INFORMS TutORials in Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Easy Affine Markov Decision Processes: Properties and Applications

Jie Ning

To cite this entry: Jie Ning. Easy Affine Markov Decision Processes: Properties and Applications. *In* INFORMS TutORials in Operations Research. Published online: 03 Oct 2017; 28-47.
https://doi.org/10.1287/educ.2017.0170

TUTORIALS IN
OPERATIONS RESEARCH
INFORMS 2017

 | ISBN 978-0-9906153-0-9
https://doi.org/10.1287/educ.2017.0170

Easy Affine Markov Decision Processes: Properties and Applications

Jie Ning
Weatherhead School of Management, Case Western Reserve University, Cleveland, Ohio 44106,
jie.ning@case.edu

Abstract This tutorial introduces a class of *decomposable affine Markov decision processes* (MDPs) that have continuous multidimensional endogenous states and actions, and an exogenous state that follows an exogenous Markov chain. We show that, unlike most MDPs with continuous state and actions, decomposable affine MDPs are free of the curse of dimensionality and can be solved easily and exactly. These nice properties are attributed to its affine dynamics and affine single-period rewards, its *decomposable* action space, and the polyhedral features of the decomposed action space. Exploiting its structure, we demonstrate that a decomposable affine MDP with a finite-horizon criterion has a value function that is affine in the endogenous state and has an *extremal* optimal policy; the value function and the extremal optimal policy are determined by the solution of a set of *auxiliary equations*. At the end of the tutorial, we illustrate the potential applicability of decomposable affine MDPs using the examples of fishery management and dynamic capacity portfolio management.

Keywords Markov decision process; continuous states and actions; multidimensional states and actions; affine value function; extremal policy; curse of dimensionality

1. Introduction

A Markov decision process (MDP) is a natural tool for analyzing sequential decision making in discrete time, but its application to problems with continuous multidimensional state and actions is significantly limited. This is because it is generally extremely challenging to solve such an MDP, either analytically or numerically, unless the dimensions of the state and action vectors are very small.

In this tutorial, we introduce the class of *decomposable affine MDPs*, which have a continuous vector-valued endogenous (i.e., controlled) state, an exogenous state that follows an exogenous Markov chain, and continuous vector-valued actions. We consider such an MDP with a finite-horizon criterion and show that it can be solved easily and exactly, regardless of the dimensions of the endogenous state and action. These results are made possible by the nice structure of decomposable affine MDPs—namely, affine dynamics and affine single-period reward, *decomposable* action space (explained in Sections 2.3 and 3), and polyhedral features of the decomposed action space. We explain the roles of these properties in ensuring the tractability of decomposable affine MDPs and show that they give rise to a value function that is an affine function of the endogenous state as well as an *extremal* optimal policy that is also affine in the endogenous state.

In the MDP literature, there have been significant efforts and progress in tackling MDPs with continuous multidimensional state and actions. One stream of studies focuses on developing generic algorithms that provide good approximate solutions (see Powell [10], Zéphyr et al. [17] for the approximation literature). This tutorial belongs to another stream, one that identifies special classes of MDPs with nice structural properties that allow easy and exact analytical solutions.

A well-known class of MDPs that have nice structural properties and admit analytical solutions is the linear-quadratic-Gaussian (LQG) models. They have *linear* dynamics, concave *quadratic* immediate rewards, and normally distributed random elements (*Gaussian*). These features lead to an optimal policy that iss:exts an affine function of the state and a value function that is a quadratic function of the state. While much of the LQG literature studies continuous-time models, the seminal papers (Simon [11], Theil [15]) consider discrete-time models.

Several classes of MDPs have the property that a myopic policy is optimal; see Denardo and Rothblum [4, 5], Sobel [12, 13], and the references therein. These papers show that, with proper reformulation, the intertemporal dependence in such an MDP can be eliminated, and the original dynamic problem reduces to a single-period static problem. Thus, a myopic policy is optimal and can be easily computed.

A decomposable affine MDP is similar to the MDPs that admit a myopic optimal policy in the following sense. The affinity of its value function and an optimal policy implies that the solution to a decomposable affine MDP is fully specified if the coefficients are known. As we will see in this tutorial, under a finite-horizon criterion, these coefficients are solutions of a set of recursive *auxiliary equations* that depend only on the exogenous state. Thus, the exogenous state can be viewed as the core of a decomposable affine MDP, which plays a key role in determining the value function and an optimal policy. The endogenous state, on the other hand, is more peripheral in the affine functions. This is similar in spirit to myopic MDPs, whose intertemporal dependence disappears after reformulation. The dependence of a decomposable affine MDP on the endogenous state is suppressed after exploiting its structural properties. A solution is fully specified by solving the auxiliary equations that are indexed by the exogenous state, instead of the dynamic program whose indexes span the entire state space.

The remainder of the tutorial begins with a simple example in Section 2 that illustrates the key features and results of decomposable affine MDPs. Then, in Section 3, we build on the insights from Section 2 and present the general model of decomposable affine MDPs and general results. In Section 4, we present applications of decomposable affine MDPs in fishery management and dynamic capacity portfolio management. We conclude the tutorial in Section 5. Interested readers are referred to Ning and Sobel [8, 9] for an in-depth analysis of decomposable affine MDPs with finite- and infinite-horizon criteria with discounting.

2. Decomposable Affine MDPs: An Example

In this section, we present and analyze a simple example and its extension to illustrate the crux of decomposable affine MDPs. We formulate the model in Section 2.1. In Section 2.2, we analyze the model and show how its key features give rise to an affine value function and an extremal optimal policy. In Section 2.3, we introduce an extension to the model and discuss important features of the extension that preserve the structure of the earlier results. The insights developed in this section will be used in Section 3 to specify and solve a general decomposable affine MDP.

2.1. Basic Model

Consider a firm that makes two types of products and faces stochastic market prices. Each period t, the firm observes the market prices and the available production capacity, and chooses the production quantities for the two products subject to the following considerations. First, a unit of production capacity can make at most a unit of product 1 or a unit of product 2. Thus, the total amount of the two products made cannot exceed the available production capacity. Let K_t denote the capacity level in period t, and let q_{jt} denote the production quantity of product j $(j = 1, 2)$. Then

$$q_{1t} + q_{2t} \leq K_t. \tag{1}$$

Second, each period the firm wants to use at least fraction $\alpha \in (0,1)$ of the capacity to maintain a relatively stable workforce:

$$q_{1t} + q_{2t} \geq \alpha K_t. \tag{2}$$

Third, the capacity depreciates as a result of production wear and tear. Given production quantities (q_{1t}, q_{2t}), K_t units of capacity depreciate to $K_t - \lambda_1 q_{1t} - \lambda_2 q_{2t}$, where $\lambda_j \in (0,1]$ reflects the depreciation caused by making product j $(j = 1,2)$. Thus, the firm needs to account for the depreciation consequence when choosing the production quantities.

Because of depreciation, the capacity level at the beginning of period $t+1$ is

$$K_{t+1} = K_t - \lambda_1 q_{1t} - \lambda_2 q_{2t}. \tag{3}$$

Let p_{jt} denote the profit margin (i.e., market price net of the production cost) of product j in period t $(j = 1,2)$. Assume that the process $\{(p_{1t}, p_{2t})\colon t = 1,2,\ldots\}$ is a Markov chain. Given (p_{1t}, p_{2t}), there is a random vector $W(p_{1t}, p_{2t})$ with a probability distribution that depends only on (p_{1t}, p_{2t}) such that the conditional distribution of $(p_{1,t+1}, p_{2,t+1})$ is the same as the probability distribution of $W(p_{1t}, p_{2t})$: $(p_{1,t+1}, p_{2,t+1}) \,|\, (p_{1t}, p_{2t}) \sim W(p_{1t}, p_{2t})$.

Assume that the firm makes the production decision over T periods to maximize the expected value of the present value (EPV) of its earnings. Let β denote the single-period discount factor, and let $\mathbb{E}$ denote the expectation operator. The firm needs to solve the following MDP:

$$\max_{\{(q_{1t}, q_{2t})\colon t=1,\ldots,T\}} \mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1}(p_{1t}q_{1t} + p_{2t}q_{2t})\right] \tag{4a}$$

subject to, for $t = 1,\ldots,T$,

$$q_{1t} + q_{2t} \leq K_t, \tag{4b}$$
$$q_{1t} + q_{2t} \geq \alpha K_t, \tag{4c}$$
$$q_{1t},\ q_{2t} \geq 0, \tag{4d}$$
$$K_{t+1} = K_t - \lambda_1 q_{1t} - \lambda_2 q_{2t}, \tag{4e}$$
$$(p_{1,t+1}, p_{2,t+1}) \,|\, (p_{1t}, p_{2t}) \sim W(p_{1t}, p_{2t}). \tag{4f}$$

2.2. Key Features of the MDP, Value Function, and Optimal Policy

The MDP in (4) has three state variables: the endogenous state variable $K_t \geq 0$ and the exogenous state variables p_{1t} and p_{2t}. The two action variables are q_{1t} and q_{2t}. Let $v_T(K, p_1, p_2)$ denote the optimal value of (4a) given the initial state $(K_1 = K, p_{11} = p_1, p_{21} = p_2)$—namely,

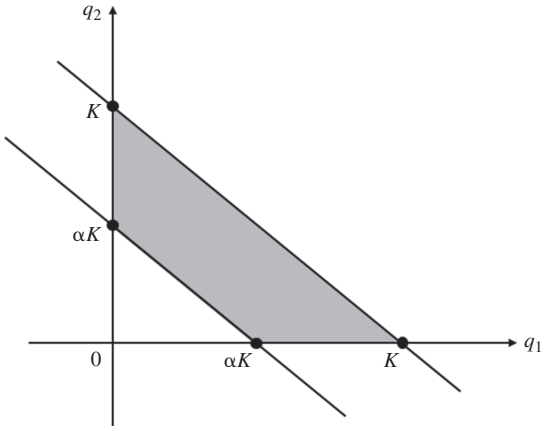
$$v_T(K, p_1, p_2) = \max \mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1}(p_{1t}q_{1t} + p_{2t}q_{2t}) \,\middle|\, (K_1, p_{11}, p_{21}) = (K, p_1, p_2)\right]. \tag{5}$$

Similarly, let $v_n(K, p_1, p_2)$ denote the optimal objective value of (4) with planning horizon $n = 1,\ldots,T-1$ and initial state $(K_1, p_{11}, p_{21}) = (K, p_1, p_2)$. Thus, for $n = 1,\ldots,T$, v_n and v_{n-1} satisfy the following dynamic program:

$$v_n(K, p_1, p_2) = \max_{\substack{(q_1, q_2)\colon q_1, q_2 \geq 0 \\ q_1+q_2 \leq K, \\ q_1+q_2 \geq \alpha K}} \left\{p_1 q_1 + p_2 q_2 + \beta\, \mathbb{E}\left[v_{n-1}(K - \lambda_1 q_1 - \lambda_2 q_2, W(p_1, p_2))\right]\right\}, \tag{6}$$

where $v_0(\cdot,\cdot,\cdot) \equiv 0$. Henceforth, v_n is termed the value function of the MDP with n periods remaining in the horizon.

