Finite-Horizon Approximate Linear Programs for Capacity Allocation over a Rolling Horizon

Thomas W.M. Vossen, Fan You, Dan Zhang
Leeds School of Business, University of Colorado at Boulder
UCB 419, 995 Regent Dr, Boulder, CO 80309
{thomas.vossen, fan.you, dan.zhang}@colorado.edu

Approximate linear programs have been used extensively to approximately solve stochastic dynamic programs that suffer from the well-known curse of dimensionality. Due to canonical results establishing the optimality of stationary value functions and policies for infinite-horizon dynamic programs, the literature has largely focused on approximation architectures that are stationary over time. In a departure from this literature, we apply a non-stationary approximation architecture to an infinite-dimensional linear programming formulation of the stochastic dynamic programs. We solve the resulting problems using a finite-horizon approximation. Such finite-horizon approximations are common in the theoretical analysis of infinite-horizon linear programs, but have not been considered in the approximate linear programming literature. We illustrate the approach on a rolling-horizon capacity allocation problem using an affine approximation architecture. We obtain three main results. First, non-stationary approximations can substantially improve upper bounds on the optimal revenue. Second, the upper bounds from the finite-horizon approximation are monotonically decreasing as the horizon length increases, and converge to the upper bound from the infinite-horizon approximation. Finally, the improvement does not come at the expense of tractability, as the resulting approximate linear programs admit compact representations and can be solved efficiently. The resulting approximations also produce strong heuristic policies and significantly reduce optimality gaps in numerical experiments.

History: January 14, 2022

1. Introduction

Linear programming-based approximate dynamic programming is a popular method to approximately solve stochastic dynamic programs that suffer from the well-known curse of dimensionality. A successful implementation of the method requires the choice of an appropriate approximation architecture. Typically, the value function is approximated by a weighted sum of pre-selected basis functions, an approach introduced by Schweitzer and Seidmann (1985) and theoretically analyzed in the seminal work of de Farias and Van Roy (2003). The method entails solving a large-scale linear program, often called the approximate linear program (ALP), to determine the weights on the basis functions.

Much of the existing literature on ALPs dealing with infinite-horizon dynamic programs assumes that the weights are stationary; that is, the weights are time-independent and constant across time periods; see, e.g., Adelman and Mersereau (2008) and Diamant et al. (2018) . The stationarity restriction can be justified by canonical results for infinite-horizon dynamic programs. Specifically, for a broad class of practically relevant infinite-horizon stochastic dynamic programs, it is known that there exist optimal value functions and optimal control policies that are stationary; see, e.g., Puterman (1994).

However, the optimality of stationary value functions and policies only applies to *exact* dynamic programming. Imposing stationarity restrictions on an approximation architecture (which itself restricts the value function) will necessarily lead to a weaker approximation. It is trivial to show that requiring the weights to be the same across periods would lead to a weaker bound compared with an approximation without such restrictions. This consideration motivates us to consider an infinite-dimensional formulation of the stochastic dynamic program, where the value functions are time-dependent, and investigate both analytically and numerically the bound improvements in comparison to a time-independent formulation. Applying the approximation architecture to this formulation results in an infinite-dimensional ALP.

We solve the infinite-dimensional ALP using a so-called finite-horizon approximation, where the weights are time-dependent within a chosen time horizon and are stationary afterwards. The use of such finite-horizon approximations for infinite-dimensional linear programming problems is well-established in the literature (Grinold 1977, Chand et al. 2002, Ghate 2010), and we note that this approximation is similar to the finite-horizon approximations in the analysis of infinite-horizon discounted stochastic dynamic programming problems. In both settings, the key to the analysis is the vanishing difference in value functions when the horizon length is sufficiently large.

We illustrate our approach on a rolling-horizon capacity allocation problem. In our setting, a single service provider with finite daily capacity can book requests from multiple customer classes at most M days in advance, where M is the length of the booking horizon. Customers can be booked in any of the M days, each with a corresponding reward. At the beginning of each day, the service provider observes both the number of bookings in the planning horizon and the incoming demand. The provider decides how to allocate the incoming demand to different days in the booking horizon, and unallocated demand is lost. The objective of the service provider is to maximize total discounted revenue over the infinite time horizon, and the problem can be formulated as a discrete-time infinite-horizon discounted reward Markov decision process.

