigg(c^{di}(e,a^o) \Big[y^{\mathbb{A}di}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}di}(e,z) \Big] 1^{di}(e) \Bigg) \Bigg\}. \tag{54b}$$

Collect terms and use (51c) to obtain (39) and (40), which completes the induction. □

Theorem 6. *With τ periods remaining, an optimal single-period decision rule consists of $a^o = A^{\tau-1,o}(s^o,e)$ and $a^{di}(e) = A^{\tau-1,di}(e)$ $(i = 1, \ldots, n_d)$, where $A^{\tau-1,di}(e)$ is the $J^{\tau-1,di}(e)$-th extreme point in $\mathcal{X}^i_{(s_i^d,e,a^o)}$ with $a^0 = A^{\tau-1,o}(s^o,e)$.*

Proof. Use v^τ in (39) and arguments similar to those in the proof of Theorem 2 to conclude that $a^{di}(e) = A^{\tau-1,di}(e)$ optimizes linear program (52b), and $a^o = A^{\tau-1,o}(s^o,e)$ is optimal in (40d). □