FIGURE 1. Feasible set of actions (q_1, q_2) given states (K, p_1, p_2).



Given state (K, p_1, p_2), let $(Q_{n1}(K, p_1, p_2), Q_{n2}(K, p_1, p_2))$ denote an optimal solution for (q_1, q_2) in dynamic program (6). Functions (Q_{n1}, Q_{n2}) comprise an optimal single-period decision rule with n periods remaining in the horizon. Arrange the optimal single-period decision rules in clock time as $\{(Q_{n1}, Q_{n2})\colon n = T, T-1, \ldots, 1\}$; this sequence is an optimal policy for the MDP.

Before solving for its value function and an optimal policy, let us first take a closer look at this MDP. It has several important features. First, the single-period reward, $p_1 q_1 + p_2 q_2$, is an affine function of the action (q_1, q_2) and the endogenous state K. Second, the dynamical equation for the endogenous state K_t is an affine function of the endogenous state and action (see (3)). Third, the constraints on the action are affine functions of the endogenous state and action (see (1) and (2)). Fourth, given state (K, p_1, p_2), the set of feasible actions is a bounded polyhedron (see Figure 1). Fifth, for all $K \geq 0$, this polyhedron always has four extreme points. Sixth, the coordinates of the four extreme points, $\{(\alpha K, 0), (0, \alpha K), (K, 0), (0, K)\}$, are linear functions of K for all $K \geq 0$. The last two features are particularly noteworthy because typically, the number of the extreme points varies with the parameters of the feasible region—in this case, K. Furthermore, the coordinates of the extreme points are typically piecewise linear, rather than linear, in the parameter K.

Given the affine single-period reward, affine dynamics, and affine constraints on actions, one might guess that the value function is affine. Next we show that while this intuition is indeed well founded, all six aforementioned features play an important role. Specifically, without the features of the feasible region and its extreme points, the affinity of the value function would break down.

We want to prove that the value function is affine in the endogenous state. That is, for all $n = 0, 1, \ldots, T$, there exist functions f_n and g_n of (p_1, p_2), such that

$$v_n(K, p_1, p_2) = f_n(p_1, p_2) K + g_n(p_1, p_2). \tag{7}$$

We shall prove (7) by induction. Because $v_0(\cdot, \cdot, \cdot) \equiv 0$, (7) is true for $n = 0$ with $f_0(\cdot, \cdot) = g_0(\cdot, \cdot) \equiv 0$. This initiates the induction.

Assume that (7) is true for $n-1$. Dynamic program (6) and the inductive assumption about v_{n-1} imply

$$\begin{aligned} &v_n(K, p_1, p_2) \\ &= \max_{\substack{(q_1, q_2)\colon q_1, q_2 \geq 0 \\ q_1 + q_2 \leq K, \\ q_1 + q_2 \geq \alpha K}} \left\{ p_1 q_1 + p_2 q_2 + \beta \, \mathbb{E}\left[f_{n-1}(W(p_1, p_2))(K - \lambda_1 q_1 - \lambda_2 q_2) + g_{n-1}(W(p_1, p_2)) \right] \right\} \end{aligned}$$

$$= \beta\,\mathbb{E}[f_{n-1}(W(p_1,p_2))]K + \beta\,\mathbb{E}[g_{n-1}(W(p_1,p_2))] \tag{8a}$$

$$+ \max_{\substack{(q_1,q_2):\, q_1,q_2\geq 0\\ q_1+q_2\leq K,\\ q_1+q_2\geq \alpha K}} \{p_1q_1 + p_2q_2 - \beta\,\mathbb{E}[f_{n-1}(W(p_1,p_2))](\lambda_1 q_1 + \lambda_2 q_2)\}. \tag{8b}$$

The optimization in (8b) is a linear program with decision variables (q_1, q_2) and the feasible region shown in Figure 1. Its optimal objective function value is achieved at an extreme point. Thus,

$$\begin{aligned} v_n(K,p_1,p_2) = {} & \beta\,\mathbb{E}[f_{n-1}(W(p_1,p_2))]K + \beta\,\mathbb{E}[g_{n-1}(W(p_1,p_2))] \\ & + \max\{(p_1 - \beta\lambda_1\,\mathbb{E}[f_{n-1}(W(p_1,p_2))])\alpha K, \\ & \quad (p_2 - \beta\lambda_2\,\mathbb{E}[f_{n-1}(W(p_1,p_2))])\alpha K, \\ & \quad (p_1 - \beta\lambda_1\,\mathbb{E}[f_{n-1}(W(p_1,p_2))])K, \\ & \quad (p_2 - \beta\lambda_2\,\mathbb{E}[f_{n-1}(W(p_1,p_2))])K\}, \end{aligned} \tag{9}$$

where the four maximands correspond to the four extreme points $\{(\alpha K, 0), (0, \alpha K), (K, 0), (0, K)\}$.

Because $K \geq 0$, we can pull K out of the maximization in (9) and compare only the coefficients. Thus, we can collect terms to obtain

$$v_n(K,p_1,p_2) = f_n(p_1,p_2)K + g_n(p_1,p_2), \tag{10a}$$

where

$$\begin{aligned} f_n(p_1,p_2) = {} & \beta\mathbb{E}[f_{n-1}(W(p_1,p_2))] \\ & + \max\{(p_1 - \beta\lambda_1\mathbb{E}[f_{n-1}(W(p_1,p_2))])\alpha,\ (p_2 - \beta\lambda_2\mathbb{E}[f_{n-1}(W(p_1,p_2))])\alpha, \\ & \quad (p_1 - \beta\lambda_1\mathbb{E}[f_{n-1}(W(p_1,p_2))]),\ (p_2 - \beta\lambda_2\mathbb{E}[f_{n-1}(W(p_1,p_2))])\}, \end{aligned} \tag{10b}$$

$$g_n(p_1,p_2) = \beta\mathbb{E}[g_{n-1}(W(p_1,p_2))]. \tag{10c}$$

Thus, the induction continues, and the value function satisfies (7) for $n = 0, 1, \ldots, T$. Furthermore, functions f_n and g_n satisfy the recursion in (10b) and (10c) with $f_0(\cdot,\cdot) = g_0(\cdot,\cdot) \equiv 0$. This implies that $g_n(\cdot,\cdot) \equiv 0$ for all $n = 0, 1, \ldots, T$. Henceforth, we shall refer to (10b) and (10c), which determine the coefficients of the value function, as the *auxiliary equations*.

In addition to proving the affinity of v_n, the preceding analysis also implies an optimal single-period decision rule. From (8), an optimal (p_1, p_2) solves the linear program (8b). Thus, given state (K, p_1, p_2), an optimal single-period decision rule (Q_{n1}, Q_{n2}) sets (q_1, q_2) to an optimal extremal point of the polyhedron. Henceforth, such a decision rule is termed *extremal*, and the subsequent policy is termed an *extremal* policy. Because the extreme points are linear in the endogenous state K, an extremal policy is also an affine policy.

Note that the optimality of an extreme point in (8b) implies that its corresponding maximand achieves the optimum in the auxiliary equation (10b) for f_n. The reverse statement is also true. Thus, the solution to the auxiliary equations determines not only the coefficients in the affine value function but also the optimality of an extreme point in the extremal decision rule.

The following proposition summarizes these results.

Proposition 1. (1) *For $n = 1, \ldots, T$, the value function satisfies*

$$v_n(K,p_1,p_2) = f_n(p_1,p_2)K + g_n(p_1,p_2),$$

where f_n and g_n satisfy the recursive auxiliary equations (10b) and (10c) with $f_0(\cdot,\cdot) = g_0(\cdot,\cdot) \equiv 0$.

(2) *With n periods remaining, an optimal single-period decision rule (Q_{n1}, Q_{n2}) sets (q_1, q_2) to the extreme point that corresponds to an optimal maximand in (10b).*

We conclude this subsection by discussing how the aforementioned six features of the MDP ensure an affine value function and an extremal optimal policy. From the induction, the affine reward and affine dynamics ensure that the objective function of the optimization is affine in (6). Together with the affine constraints, they ensure that the optimization is a linear program. The boundedness of the polyhedron then ensures that there always exists an optimal solution to the linear program. The fact that the number of extreme points is invariant with respect to K ensures that the number of maximands in (9) does not depend on K. Finally, the linearity of the extreme points in K and the nonnegativity of K allow us to pull K out of the maximization in (9) and compare only the coefficients. This completes the induction and ensures that if v_{n-1} is affine, then v_n is affine.

2.3. Extension: Capacity Expansion

The model presented in Section 2.1 and analyzed in Section 2.2 is simple in that we have only one endogenous state variable. In this subsection, we extend the model by incorporating a second endogenous state variable and discuss conditions that preserve the affinity of the value function and the extremal feature of an optimal policy.

Assume that the firm in Section 2.1 can spend up to B units of cash to expand its production capacity over the T-period planning horizon. Let i_t denote the amount of capacity investment in period t, and let b_t denote the amount of budget available for investment at the beginning of period t. Then $b_1 = B$, and

$$0 \le i_t \le b_t, \quad b_{t+1} = b_t - i_t. \tag{11}$$

Let y_t denote the amount of capacity installed per unit of cash investment in period t—namely, the investment yield. Assume that the newly installed capacity in period t can be used for production in period $t+1$. Thus, the production decisions are subject to the same constraints as before, $q_{1t} + q_{2t} \le K_t$ and $q_{1t} + q_{2t} \ge \alpha K_t$. The dynamics of the capacity changes to

$$K_{t+1} = K_t - \lambda_1 q_{1t} - \lambda_2 q_{2t} + y_t i_t. \tag{12}$$

Similar to the profit margins (p_{1t}, p_{2t}), the investment yield y_t is exogenous. Assume that the process $\{(p_{1t}, p_{2t}, y_t)\colon t = 1, 2, \ldots\}$ is an exogenous Markov chain. Given (p_{1t}, p_{2t}, y_t), there is a random vector $Z(p_{1t}, p_{2t}, y_t)$ with a probability distribution that depends only on (p_{1t}, p_{2t}, y_t) such that the condition distribution of $(p_{1,t+1}, p_{2,t+1}, y_{t+1})$ is the same as the distribution of $Z(p_{1t}, p_{2t}, y_t)$: $(p_{1,t+1}, p_{2,t+1}, y_{t+1}) \,|\, (p_{1t}, p_{2t}, y_t) \sim Z(p_{1t}, p_{2t}, y_t)$.