The problem we consider is faced by many businesses that allocate limited capacity to different types of customers over time. In such settings, it is customary to satisfy demand using capacity on different days.

One example is medical appointment scheduling, where patients requesting appointments can be placed in different days depending on capacity availability. Most patients have a time preference for their appointments, therefore the "rewards" from different days can differ. In appointment scheduling, the key trade-off is between allocating arriving customers to different dates taking into account their preferences and rejecting them to reserve capacity for future customers with potentially higher values. In this context, rolling-horizon dynamic programming formulations are adopted in both Patrick et al. (2008) and Liu et al. (2010), among others. Rohleder and Klassen (2002) conduct a simulation study for rolling-horizon appointment scheduling problems. Our framework can also be applied to lead-time quotation in so-called available-to-promise systems, which allow manufacturers to dynamically (re)allocate resources to promise and fulfill customer orders; see Reindorp and Fu (2011). Other applications include scheduling (Ovacik and Uzsoy 1995) and production planning (Sridharan et al. 1987), even though these applications have additional details not considered in our problem formulation. More broadly, rolling-horizon decision making is very popular in practice and many practical problems can be formulated as rolling-horizon dynamic programs. A relevant literature review is given by Chand et al. (2002), and an early theoretical treatment is offered by Sethi and Sorger (1991). Finally, our overall approach can also be used in other applications that involve decision making over an infinite horizon, such as the multi-class queueing and inventory control problems discussed in Brown and Haugh (2017) or dynamic product promotion (Ye et al. 2018).

Even though we consider a revenue maximization objective in the current paper, our model can also handle cost minimization objectives by treating cost as negative revenue. For appointment scheduling problems, both revenue maximization and cost minimization objectives have been considered in the literature; see, e.g., Liu et al. (2010) and Patrick et al. (2008). The paper by Patrick et al. (2008) is particularly relevant. The motivation of Patrick et al. (2008) is a multi-priority scheduling problem for image scanning facilities where patients have different target service dates. They point out that the problem formulation is quite generic and can be applied to many other healthcare settings as well. Our model captures a stylized version of the problem they consider, largely to streamline exposition and emphasize key insights. Even though we ignore certain operational features such as demand carryover or surge capacity, our framework can be extended to incorporate these as well.

We apply a finite-horizon approximation together with an affine approximation, where the value function is represented by an affine function of the state vector. In contrast to stationary approximations that are customary in most of the existing literature, the finite-horizon approximation can capture certain time dynamics. We show that, as a result, the approximation produces a tighter upper bound on the optimal revenue. Moreover, we introduce novel techniques to construct lower bounds for a given horizon length, and derive analytical expressions for the gap between these upper and lower bounds. This allows us to establish convergence to an infinite-horizon approximation bound for moderate horizon lengths, and can help guide the selection of an appropriate horizon length. In addition, we also show that the finite-horizon approximation enables heuristic policies that can improve revenue performance. In these heuristic control methods, the ALP is solved starting from the current state. Then, we probabilistically select an allocation that is used in the current period. This process is repeated in the following periods. This type of heuristic control policy can be viewed as a form of model predictive control (Mayne 2014); see also the extensive discussion in Bertsekas (2001). We perform a numerical study to verify our theoretical results and to investigate the performance of heuristic policies that are based on the finite-horizon approximation. Our experiments confirm that the finite horizon approximation produces both tighter bounds and stronger policies than the stationary affine approximation. Overall, the combination of those improvements leads to approximations that drastically reduce optimality gaps relative to the stationary affine approximation.

An important contribution of our work is that these improvements do not come at the expense of tractability. Specifically, we derive compact reduced formulations for the resulting ALPs, whose dimensions are linear in the problem inputs. The size of the resulting ALPs is proportional to the length of the booking horizon M, the number of customer classes N, and the length of the planning horizon T of our choice. When $T=0$, the finite-horizon approximation reduces to the stationary affine approximation. This is related to the recent work of Tong and Topaloglu (2013) and Vossen and Zhang (2015), which shows that certain ALPs for network revenue management problems (Adelman 2007) admit compact representations. However, the problem setup is quite different in the present paper. The network revenue management problem is a finite-horizon problem that assumes at most one customer arrival in each period, while the problem in this paper is formulated as an infinite-horizon problem with "block" demand from multiple customer classes. Block demand models are typically only considered for single-resource problems in the revenue management literature (Brumelle and McGill 1993, Robinson 1995). Therefore, the state dynamics and action space are more complicated than the corresponding dynamic programs in network revenue management.

While our approach is closely related to the fluid optimization approach proposed in Bertsimas and Mišić (2016), our results extend their work along several dimensions. First, Bertsimas and Mišić (2016) consider decomposable MDPs with small action spaces, whereas our framework in principle applies to any infinite-horizon MDP. Second, the fluid approximation in Bertsimas and Mišić (2016) can be viewed as a non-stationary separable piecewise linear value function approximation, while our framework can be applied in conjunction with any functional approximation scheme. Moreover, the convergence of the finite-horizon approximation of the infinite-horizon fluid approximation is only observed empirically in the numerical study in Bertsimas and Mišić (2016), while we derive analytical bounds that provide strong support for fast convergence. In addition, we propose more general methods for constructing heuristic control policies. Overall, our work therefore generates several new insights on the benefits of finite-horizon approximations for complex infinite-dimensional dynamic programs.

A natural question is how finite-horizon approximations compare with stationary approximations that use a stronger functional approximation. One powerful functional approximation is the separable piecewise linear approximation; see, e.g., Farias and Van Roy (2007) and Meissner and Strauss (2012). Instead of a constant marginal value for each available booking and arriving customer in each period in the affine approximation, the separable piecewise linear approximation considers marginal values that depend on the booking level and number of arriving customers. The resulting ALPs are much larger in comparison to the affine ALPs, and tend to be much harder to solve. We conduct a numerical study to compare affine finite-horizon approximation and the stationary separable piecewise linear approximation (noting that finite-horizon separable piecewise linear approximation are computationally prohibitive); the details are relegated to Appendix A. Our numerical study shows that the affine finite-horizon approximation significantly outperforms the separable piecewise linear approximation, in that it produces stronger bounds (except for very small problem instances) and solves to optimality orders of magnitude faster. This further illustrates a key benefit of using finite horizon approximations: instead of using stronger functional approximations in a stationary setting (which can still be computationally prohibitive), using a finite horizon approximation based on simpler approximation architectures (which are more likely to admit compact reformulations) can lead to improved performance.

The remainder of the paper is organized as follows. Section 2 introduces the model formulation. Section 3 presents the framework of linear programming based approximate dynamic programming, and describes the key ideas behind our approach. In Section 4, we apply these concepts using an affine value approximation and derive reduced formulations for the resulting ALPs. Section 5 considers the bound improvements from the finite horizon approximation. Section 6 proposes heuristic policy alternatives from the finite horizon approximation. Section 7 reports numerical results and Section 8 concludes. Appendix A reports a numerical comparison between the (stationary) separable piecewise linear approximation and the finite horizon affine approximation. Appendix B contains all technical proofs.

2. The Rolling-Horizon Capacity Allocation Problem

We consider a rolling-horizon capacity allocation problem, where each customer can reserve a unit of capacity at most M days in advance. We refer to M as the *booking horizon*. The capacity on each day is C units, and customer requests can come from N customer classes. For notational convenience, we introduce the index sets $\mathcal{M}=\{1,\dots,M\}$, $\mathcal{N}=\{1,\dots,N\}$, and $\mathcal{M}^-=\{1,\dots,M-\ 1\}$.

At the beginning of each day, a scheduler observes the number of bookings over the M-day booking horizon, denoted by the vector $\mathbf{b}=(b_1,\dots,b_M)$, as well as the demand vector $\mathbf{d}=(d_1,\dots,d_N)$. The demand from each class is independent and follows a discrete distribution with bounded support. The upper bound on demand for each class is D. The distribution of class-n demand is given by $p_n(\cdot)$ for each n. It follows that the joint distribution of demand vector $\mathbf{d}$ is

$$p(\mathbf{d})=\prod_{n\in\mathcal{N}}p_n(d_n).$$

Given this information, the scheduler decides upon an allocation $\mathbf{a}\in\mathbb{Z}_+^{M\times N}$ of the remaining capacity to the incoming demand, where $\mathbb{Z}_+$ denotes the set of nonnegative integers. The reward is linear, in that each unit of allocated class-n demand to day m will provide a reward $v_{m,n}$. Therefore, the total reward for the allocation $\mathbf{a}$ is $r(\mathbf{a})=\sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}}v_{m,n}a_{m,n}$. After allocating the demand, the number of bookings at the beginning of the next day will be

$$\left(b_2+\sum_{n\in\mathcal{N}}a_{2,n},\dots,b_M+\sum_{n\in\mathcal{N}}a_{M,n},0\right).$$