Assume that any investment budget remaining at the end of the planning horizon is lost. The goal of the firm is to make production and investment decisions to maximize the EPV of its earnings: $\mathbb{E}[\sum_{t=1}^{T} \beta^{t-1}(p_{1t}q_{1t} + p_{2t}q_{2t})]$. The problem is the following MDP:

$$\max_{\{(q_{1t}, q_{2t}, i_t)\colon t=1,\ldots,T\}} \mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1}(p_{1t}q_{1t} + p_{2t}q_{2t})\right] \tag{13a}$$

subject to, for $t = 1, \ldots, T$,

$$q_{1t} + q_{2t} \le K_t, \tag{13b}$$
$$q_{1t} + q_{2t} \ge \alpha K_t, \tag{13c}$$
$$i_t \le b_t, \tag{13d}$$
$$q_{1t},\ q_{2t},\ i_t \ge 0, \tag{13e}$$
$$K_{t+1} = K_t - \lambda_1 q_{1t} - \lambda_2 q_{2t} + y_t i_t, \tag{13f}$$
$$b_{t+1} = b_t - i_t, \tag{13g}$$
$$(p_{1,t+1}, p_{2,t+1}, y_{t+1}) \,|\, (p_{1t}, p_{2t}, y_t) \sim Z(p_{1t}, p_{2t}, y_t). \tag{13h}$$

2.3.1. Key Features of the MDP. The MDP in (13) has two endogenous state variables, K_t and b_t; three exogenous state variables, p_{1t}, p_{2t}, and y_t; and three decision variables, q_{1t}, q_{2t}, and i_t. Similar to the MDP in (4), its single-period reward, the dynamics of the endogenous state, and the constraints are all affine in the endogenous state and action. Furthermore, given the endogenous state $(K_t, b_t) = (K, b)$, the affine constraints form two bounded polyhedra: $\{(q_1, q_2)\colon q_1 + q_2 \le K, q_1 + q_2 \ge \alpha K, q_1, q_2 \ge 0\}$ and $\{i\colon 0 \le i \le b\}$. The first polyhedron is the same as the earlier one depicted in Figure 1 and, thus, has the same properties as before. The second one is simply the closed interval $[0, b]$ and also has those desirable properties. That is, for all $b \in [0, B]$, it always has two extreme points, 0 and b, and these extreme points are linear in b.

In addition to these properties, which are the same as those in Section 2.2, the MDP in (13) has two other important features. First, each bounded polyhedron involves only one endogenous state variable as the parameter; the first one depends only on K and the second one only on b. Second, the polyhedra are for sets of action variables that are mutually exclusive; the first one is for (q_1, q_2) and the second one is for i. Henceforth, we refer to these two features as the *decomposability* of the set of feasible actions. In this MDP, the action vector decomposes into two subvectors, (q_1, q_2) and i, and the endogenous state decomposes into its two elements, K and b. Each action subvector is constrained within a bounded polyhedron that is parameterized by an endogenous state variable. As we shall see below and discuss at the end of this subsection, decomposability plays an important role in preserving the affinity of the value function and the extremal property of an optimal policy.

2.3.2. Value Function. Let $\bar{v}_n(K, b, p_1, p_2, y)$ denote the value function of MDP (13) with n periods remaining and initial state (K, b, p_1, p_2, y). It satisfies the following dynamic program with $\bar{v}_0(\cdot,\cdot,\cdot,\cdot,\cdot) \equiv 0$ and

$$\bar{v}_n(K, b, p_1, p_2, y) = \max_{\substack{(q_1, q_2, i)\colon q_1, q_2 \ge 0 \\ q_1+q_2 \le K, \\ q_1+q_2 \ge \alpha K \\ 0 \le i \le b}} \{p_1 q_1 + p_2 q_2 + \beta\, \mathbb{E}[\bar{v}_{n-1}(K - \lambda_1 q_1 - \lambda_2 q_2 + iy, b - i, Z(p_1, p_2, y))]\}. \quad (14)$$

As in Section 2.2, we shall show that $\bar{v}_n$ is affine in the endogenous state using induction. Note that $\bar{v}_0(\cdot,\cdot,\cdot,\cdot,\cdot) \equiv 0$, which initiates the induction. Assume that $\bar{v}_{n-1}(K, b, p_1, p_2, y)$ is affine in K and b; i.e., there exist functions $\bar{f}_{1,n-1}$, $\bar{f}_{2,n-1}$, and $\bar{g}_{n-1}$ such that

$$\bar{v}_{n-1}(K, b, p_1, p_2, y) = \bar{f}_{1,n-1}(p_1, p_2, y)K + \bar{f}_{2,n-1}(p_1, p_2, y)b + \bar{g}_{n-1}(p_1, p_2, y). \quad (15)$$

From (14),

$$\begin{aligned}\bar{v}_n(K, b, p_1, p_2, y) &= \max_{\substack{(q_1, q_2, i)\colon q_1, q_2 \ge 0 \\ q_1+q_2 \le K, \\ q_1+q_2 \ge \alpha K \\ 0 \le i \le b}} \{p_1 q_1 + p_2 q_2 + \beta\, \mathbb{E}[\bar{f}_{1,n-1}(Z(p_1, p_2, y))(K - \lambda_1 q_1 - \lambda_2 q_2) \\ &\qquad + \bar{f}_{2,n-1}(Z(p_1, p_2, y))(b - i) + \bar{g}_{n-1}(Z(p_1, p_2, y))]\} \\ &= \beta\, \mathbb{E}[\bar{f}_{1,n-1}(Z(p_1, p_2, y))]K + \beta\, \mathbb{E}[\bar{f}_{2,n-1}(Z(p_1, p_2, y))]b + \beta\, \mathbb{E}[\bar{g}_{n-1}(Z(p_1, p_2, y))] \quad (16a) \\ &\quad + \max_{\substack{(q_1, q_2, i)\colon q_1, q_2 \ge 0 \\ q_1+q_2 \le K, \\ q_1+q_2 \ge \alpha K \\ 0 \le i \le b}} \{p_1 q_1 + p_2 q_2 - \beta\, \mathbb{E}[\bar{f}_{1,n-1}(Z(p_1, p_2, y))](\lambda_1 q_1 + \lambda_2 q_2) \quad (16b) \\ &\qquad - \beta\, \mathbb{E}[\bar{f}_{2,n-1}(Z(p_1, p_2, y))]i\}.\end{aligned}$$

The optimization in (16b) is a linear program with decision variables (q_1, q_2, i). Given that (q_1, q_2) and i are constrained by separate polyhedra, we can decompose (16b) into two linear programs, one with decision variables (q_1, q_2) and the other with decision variable i:

$$\begin{aligned}\bar{v}_n(K, b, p_1, p_2, y) &= \beta\, \mathbb{E}[\bar{f}_{1,n-1}(Z(p_1, p_2, y))]K + \beta\, \mathbb{E}[\bar{f}_{2,n-1}(Z(p_1, p_2, y))]b + \beta\, \mathbb{E}[\bar{g}_{n-1}(Z(p_1, p_2, y))] \quad (17a)\end{aligned}$$

$$+ \max_{\substack{(q_1,q_2):\, q_1,\, q_2 \geq 0 \\ q_1+q_2 \leq K, \\ q_1+q_2 \geq \alpha K}} \left\{ p_1 q_1 + p_2 q_2 - \beta\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))](\lambda_1 q_1 + \lambda_2 q_2) \right\} \tag{17b}$$

$$- \max_{0 \leq i \leq b} \left\{ \beta\, \mathbb{E}[\bar{f}_{2,\,n-1}(Z(p_1,p_2,y))] i \right\}. \tag{17c}$$

The optimal objective value of linear program (17b) is achieved at one of its four extreme points $\{(\alpha K, 0), (0, \alpha K), (K, 0), (0, K)\}$. The optimal objective value of linear program (17c) is achieved at one of its two extreme points $\{0, b\}$. Thus,

$$\begin{aligned}
&\bar{v}_n(K, b, p_1, p_2, y) \\
&= \beta\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))] K + \beta\, \mathbb{E}[\bar{f}_{2,\,n-1}(Z(p_1,p_2,y))] b + \beta\, \mathbb{E}[\bar{g}_{n-1}(Z(p_1,p_2,y))] \\
&\quad + \max \left\{ \left(p_1 - \beta\lambda_1\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))]\right) \alpha K,\ \left(p_2 - \beta\lambda_2\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))]\right) \alpha K, \right. \\
&\qquad \left. \left(p_1 - \beta\lambda_1\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))]\right) K,\ \left(p_2 - \beta\lambda_2\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))]\right) K \right\} \\
&\quad - \max \left\{ 0,\ \beta\, \mathbb{E}[\bar{f}_{2,\,n-1}(Z(p_1,p_2,y))] b \right\}.
\end{aligned} \tag{18}$$

Because $K,\ b \geq 0$, we can pull K and b out of their respective maximizations in (18) and compare only the coefficients. Thus, we collect terms to obtain

$$\bar{v}_n(K, b, p_1, p_2, y) = \bar{f}_{1n}(p_1, p_2, y) K + \bar{f}_{2n}(p_1, p_2, y) + \bar{g}_n(p_1, p_2, y), \tag{19a}$$

where

$$\begin{aligned}
\bar{f}_{1n}(p_1,p_2) &= \beta\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))] \\
&\quad + \max \left\{ \left(p_1 - \beta\lambda_1\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))]\right) \alpha,\ \left(p_2 - \beta\lambda_2\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))]\right) \alpha, \right. \\
&\qquad \left. \left(p_1 - \beta\lambda_1\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))]\right),\ \left(p_2 - \beta\lambda_2\, \mathbb{E}[\bar{f}_{1,\,n-1}(Z(p_1,p_2,y))]\right) \right\},
\end{aligned} \tag{19b}$$

$$\bar{f}_{2n}(p_1,p_2,y) = \beta\, \mathbb{E}[\bar{f}_{2,\,n-1}(Z(p_1,p_2,y))] - \max\{0,\ \beta\, \mathbb{E}[\bar{f}_{2,\,n-1}(Z(p_1,p_2,y))]\}, \tag{19c}$$

$$\bar{g}_n(p_1,p_2,y) = \beta\, \mathbb{E}[\bar{g}_{n-1}(Z(p_1,p_2,y))]. \tag{19d}$$

Thus, the induction continues, and the value function satisfies (19) for $n = 0, 1, \ldots, T$. Furthermore, functions $\bar{f}_{1n}$, $\bar{f}_{2n}$, and $\bar{g}_n$ satisfy the recursive *auxiliary equations* (19b) and (19d) with $\bar{f}_{1,\,0}(\cdot,\cdot,\cdot) = \bar{f}_{2,\,0}(\cdot,\cdot,\cdot) = \bar{g}_0(\cdot,\cdot,\cdot) \equiv 0$. This implies that $\bar{g}_n(\cdot,\cdot,\cdot) \equiv 0$ for all $n = 0, 1, \ldots, T$.

2.3.3. Optimal Policy. Let $(\bar{Q}_{n1}, \bar{Q}_{n2}, I)$ denote an optimal single-period decision rule for (q_1, q_2, i) with n periods remaining. From the preceding analysis, it is optimal to set (q_1, q_2) to an optimal extreme point of the polyhedron $\{(q_1, q_2)\colon\ q_1 + q_2 \leq K, q_1 + q_2 \leq \alpha K, q_1, q_2 \geq 0\}$, and it is optimal to set i to an extreme point of the closed internal $[0, b]$. Thus, $(\bar{Q}_{n1}, \bar{Q}_{n2}, I)$ is *extremal*, and the extremal property of an optimal policy is preserved. As in the basic model, the optimality of an extremal point is given by the solution to the auxiliary equations (19b) and (19c).

The following proposition summarizes these results.

Proposition 2. (1) *For $n = 1, \ldots, T$, the value function*

$$\bar{v}_n(K, b, p_1, p_2, y) = \bar{f}_{1n}(p_1, p_2, y) K + \bar{f}_{2n}(p_1, p_2, y) b + \bar{g}_n(p_1, p_2, y), \tag{20}$$

where $\bar{f}_{1n}$, $\bar{f}_{2n}$, and $\bar{g}_n$ satisfy the recursive auxiliary equations (19b) and (19d) with $\bar{f}_{1,0}(\cdot,\cdot,\cdot) = \bar{f}_{2,0}(\cdot,\cdot,\cdot) = \bar{g}_0(\cdot,\cdot,\cdot) \equiv 0$.

(2) *With n periods remaining, an optimal single-period decision rule $(\bar{Q}_{n1}, \bar{Q}_{n2}, \bar{I}_n)$ sets (q_1, q_2) to the extreme point that corresponds to an optimal maximand in (19b), and it sets i to the extreme point that corresponds to an optimal maximand in (19c).*

We conclude the analysis of the model extension by discussing the role of decomposability. In (16b), it allows us to separate the linear program about (q_1, q_2, i) into two independent linear programs. As a result, in each of these smaller linear programs, we can take advantage of the features of the associated polyhedron and apply the argument in Section 2.2 to establish the linearity in an endogenous state variable. If the set of feasible actions were not decomposable, then the solution to the linear program (16b) would be piecewise linear in K and b, because one would have to compare the terms involving K with those involving b. Then the features of each polyhedron would not be useful, and the affinity of the value function would break down.

3. Decomposable Affine MDPs: General Model and Results

In this section, we build on the insights from Section 2 and consider decomposable affine MDPs in general. We first characterize a generic decomposable affine MDP in Section 3.1, and then we show that it has an affine value function and an extremal optimal policy in Section 3.2. In Section 3.3, we discuss the significant implications of the results. For simplicity, we only consider the finite-horizon criterion in this tutorial. Interested readers are referred to Ning and Sobel [8] and [9] for infinite-horizon results.

3.1. Model for Decomposable Affine MDPs

The examples in Section 2 illustrate three key characteristics of a decomposable affine MDP: affinity, decomposability, and that each polyhedron is bounded and has a fixed number of extreme points that are linear in an endogenous state variable. We now generalize these insights to a generic decomposable affine MDP. We first introduce the notation, then formulate each of the three characteristics. While the discussion may appear to be mathematical, the key insights remain the same.

3.1.1. Notation. Consider a time-homogeneous MDP in discrete time. The state in period t is (s_t, e_t), where s_t is the endogenous (i.e., controlled) state, and e_t is the exogenous state. The endogenous state is a nonnegative n-by-1 vector (i.e., $s_t \in \Re_+^{n\times 1}$), and the exogenous state takes values in a set Ω (i.e., $e_t \in \Omega$). For simplicity, we henceforth assume that Ω is a finite set with ω elements; i.e., its cardinality $|\Omega| = \omega$.

The action in period t is a_t, which is an m-by-1 vector. Given state $(s_t = s, e_t = e)$, action a_t is constrained within the feasibility set $\mathcal{B}_{(s,e)} \subset \Re^{m\times 1}$. The immediate reward in period t is R_t, which may be random. Given $(s_t = s, e_t = e, a_t = a)$, the expected single-period reward $\mathbb{E}[R_t] = r(s, e, a)$, where $\mathbb{E}$ is the expectation operator.

Next we introduce notation for the dynamics of the endogenous and exogenous states. Given $(s_t = s, e_t = e, a_t = a)$, the endogenous state in period $t+1$ is a random vector (r.v.) whose distribution depends on (s, e, a). That is, $s_{t+1} \,|\, (s_t = s, e_t = e, a_t = a) \sim T(s, e, a)$, where $T(s, e, a)$ is an r.v. whose distribution is fully specified by (s, e, a). The exogenous state follows a finite Markov chain with state space Ω, and, as is indicated by its exogeneity, its transition probability is unaffected by the endogenous state and action. Let p_{ez} denote the one-step transition probability of the Markov chain for the exogenous state $(e, z \in \Omega)$. Let $\xi(e)$ denote the random variable that has the same distribution as e_{t+1} given $e_t = e$. Then $\xi(e) = z$ with probability p_{ez}.

3.1.2. Affinity. The expected single-period reward and dynamics of a decomposable affine MDP are affine functions of the endogenous state and action. Formally, for all state $(s, e) \in \Re_+^{n\times 1} \times \Omega$ and action $a \in \mathcal{B}_{(s,e)}$,

$$T(s, e, a) \sim \mathbf{Y}^{\mathbb{S}}(e)s + \mathbf{Y}^{\mathbb{A}}(e)a + \mathbf{Y}^{o}(e), \tag{21a}$$

$$r(s, e, a) = y^{\mathbb{S}}(e)s + y^{\mathbb{A}}(e)a + y^{o}(e). \tag{21b}$$

Here, $\mathbf{Y}^{\mathbb{S}}(e)$, $\mathbf{Y}^{\mathbb{A}}(e)$, and $\mathbf{Y}^{o}(e)$ are random and take values in $\Re^{n\times n}$, $\Re^{n\times m}$, and $\Re^{n\times 1}$, $y^{\mathbb{S}}(e)\in\Re^{1\times n}$, $y^{\mathbb{A}}(e)\in\Re^{1\times m}$, and $y^{o}(e)\in\Re$. Boldface is used as a visual reminder that $\mathbf{Y}^{\mathbb{S}}(e)$, $\mathbf{Y}^{\mathbb{A}}(e)$, and $\mathbf{Y}^{o}(e)$ are random, unlike $y^{\mathbb{S}}(e)$, $y^{\mathbb{A}}(e)$, and $y^{o}(e)$, which are deterministic. Superscripts $\mathbb{S}$ and $\mathbb{A}$ indicate coefficients of s and a, and superscript o indicates the constant in the affine functions.

3.1.3. Decomposability. Recall that $\mathcal{B}_{(s,e)}$ is the set of feasible actions given state (s,e). Let s_i denote element i of the endogenous state s. A decomposable affine MDP satisfies the following decomposability condition for all $(s,e)\in\Re_+^{n\times 1}\times\Omega$:

$$\mathcal{B}_{(s,e)}=\bigtimes_{i=1}^{n}\mathcal{B}^i_{(s_i,e)}. \tag{22}$$

That is, given (s,e), the action vector a can be partitioned into n subvectors, $a^i(e)$ $(i=1,\ldots,n)$, and set $\mathcal{B}_{(s,e)}$ can be partitioned into n subsets, $\mathcal{B}^i_{(s_i,e)}$, such that $a^i(e)\in\mathcal{B}^i_{(s_i,e)}$.

Two things are noteworthy about the decomposability assumption. First, set $\mathcal{B}^i_{(s_i,e)}$ depends on s only via s_i. Thus, action variables in $a^i(e)$ are constrained only by one endogenous state variable s_i. Second, as indicated by the notation, the partitioning of a may vary with the exogenous state e but not with s.

For the model in Section 2.3, state $s=(K,b)$, $e=(p_1,p_2,y)$, $a=(q_1,q_2,i)$, and $\mathcal{B}_{(s,e)}=\{(q_1,q_2,i)\colon q_1+q_2\le K,\ q_1+q_2\ge\alpha K,\ i\le b,\ q_1,q_2,\ i\ge 0\}$. For all e, action a is partitioned into two subvectors, $a^1(e)=(q_1,q_2)$ and $a^2(e)=i$, and set $\mathcal{B}_{(s,e)}$ is partitioned into two subsets, $\mathcal{B}^1_{(K,e)}=\{(q_1,q_2)\colon q_1+q_2\le K,\ q_1+q_2\ge\alpha K,\ q_1,q_2\ge 0\}$ and $\mathcal{B}^2_{(b,e)}=\{i\colon 0\le i\le b\}$, such that $a\in\mathcal{B}_{(s,e)}$ can be written as $a^1(e)\in\mathcal{B}^1_{(K,e)}$ and $a^2(e)\in\mathcal{B}^2_{(b,e)}$.

Henceforth, we use superscript i to indicate entities that are related to subvector $a^i(e)$, and we use $K^i(e)$ to denote the dimension of $a^i(e)$. Then, for $i=1,\ldots,n$, $a^i(e)\in\mathcal{B}^i_{(s_i,e)}\subset\Re^{K^i(e)\times 1}$, and $\sum_{i=1}^n K^i(e)=m$.

3.1.4. Compound Affinity and Decomposability. This last condition for decomposable affine MDPs deals with the structure of each subset $\mathcal{B}^i_{(s_i,e)}$ $(i=1,\ldots,n)$. It consists of three parts. First, for all $(s_i,e)\in\Re_+\times\Omega$, $\mathcal{B}^i_{(s_i,e)}$ is a bounded polyhedron. Second, the number of extreme points of $\mathcal{B}^i_{(s_i,e)}$ does not depend on s_i. Let $\mathcal{X}^i_{(s_i,e)}$ denote the set of extreme points of $\mathcal{B}^i_{(s_i,e)}$; then $|\mathcal{X}^i_{(s_i,e)}|$ denotes the number of extreme points and is invariant with respect to s_i. For expository simplicity, we henceforth use $|\mathcal{X}^i_e|$ instead of $|\mathcal{X}^i_{(s_i,e)}|$ to denote the cardinality of $\mathcal{X}^i_{(s_i,e)}$. Third, there exist a $K^i(e)$-by-$|\mathcal{X}^i_e|$ matrix $M^i(e)$ and a scalar $c^i(e)$ such that

$$\mathcal{X}^i_{(s_i,e)}=\bigl\{M^i_{\cdot k}(e)s_i+c^i(e)1^i(e),k=1,\ldots,|\mathcal{X}^i_e|\bigr\}, \tag{23}$$

where $M^i_{\cdot k}(e)$ is the kth column of $M^i(e)$, and $1^i(e)$ is a column vector of all ones and has the same dimension as $a^i(e)$. Thus, each extreme point of $\mathcal{B}^i_{(s_i,e)}$ is affine with respect to $s_i\in\Re_+$, and the constant vector has identical elements. Henceforth, we shall refer to $M^i_{\cdot k}(e)x+c^i(e)1^i(e)$ as the *kth extreme point in* $\mathcal{X}^i_{(s_i,e)}$.

For the model in Section 2.3, $\mathcal{B}^1_{(K,e)}$ and $\mathcal{B}^2_{(b,e)}$ are bounded polyhedra. The set of extreme points of $\mathcal{B}^1_{(K,e)}$ is $\mathcal{X}^1_{(K,e)}=\{(\alpha K,0),(0,\alpha K),(K,0),(0,K)\}$, and that of $\mathcal{B}^2_{(b,e)}$ is $\mathcal{X}^2_{(b,e)}=\{0,b\}$. Thus, for all e, $|\mathcal{X}^1_e|=4$ and $|\mathcal{X}^2_e|=2$, which do not depend on the endogenous state. Furthermore, there exist

$$M^1(e)=\begin{pmatrix}\alpha & 0 & 1 & 0\\ 0 & \alpha & 0 & 1\end{pmatrix},\quad c^1(e)=0,$$
$$M^2(e)=(0\ 1),\quad c^2(e)=0,$$

such that the extreme points in $\mathcal{X}^1_{(K,e)}$ and $\mathcal{X}^2_{(b,e)}$ satisfy (23).

3.2. Value Function and Optimal Policy

Consider a decomposable affine MDP that has a finite-horizon criterion with horizon length T and single-period discount factor β. The goal is to find an optimal policy that maximizes the EPV of rewards $\mathbb{E}[\sum_{t=1}^{T} \beta^{t-1} R_t]$.

Let v^τ denote the value function with τ periods remaining ($\tau = 1, \ldots, T$), and let $A^{\tau i}(\cdot, e)$ denote an optimal single-period decision rule for subvector $a^i(e)$. Then v_τ satisfies

$$v^\tau(s,e) = \max_{a \in \mathcal{B}_{(s,\,e)}} \left\{ r(s,e,a) + \beta\, \mathbb{E}[v^{\tau-1}(T(s,e,a),\ \xi(e))] \right\}, \tag{24}$$

and $(A^{\tau 1}(s,e), \ldots, A^{\tau n}(s,e))$ must achieve the maximum in (24).

For the analysis below, it is convenient to write (21) in terms of subvectors $a^i(e)$:

$$T(s,e,a) \sim \mathbf{Y}^{\mathbb{S}}(e)s + \sum_{i=1}^{n} \mathbf{Y}^{\mathbb{A}i}(e) a^i(e) + \mathbf{Y}^{o}(e), \tag{25a}$$

$$r(s,e,a) = y^{\mathbb{S}}(e)s + \sum_{i=1}^{n} y^{\mathbb{A}i}(e) a^i(e) + y^{o}(e), \tag{25b}$$

where $\mathbf{Y}^{\mathbb{A}i}(e)$ is the submatrix of $\mathbf{Y}^{\mathbb{A}}(e)$ 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 $a^i(e)$.

From (25), dynamic program (24) can be written as

$$\begin{aligned} v^\tau(s,e) = {} & y^{\mathbb{S}}(e)s + y^{o}(e) \\ & + \max_{a \in \mathcal{B}_{(s,\,e)}} \left\{ \sum_{i=1}^{n} y^{\mathbb{A}i}(e) a^i(e) + \beta\, \mathbb{E}\left[v^{\tau-1}\left(\mathbf{Y}^{\mathbb{S}}(e)s + \sum_{i=1}^{n} \mathbf{Y}^{\mathbb{A}i}(e) a^i(e) + \mathbf{Y}^{o}(e),\ \xi(e) \right) \right] \right\}. \end{aligned} \tag{26}$$

Define $Y^{\mathbb{S}}(e,z) = \mathbb{E}[\mathbf{Y}^{\mathbb{S}}(e) \,|\, \xi(e) = z] \in \Re^{n \times n}$, $Y^{\mathbb{A}i}(e,z) = \mathbb{E}[\mathbf{Y}^{\mathbb{A}i}(e) \,|\, \xi(e) = z] \in \Re^{n \times K^i(e)}$ for $i = 1, \ldots, n$, and $Y^{o}(e,z) = \mathbb{E}[\mathbf{Y}^{o}(e) \,|\, \xi(e) = z] \in \Re^{n \times 1}$.

The following theorem states that a decomposable affine MDP has a value function that is affine in the endogenous state. Its proof is similar in spirit to the analysis in Section 2.3. After the proof, we discuss the roles of the affinity, decomposability, and compound affinity and decomposability conditions in ensuring the affinity of the value function.

Theorem 1. *For $\tau = 1, 2, \ldots, T$, the value function with τ periods remaining has the affine representation*

$$v^\tau(s,e) = f^\tau(e)s + g^\tau(e), \tag{27}$$

where $f^\tau(e) = (f_1^\tau(e), \ldots, f_n^\tau(e)) \in \Re^{1 \times n}$ 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 $\tau \in \mathbb{N}$,

$$\begin{aligned} f_i^\tau(e) = {} & y_i^{\mathbb{S}}(e) + \beta \sum_{z \in \Omega} p_{ez} f^{\tau-1}(z) Y_{\cdot i}^{\mathbb{S}}(e,z) \\ & + \max_{k=1,\ldots,|\mathcal{X}_e^i|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z \in \Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \right] M_{\cdot k}^i(e) \right\}, \end{aligned} \tag{28a}$$

$$\begin{aligned} g^\tau(e) = {} & y^{o}(e) + \beta \sum_{z \in \Omega} p_{ez} [f^{\tau-1}(z) Y^{o}(e,z) + g^{\tau-1}(z)] \\ & + \sum_{i=1}^{n} \left[c^i(e) \left(y^{\mathbb{A}i}(e) + \beta \sum_{z \in \Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \right) 1^i(e) \right]. \end{aligned} \tag{28b}$$

Proof. Initiate a proof by induction on τ by confirming (27) 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 $\tau \in \{1,\dots,T\}$, 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 (26) yields

$$\begin{aligned}
&v^\tau(s,e) - y^{\mathbb{S}}(e)s - y^o(e)\\
&= \max_{a\in\mathcal{B}_{(s,e)}} \left\{ \sum_{i=1}^n y^{\mathbb{A}i}(e)a^i(e) + \beta\mathbb{E}\left[v^{\tau-1}\left(\mathbf{Y}^{\mathbb{S}}(e)s + \sum_{i=1}^n \mathbf{Y}^{\mathbb{A}i}(e)a^i(e) + \mathbf{Y}^o(e),\ \xi(e)\right)\right]\right\}\\
&= \max_{a\in\mathcal{B}_{(s,e)}} \left\{ \sum_{i=1}^n y^{\mathbb{A}i}(e)a^i(e) + \beta\mathbb{E}\left[f^{\tau-1}(\xi(e))\left(\mathbf{Y}^{\mathbb{S}}(e)s + \sum_{i=1}^n \mathbf{Y}^{\mathbb{A}i}(e)a^i(e) + \mathbf{Y}^o(e)\right) + g^{\tau-1}(\xi(e))\right]\right\}.
\end{aligned} \tag{29}$$

Recall that $Y^{\mathbb{S}}(e,z) = \mathbb{E}[\mathbf{Y}^{\mathbb{S}}(e) \,|\, \xi(e) = z]$; thus

$$\mathbb{E}[f^{\tau-1}(\xi(e))\mathbf{Y}^{\mathbb{S}}(e)] = \sum_{z\in\Omega} \mathbb{E}[f^{\tau-1}(\xi(e))\mathbf{Y}^{\mathbb{S}}(e) \,|\, \xi(e) = z]p_{ez} = \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{S}}(e,z). \tag{30}$$

Similarly, $\mathbb{E}[f^{\tau-1}(\xi(e))\mathbf{Y}^{\mathbb{A}i}(e)] = \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \in \Re^{1\times K^i(e)}$, $\mathbb{E}[f^{\tau-1}(\xi(e))\mathbf{Y}^o(e)] = \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^o(e,z) \in \Re$, and $\mathbb{E}[g^{\tau-1}(\xi(e))] = \sum_{z\in\Omega} p_{ez} g^{\tau-1}(z) \in \Re$. Therefore, (29) is

$$\begin{aligned}
&v^\tau(s,e) - y^{\mathbb{S}}(e)s - y^o(e)\\
&\quad= \max_{a\in\mathcal{B}_{(s,e)}} \left\{ \sum_{i=1}^n y^{\mathbb{A}i}(e)a^i(e) + \beta \sum_{z\in\Omega} p_{ez}\left[f^{\tau-1}(z)Y^{\mathbb{S}}(e,z)s + \sum_{i=1}^n f^{\tau-1}(z)Y^{\mathbb{A}i}(e,z)a^i(e)\right.\right.\\
&\qquad\qquad \left.\left. + f^{\tau-1}(z)Y^o(e,z) + g^{\tau-1}(z)\right]\right\} &\text{(31a)}\\
&\quad= \beta \sum_{z\in\Omega} p_{ez}\left[f^{\tau-1}(z)Y^{\mathbb{S}}(e,z)s + f^{\tau-1}(z)Y^o(e,z) + g^{\tau-1}(z)\right] &\text{(31b)}\\
&\qquad + \max_{a\in\mathcal{B}_{(s,e)}} \left\{ \sum_{i=1}^n y^{\mathbb{A}i}(x)a^i(e) + \beta \sum_{z\in\Omega} p_{xz} f^{\tau-1}(z) \sum_{i=1}^n Y^{\mathbb{A}i}(e,z)a^i(e)\right\}. &\text{(31c)}
\end{aligned}$$

The remainder of the proof shows that (31c) is affine in s.

Because the maximand of (31c) is linear and $\mathcal{B}_{(s,e)}$ consists of separate polyhedra, we can write (31c) as the sum of n suboptimizations. The i-th one has the components of $a^i(e)$ as variables, $\mathcal{B}^i_{(s_i,e)}$ as the feasible region, and $y^{\mathbb{A}i}(x)a^i(e) + \beta\sum_{z\in\Omega} p_{xz} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z)a^i(e)$ as the maximand. Let Z_i denote the optimal value of the i-th optimization. Then (31c) equals $\sum_{i=1}^n Z_i$, and

$$Z_i = \max_{a^i(e)\in\mathcal{B}^i_{(s_i,e)}} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z)\right] a^i(e)\right\}. \tag{32}$$

This is a linear program with a bounded feasible region. Thus, the optimal objective value is obtained at an extreme point. From the representation for the extreme points in (23), it follows that

$$\begin{aligned}
Z_i &= \max_{k=1,\dots,|\mathcal{X}^i_e|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z)\right] (M^i_{\cdot k}(e)s_i + c^i(e)1^i(e))\right\} &\text{(33a)}\\
&= \max_{k=1,\dots,|\mathcal{X}^i_e|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z)\right] M^i_{\cdot k}(e)s_i &\text{(33b)}\right.\\
&\qquad \left. + \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z)\right] c^i(e)1^i(e)\right\}. &\text{(33c)}
\end{aligned}$$

Because the summand in (33c) is a constant with respect to k, we can take it out of the maximization and consider only the first summand in (33b). Because $s_i \geq 0$ in (33b), the maximization further reduces to a maximization over the coefficients. Thus, Z_i satisfies

$$Z_i = \left[\max_{k=1,\ldots,|\mathcal{X}_e^i|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \right] M^i_{\cdot k}(e) \right\} \right] s_i \tag{34a}$$

$$+ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \right] 1^i(e) c^i(e). \tag{34b}$$

Therefore, (29) is

$$v^\tau(s,e) = y^{\mathbb{S}}(e)s + y^o(e) + \beta \sum_{z\in\Omega} p_{ez} \left[f^{\tau-1}(z) Y^{\mathbb{S}}(e,z) s + f^{\tau-1}(z) Y^o(e,z) + g^{\tau-1}(z) \right] + \sum_{i=1}^n Z_i. \tag{35}$$

Use (34) and collect terms in (35) 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 (28). □

The inductive proof for Theorem 1 illustrates the roles of the affinity, decomposability, and compound affinity and decomposability conditions in ensuring the affinity of the value function. First, the affinity of the dynamics and expected single-period reward yields an affine objective function. Second, the decomposability feature leads to a partition in the feasibility set. Paired with the affine objective function, this allows decomposing optimization (31c) over $\mathcal{B}_{(s,e)}$ into n smaller optimizations over $\mathcal{B}^1_{(s_1,e)}, \ldots, \mathcal{B}^n_{(s_n,e)}$. Third, the polyhedral feature of $\mathcal{B}^i_{(s_i,e)}$ makes each of the smaller optimizations a linear program. Finally, the boundedness of $\mathcal{B}^i_{(s_i,e)}$, the features of its extreme points, and $s_i \geq 0$ ensure that the optimal objective is linear in s_i. Among all the features, we emphasize the linearity of the extreme points in s_i. In general, an extreme point of a linear program is piecewise linear in s_i, which would cause the affinity of the value function to break down. This explains why numerous linear programming formulations of dynamic models have *nonlinear* value functions (see material on sensitivity analysis in, e.g., Dantzig [3] and Mathur and Solow [6]).

The next theorem specifies an optimal policy and shows that it is extremal. Let $J_i^{\tau-1}(e)$ be an index $k \in \{1,\ldots,|\mathcal{X}_e^i|\}$ that achieves the maximum in (28a):

$$J_i^{\tau-1}(e) \in \operatorname*{arg\,max}_{k=1,\ldots,|\mathcal{X}_e^i|} \left\{ \left[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z) \right] M^i_{\cdot k}(e) \right\}. \tag{36}$$

Recall that $A^{\tau i}(\cdot, e)$ is an optimal single-period decision rule for $a^i(e)$ with τ periods remaining in the horizon.

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

$$A^{\tau i}(s,e) = M^i_{\cdot, J_i^{\tau-1}(e)}(e) \times s_i + c^i(e) 1^i(e). \tag{37}$$

Proof. Use $v^\tau(s,e) = f^\tau(e)s + g^\tau(e)$ (Theorem 1) in dynamic program (26) to obtain (31). Thus, the solution of (32) implies that (37) is optimal. □

From Theorem 2, if multiple maximands on the right side of (36) achieve the optimum, then each of the corresponding extreme points is optimal. This implies that a tie is broken arbitrarily, and, as a result, multiple optimal policies may give rise to the unique value function.

3.3. Implications for Computation and Analysis

Theorems 1 and 2 have a couple of noteworthy features. First, they provide an *exact* solution to an MDP that has continuous vector-valued state and actions. Second, this solution is obtained analytically rather than numerically. This is striking because most MDPs with continuous states and actions are solved *approximately* using numerical methods. As discussed earlier, these results are made possible by the defining features of decomposable affine MDPs: affinity, decomposability, and compound affinity and decomposability.

From Theorem 1, a full specification of the value function depends on the specification of coefficients f^τ and g^τ. From Theorem 2, a full specification of an optimal policy depends on f^τ and g^τ as well, because the optimality of the k-th extreme point of $\mathcal{B}^i_{(s_i,e)}$ depends on whether $[y^{\mathbb{A}i}(e) + \beta \sum_{z\in\Omega} p_{ez} f^{\tau-1}(z) Y^{\mathbb{A}i}(e,z)] M^i_{\cdot k}(e)$ achieves the maximum in (36). Thus, instead of solving the original dynamic program (26), we can take advantage of Theorems 1 and 2 and compute only f^τ and g^τ to specify the value function and an optimal policy.

This insight has significant computational implications, because f^τ and g^τ can be obtained by solving the recursive auxiliary equations (28). Unlike the original dynamic program (26), whose domain is the entire state space $\Re^n_+ \times \Omega$, the auxiliary equations are only on $\{1,\ldots,n\} \times \Omega$. Furthermore, unlike the optimization in (26), which is over the continuous feasibility set $\mathcal{B}_{(s,e)}$, the optimization in the auxiliary equations is over a finite set $\mathcal{X}^i_{(s_i,e)}$. As a result, the common "curse of dimensionality" caused by discrete approximation of the continuous endogenous state and action space does not arise in decomposable affine MDPs. An *exact* solution can be obtained easily. Interested readers are referred to Ning and Sobel [8] for efficient algorithms to solve decomposable affine MDPs with finite- and infinite-horizon discounted criteria.

Intuitively, discretizing the continuous action space is not necessary for decomposable affine MDPs because of the extremal property of an optimal policy. It implies that an interior point of $\mathcal{B}^i_{(s_i,e)}$ cannot strictly dominate an optimal extreme point. Thus, discretization is not needed, and the curse of dimensionality due to a continuous action space is exorcised.

Finally, we note that while the curse of dimensionality is significantly alleviated thanks to the auxiliary equations, the dimensions of the endogenous state and action still play an important role in determining the size of the equations one needs to solve. First, the variables in the auxiliary equations are vector-valued functions (f^τ, g^τ), in which $f^\tau(e) \in \Re^{n\times 1}$. Thus, a high-dimensional endogenous state vector leads to a large set of auxiliary equations. Second, solving the auxiliary equations requires comparing the extreme points in each of the n polyhedra. Keeping n fixed, an increase in the dimension of the action vector may increase the number of extreme points in a polyhedron, thus increasing the size of the auxiliary equations.

4. Applications

In this section, we present two applications of decomposable affine MDPs. The one in Section 4.1 is a straightforward application of Section 3 in fishery management. The application in Section 4.2 is in dynamic capacity portfolio management and extends Section 3 slightly by allowing a salvage value function that is affine in the endogenous state.

4.1. Fishery

An important research topic in resource management and economics is the optimal harvesting of biologically renewable resources such as fish and forests. In this subsection, we study the optimal harvesting of a fish species. Ideally, in such studies, the age structure of the harvested population should be incorporated in the model, because it plays an important role in determining the natural mortality and growth of the population (Anderson et al. [1], Berkeley et al. [2]). However, few fishery management models include age structure because of the daunting computational challenge of the curse of dimensionality. In the following,

we show that the incorporation of age structure in decomposable affine MDPs is free of the curse.

For simplicity, we study a single species with three age classes and consider the optimal policy for age-specific harvesting.[1] Let s_{it} and a_{it} $(i=1,2,3,\ t=0,1,\ldots)$ denote the numbers of age i fish before and after harvesting in year t, respectively. Thus, the number of age i fish that are harvested in year t is $s_{it}-a_{it}$. The endogenous state in year t is (s_{1t},s_{2t},s_{3t}), and the action is (a_{1t},a_{2t},a_{3t}). The exogenous state variable e_t includes relevant information on environmental conditions and economics of the fish markets in year t.

The sequence of events is as follows. In fishing year t, harvesting takes place at the beginning of the year, then the unharvested fish reproduce, and finally, natural mortality occurs. Let $\lambda_i(e_t)$ denote the fecundity rate of age i fish in year t (after accounting for the survival rate of larvae), and let $\theta_i(e_t)\in(0,1)$ denote the natural survival rate of age i fish. Both $\lambda_i(e_t)$ and $\theta_i(e_t)$ are random variables. The dynamics of the fish population in each age class are

$$\begin{aligned} s_{1,\,t+1} &\sim \sum_{i=1}^{3}\lambda_i(e_t)a_{it},\\ s_{2,\,t+1} &\sim \theta_1(e_t)a_{1t},\\ s_{3,\,t+1} &\sim \theta_2(e_t)a_{2t}+\theta_3(e_t)a_{3t}. \end{aligned}$$

These equations imply affine dynamics as in (21a) with

$$\mathbf{Y}^{\mathbb{S}}(\cdot)\equiv 0,\qquad \mathbf{Y}^{\mathbb{A}}(e)\sim\begin{pmatrix}\lambda_1(e) & \lambda_2(e) & \lambda_3(e)\\ \theta_1(e) & 0 & 0\\ 0 & \theta_2(e) & \theta_3(e)\end{pmatrix},\qquad \mathbf{Y}^{o}(\cdot)\equiv 0.$$

Let $w_i(e_t)$ denote the average net profit per age i fish in year t when the exogenous state is e_t. The net profit in year t is then $\sum_{i=1}^{3}w_i(e_t)(s_{it}-a_{it})$. Thus, the expected single-period reward is $r(s,e,a)=w(e)(s-a)$, which satisfies (21b) with

$$y^{\mathbb{S}}(e)=w(e),\qquad y^{\mathbb{A}}(e)=-w(e),\qquad y^{o}(e)=0.$$

Consider a schooling fishery whose fishing cost is approximately a linear function of the harvest (Tahvonen [14]). In year t, given state $(s_{1t},s_{2t},s_{3t})=s$ and $e_t=e$, the feasible set of action is $\mathcal{B}_{(s,e)}=\{a\in\Re^3_+\colon a\le s\}$. Thus, $\mathcal{B}_{(s,e)}$ satisfies decomposability condition (22) with $\mathcal{B}^i_{(s_i,e)}=[0,s_i]$, $a^i(e)\equiv a_i$, and $K^i(e)\equiv 1$. Furthermore, for $i=1,2,3$, $\mathcal{B}^i_{(s_i,e)}$ is a bounded polyhedron with extreme point set $\mathcal{X}^i_{(s_i,e)}=\{0,s_i\}$. Thus, the compound affinity and decomposability condition is satisfied with $|\mathcal{X}^i_e|\equiv 2$, $M^i(e)\equiv(0,1)$, and $c^i(e)\equiv 0$ for all $e\in\Omega$ in Equation (23). Therefore, this is a decomposable affine MDP, and the results in Section 3.2 are applicable.

The decision maker wants to maximize the EPV of the annual net profit over T periods—namely, $\max\mathbb{E}[\sum_{t=1}^{T}\beta^{t-1}\sum_{i=1}^{3}w_i(e_t)(s_{it}-a_{it})]$. From Theorem 1, the value function with τ periods remaining in the horizon is $(\tau=1,2,\ldots,T)$:

$$v^{\tau}(s,e)=\sum_{i=1}^{3}f_i^{\tau}(e)s_i,\quad (s,e)\in\Re^3_+\times\Omega, \tag{39a}$$

$$f_1^{\tau}(e)=\max\{w_1(e),\beta\,\mathbb{E}[\lambda_1(e)f_1^{\tau-1}(\xi(e))+\theta_1(e)f_2^{\tau-1}(\xi(e))]\}, \tag{39b}$$

$$f_2^{\tau}(e)=\max\{w_2(e),\beta\,\mathbb{E}[\lambda_2(e)f_1^{\tau-1}(\xi(e))+\theta_2(e)f_3^{\tau-1}(\xi(e))]\}, \tag{39c}$$

$$f_3^{\tau}(e)=\max\{w_3(e),\beta\,\mathbb{E}[\lambda_3(e)f_1^{\tau-1}(\xi(e))+\theta_3(e)f_3^{\tau-1}(\xi(e))]\}. \tag{39d}$$

[1] Age-specific harvesting is a reasonable assumption with current fishing technologies and for species with age classes that are spatially segregated (e.g., shrimp in the Gulf of Mexico).

From Theorem 2, an optimal single-period decision rule for harvesting age i fish with τ periods remaining, $A^{\tau i}$, satisfies

$$A^{\tau i}(s,e) = \begin{cases} 0 & \text{if } w_i(e) \le f_i^{\tau}(e), \\ s_i & \text{otherwise,} \end{cases} \qquad (i=1,2,3). \tag{40}$$

That is, it is optimal to harvest all the age i fish if the immediate net profit w_i is higher than $\beta\,\mathbb{E}[\lambda_i(e) f_1^{\tau-1}(\xi(e)) + \theta_i(e) f_{i+1}^{\tau-1}(\xi(e))]$. Otherwise, it is optimal to harvest nothing. This "bang-bang" feature of the optimal policy arises from the fact that in this MDP, each polyhedron is simply a closed interval.

To summarize, decomposable affine MDPs exorcise the curse of dimensionality in dynamic stochastic models of fisheries, and they allow easy computation and specification of an optimal policy with multiple age classes and exogenous state variables.

4.2. Dynamic Capacity Portfolio Management

Dynamic management of a portfolio of production capacities is an important topic that has not been explored much in the stochastic capacity management literature (Van Mieghem [16]). This is to a large extent caused by the analytical complexity of such problems. In this subsection, we show that decomposable affine MDPs provide a tractable modeling framework for these studies. (See Ning and Sobel [7] for an example that uses decomposable affine MDPs to study dynamic capacity portfolio management under financial constraints.)

For expository simplicity, we consider a firm that has two production facilities, each of which makes two types of products. (See Ning and Sobel [8] for a more general model that has n facilities and facility i $(i=1,\ldots,n)$ makes m_i types of products.) Different facilities may make identical or completely different products, and the firm makes production, investment, and divestment decisions for each of them in response to the exogenous stochastic market. Henceforth, we refer to the type i product made at facility j $(j=1,2)$ as product (i,j), and we refer to the capacity at facility i as capacity i.

At the beginning of period t, the firm observes the following internal data: the available capacity at each facility, (K_{1t}, K_{2t}), and the amount of budget available for investment, b_t. It observes the following external data: the profit margin $(p_{ijt}\colon i,j=1,2)$ of each product (which may be negative); the amount of capacity i installed per unit of investment, y_{it}; and the divestment price per unit of capacity i divested, r_{it}.

With this information, the firm chooses production quantities $(q_{ijt}\colon i,j=1,2)$; the amount of budget to invest in expanding capacity i, x_{it}; and the amount of capacity i to divest, d_{it}. The sequence of events is as follows. Production and investment occur at the beginning of the period; divestment occurs at the end of the period. All earnings from production and divestment are collected at the end of the period; capacity installed from investment is ready to use at the beginning of the next period.

The firm's decisions are subject to the following constraints. First, the amount of production at facility i cannot exceed the capacity K_{it}:

$$\gamma_{i1} q_{i1t} + \gamma_{i2} q_{i2t} \le K_{it}, \quad i=1,2, \tag{41}$$

where γ_{ij} $(j=1,2)$ reflects the efficiency of capacity i in making product (i,j).

Second, the amount of capacity i divested cannot exceed the amount of capacity remaining at the end of the period. Capacity depreciates during period t in two ways before divestment occurs: natural depreciation, where the capacity deteriorates due to the passage of time, and production-induced depreciation, where the capacity deteriorates due to its use for production. Given production quantities q_{ijt}, capacity K_{it} depreciates to $\theta_i K_{it} - \sum_{j=1}^{2} \lambda_{ij} q_{ijt}$ at the end of period t, where $\theta_j \in (0,1]$ and $\lambda_{ij} \in [0,1]$ reflect the natural and production-induced

depreciation, respectively. Assume $\lambda_{ij} \le \theta_i \gamma_{ij}$ to guarantee nonnegative postdepreciation capacity. Thus, the divestment decision satisfies

$$d_{it} \le \theta_i K_{it} - \lambda_{i1} q_{i1t} - \lambda_{i2} q_{i2t}, \quad i = 1, 2. \tag{42}$$

Third, the total amount of investment in period t cannot exceed the available budget:

$$x_{1t} + x_{2t} \le b_t. \tag{43}$$

The endogenous state variables in period t are (K_{1t}, K_{2t}, b_t). Their dynamical equations are

$$K_{i,\,t+1} = \theta_i K_{it} - \lambda_{i1} q_{i1t} - \lambda_{i2} q_{i2t} + y_{it} x_{it} - d_{it}, \quad i = 1, 2, \tag{44a}$$

$$b_{t+1} = b_t - x_{1t} - x_{2t}. \tag{44b}$$

The initial investment budget is B, so $b_1 = B$.

Let $e_t = (p_{ijt}, y_{it}, r_{it}\colon i, j = 1, 2)$ denote the exogenous state. The process $e_1, e_2, \ldots$ follows an exogenous Markov chain with finite state space Ω.

In period t, the firm earns a profit of p_{ijt} per unit of product (i, j) and a revenue of r_{it} per unit of divested capacity i. Thus, its single-period reward is $\sum_{i,j}(p_{ijt} q_{ijt} + r_{it} d_{it})$. Assume that at the end of the T-period planning horizon all leftover investment budget is lost and all remaining capacity is divested at a price of $r_{i,\,T+1}$. The decision maker wants to maximize the EPV of the firm's earnings over T periods—namely,

$$\max \mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1} \left(\sum_{i,j=1}^{2} p_{ijt} q_{ijt} + r_{it} d_{it} \right) + \sum_{i=1}^{2} \beta^T r_{i,\,T+1} K_{i,\,T+1} \right]. \tag{45}$$

Note that unlike the models discussed so far, expression (45) includes a nonzero salvage value $\sum_{i=1}^{2} \beta^T r_{i,\,T+1} K_{i,\,T+1}$. As we shall see below, the linearity of this salvage value in K_{T+1} preserves the affinity of the value function and the extremal property of an optimal policy.

In the following, we first show that this problem is a decomposable affine MDP. Then we present the value function and an optimal policy. From (44) and (45), the dynamical equations of the endogenous state and the expected single-period rewards are affine functions of the endogenous state (K_{1t}, K_{2t}, b_t) and actions $(q_{ijt}, d_{it}, x_{it}\colon i, j = 1, 2)$. From (41)–(43), action variables $(q_{i1t}, q_{i2t}, d_{it})$ are constrained only by the endogenous state variable K_{it}, and action variables (x_{1t}, x_{2t}) are constrained only by b_t. Thus, given state $(K_{1t}, K_{2t}, b_t) = (K_1, K_2, b)$ and $e_t = e$, we can partition the action vector $(q_{ij}, d_i, x_i\colon i, j = 1, 2)$ into the following three subvectors:

$$a^1(e) \equiv (q_{11}, q_{12}, d_1), \qquad a^2(e) \equiv (q_{21}, q_{22}, d_2), \qquad a^3(e) \equiv (x_1, x_2), \tag{46}$$

such that, for $i = 1, 2$,

$$a^i(e) \in \mathcal{B}^i_{(K_i,\,e)} = \left\{ (q_{i1}, q_{i2}, d_i) \in \Re^3_+\colon \sum_{j=1}^{2} \gamma_{ij} q_{ij} \le K_i, d_i \le \theta_i K_i - \sum_{j=1}^{2} \lambda_{ij} q_{ij} \right\}, \tag{47a}$$

$$a^3(e) \in \mathcal{B}^{n+1}_{(b,\,e)} = \left\{ (x_1, x_2) \in \Re^2_+\colon x_1 + x_2 \le b \right\}. \tag{47b}$$

Thus, the decomposability condition is satisfied.

Finally, from (47), sets $\mathcal{B}^1_{(K_1,\,e)}$, $\mathcal{B}^2_{(K_2,\,e)}$, and $\mathcal{B}^3_{(b,\,e)}$ are bounded polyhedra. Under the assumption $\lambda_{ij} \le \theta_i \gamma_{ij}$, their respective sets of extreme points are listed in Tables 1 and 2.

TABLE 1. Extreme points of $\mathcal{X}^i_{(K_i,e)}$ for $i=1,2$.

(q_{i1},q_{i2},d_i)	(q_{i1},q_{i2},d_i)
$(0,0,0)$	$(0,0,\theta_i K_i)$
$(K_i/\gamma_{i1},0,0)$	$(K_i/\gamma_{i1},0,(\theta_i-\lambda_{i1}/\gamma_{i1})K_i)$
$(0,K_i/\gamma_{i2},0)$	$(0,K_i/\gamma_{i2},(\theta_i-\lambda_{i2}/\gamma_{i2})K_i)$

TABLE 2. Extreme points of $\mathcal{X}^3_{(b,e)}$.

(x_1,x_2)
$(0,0)$
$(b,0)$
$(0,b)$

Thus, for $i=1,2$, $\mathcal{B}^i_{(K_i,e)}$ has six extreme points for all $K_i\ge 0$ (i.e., $|\mathcal{X}^i_e|\equiv 6$), and $\mathcal{B}^3_{(b,e)}$ has three extreme points for all $b\in[0,B]$ (i.e., $|\mathcal{X}^3_e|\equiv 3$). The extreme points satisfy (23) with

$$M^i(e)\equiv\begin{pmatrix}0 & 1/\gamma_{i1} & 0 & 0 & 1/\gamma_{i1} & 0\\ 0 & 0 & 1/\gamma_{i2} & 0 & 0 & 1/\gamma_{i1}\\ 0 & 0 & 0 & \theta_i & \theta_i-\lambda_{i1}/\gamma_{i1} & \theta_i-\lambda_{i2}/\gamma_{i2}\end{pmatrix},\quad c^i(e)\equiv 0,\ i=1,2$$

$$M^3(e)\equiv\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},\quad c^3(e)\equiv 0.$$

Therefore, this problem is a decomposable affine MDP.

Following the steps as in the proof for Theorem 1 with $v^0(K_1,K_2,b,e)=r_1K_1+r_2K_2$, we obtain, for $\tau=1,\dots,T$,

$$v^\tau(K_1,K_2,b,e)=f_1^\tau(e)K_1+f_2^\tau(e)K_2+f_3^\tau(e)b, \tag{48}$$

where $(f_1^\tau(e),f_2^\tau(e),f_3^\tau(e))$ satisfy the following recursion with $f_1^0(e)=r_1$, $f_2^0(e)=r_2$, $f_3^0(e)=0$, and

$$f_i^\tau(e)=\max\{D_{i1}^{\tau-1}(e),D_{i2}^{\tau-1}(e),D_{i3}^{\tau-1}(e),D_{i4}^{\tau-1}(e),\ D_{i5}^{\tau-1}(e),D_{i6}^{\tau-1}(e)\},\quad i=1,2, \tag{49a}$$

$$f_3^\tau(e)=\max\{\beta\,\mathbb{E}[f_3^{\tau-1}(\xi(e))],\beta y_1\,\mathbb{E}[f_1^{\tau-1}(\xi(e))],\beta y_2\,\mathbb{E}[f_2^{\tau-1}(\xi(e))]\}, \tag{49b}$$

in which for $i=1,2$,

$$D_{i1}^{\tau-1}(e)=\beta\theta_i\,\mathbb{E}[f_i^{\tau-1}(\xi(e))], \tag{50a}$$

$$D_{i2}^{\tau-1}(e)=\frac{p_{i1}}{\gamma_{i1}}+\beta\,\mathbb{E}[f_i^{\tau-1}(\xi(e))]\left(\theta_i-\frac{\lambda_{i1}}{\gamma_{i1}}\right), \tag{50b}$$

$$D_{i3}^{\tau-1}(e)=\frac{p_{i2}}{\gamma_{i2}}+\beta\,\mathbb{E}[f_i^{\tau-1}(\xi(e))]\left(\theta_i-\frac{\lambda_{i2}}{\gamma_{i2}}\right), \tag{50c}$$

$$D_{i4}^{\tau-1}(e)=r_i\theta_i, \tag{50d}$$

$$D_{i5}^{\tau-1}(e)=\frac{p_{i1}}{\gamma_{i1}}+r_i\left(\theta_i-\frac{\lambda_{i1}}{\gamma_{i1}}\right), \tag{50e}$$

$$D_{i6}^{\tau-1}(e)=\frac{p_{i2}}{\gamma_{i2}}+r_i\left(\theta_i-\frac{\lambda_{i2}}{\gamma_{i2}}\right). \tag{50f}$$

In order of their appearance, label the six maximands in (49a) from one to six, and label the three maximands in (49b) from one to three. Let $J^{\tau-1,i}(e)$ denote the label of an optimal maximand in (49a) for $i=1,2$, and let $J^{\tau-1,3}(e)$ denote the label of an optimal maximand

in (49b). With τ periods remaining in the horizon, an optimal single-period decision rule $(A^{\tau 1}, A^{\tau 2}, A^{\tau 3})$ sets $a^i(e) = (q_{i1}, q_{i2}, d_i)$ to the $J^{\tau-1,i}(e)$th extreme point in $\mathcal{X}^i_{(K_i,e)}$ for $i = 1, 2$ and sets $a^3(e) = (x_1, x_2)$ to the $J^{\tau-1,3}(e)$th extreme point in $\mathcal{X}^3_{(b,e)}$. That is,

$$A^{\tau i}(K_1, K_2, b, e) = M^i_{\cdot, J^{\tau-1,i}(e)}(e) \times K_i, \quad i = 1, 2, \tag{51a}$$

$$A^{\tau 3}(K_1, K_2, b, e) = M^3_{\cdot, J^{\tau-1,i}(e)}(e) \times b. \tag{51b}$$

5. Summary

This tutorial introduces *decomposable affine MDPs*, which have continuous vector-valued endogenous states and actions and an exogenous state that follows an exogenous Markov chain. A decomposable affine MDP is characterized by dynamical equations and a single-period reward that are affine functions of the endogenous state and action, a decomposable action space, and polyhedral features of the decomposed action space.

We consider a decomposable affine MDP with a finite-horizon criterion and show that it is free of the curse of dimensionality. Specifically, it has a value function that is an affine function of the endogenous state and an optimal policy that is extremal and affine in the endogenous state. The linear coefficients in the value function and extremal optimal policy depend on the solution of a set of recursive auxiliary equations, which are indexed by the exogenous state and do not involve the endogenous state. This exorcises the curse of dimensionality due to a discrete approximation of the continuous state and action spaces. When the exogenous state follows a finite Markov chain, the auxiliary equations, and consequently the corresponding decomposable affine MDP, can be solved easily and exactly.

The applications in the tutorial imply the potential applicability of decomposable affine MDPs in large-scale problems in contexts such as fishery management and dynamic capacity portfolio management.

References

[1] C. N. K. Anderson, C. H. Hsieh, S. A. Sandin, R. Hewitt, A. Hollowed, J. Beddington, R. M. May, and G. Sugihara. Why fishing magnifies fluctuations in fish abundance. *Nature* 452(17): 835–839, 2008.

[2] S. A. Berkeley, C. Chapman, and S. M. Sogard. Maternal age as a determinant of larval growth and survival in a marine fish, *Sebastes melanops*. *Ecology* 85(5):1258–1264, 2004.

[3] G. B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ, 1963.

[4] E. V. Denardo and U. G. Rothblum. Affine dynamic programming. M. L. Puterman, ed. *Dynamic Programming and Its Applications*. Academic Press, New York, 255–267, 1979.

[5] E. V. Denardo and U. G. Rothblum. Affine structure and invariant policies for dynamic programs. *Mathematics of Operations Research* 8(3):342–365, 1983.

[6] K. Mathur and D. Solow. *Management Science: The Art of Decision Making*. Prentice-Hall, Englewood Cliffs, NJ, 1994.

[7] J. Ning and M. J. Sobel. Production and capacity management with internal financing. *Manufacturing Service Oper. Management*, 2017. Forthcoming.

[8] J. Ning and M. J. Sobel. Easy affine Markov decision processes: Algorithms and applications. Working paper, Case Western Reserve University, Cleveland, https://ssrn.com/abstract=2998786, 2017.

[9] J. Ning and M. J. Sobel. Easy affine Markov decision processes: Theory. Working paper, Case Western Reserve University, Cleveland, https://ssrn.com/abstract=2959096, 2017.

[10] W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. John Wiley & Sons, Hoboken, NJ, 2007.

[11] H. A. Simon. Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica* 24(1):74–81, 1956.

[12] M. J. Sobel. Higher-order and average reward myopic-affine dynamic models. *Mathematics of Operations Research* 15(2):299–310, 1990.

[13] M. J. Sobel. Myopic solutions of affine dynamic models. *Operations Research* 38(2):847–853, 1990.

[14] O. Tahvonen. Economics of harvesting age-structured fish populations. *Journal of Environmental Economics and Management* 58(3):281–299, 2009.

[15] H. Theil. A note on certainty equivalence in dynamic planning. *Econometrica* 25(2):346–349, 1957.

[16] J. A. Van Mieghem. Capacity management, investment and hedging: Review and recent developments. *Manufacturing & Service Operations Management* 5(4):269–302, 2003.

[17] L. Zéphyr, P. Lang, B. F. Lamond, and P. Côté. Approximate stochastic dynamic programming for hydroelectric production planning. *European Journal of Operational Research* 262(2): 586–601, 2017.