The expression above reflects the rolling horizon, as day m in the new booking horizon corresponds to day $m+1$ on the previous day. Since day M just entered the planning horizon, the number of bookings on day M equals 0.

As mentioned earlier, the key trade-off is between allocating arriving customers to different dates taking into account their preferences and rejecting them to reserve capacity for future customers with potentially higher values. We formulate the problem as an infinite-horizon discrete-time discounted Markov decision process (MDP) to systematically account for this tradeoff. We use $\beta\in(0,1)$ to denote the per period discount factor. The state space of this MDP is

$$\mathcal{S}=\{(\mathbf{b},\mathbf{d})\in\mathbb{Z}_+^{M+N-1}:b_m\le C,\quad\forall m\in\mathcal{M}^-,\ d_n\le D,\quad\forall n\in\mathcal{N}\}.$$

We omit the number of bookings on day M because it is always 0. However, we adopt the notational convention $b_M\equiv 0$ to simplify subsequent development. The action space reflects how the incoming requests are managed, where each request is either allocated a unit of capacity in the booking horizon or rejected. Thus, the set of feasible actions in state $(\mathbf{b},\mathbf{d})$ is

$$\mathcal{A}(\mathbf{b},\mathbf{d}) = \left\{ \mathbf{a} \in \mathbb{Z}_+^{M\times N} : \sum_{m\in\mathcal{M}} a_{m,n} \le d_n,\ \forall n \in \mathcal{N},\ \sum_{n\in\mathcal{N}} a_{m,n} \le C - b_m,\quad \forall m \in \mathcal{M} \right\}.$$

For notational simplicity, we also introduce the set of feasible state-action pairs

$$\mathcal{X} = \{(\mathbf{b},\mathbf{d},\mathbf{a}) : (\mathbf{b},\mathbf{d}) \in \mathcal{S}, \mathbf{a} \in \mathcal{A}(\mathbf{b},\mathbf{d})\}.$$

Finally, the state transition probabilities can be stated as

$$p(\mathbf{b}',\mathbf{d}'|\mathbf{b},\mathbf{d},\mathbf{a}) = \begin{cases} p(\mathbf{d}'), \text{ if } b'_m = b_{m+1} + \sum_{n=1}^{N} a_{m+1,n} \text{for } m \in \mathcal{M}^-, \\ 0, \qquad \text{otherwise}, \end{cases} \qquad \forall (\mathbf{b}',\mathbf{d}') \in \mathcal{S}, (\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}.$$

With these definitions, the dynamic programming optimality equations can be stated as

$$J(\mathbf{b},\mathbf{d}) = \max_{\mathbf{a}\in\mathcal{A}(\mathbf{b},\mathbf{d})} \left\{ r(\mathbf{a}) + \beta \sum_{(\mathbf{b}',\mathbf{d}')\in\mathcal{S}} p(\mathbf{b}',\mathbf{d}'|\mathbf{b},\mathbf{d},\mathbf{a}) J(\mathbf{b}',\mathbf{d}') \right\}, \qquad \forall (\mathbf{b},\mathbf{d}) \in \mathcal{S}. \tag{1}$$

The dynamic program (1) has an $(M+N-1)$-dimensional state space and an $(M\times N)$-dimensional action space, and cannot be easily solved by standard solution methods such as value iteration or policy iteration. In the remainder of the paper, we focus on the linear programming-based approximate dynamic programming approaches to solve the problem.

3. General Framework

In this section, we discuss the general ideas behind our approach. Section 3.1 introduces approximate linear programming. In the existing literature, ALPs typically use stationary value function approximations that are time-independent for infinite-horizon problems. Section 3.2 introduces non-stationary value function approximations. Section 3.3 proposes a finite-horizon approximation as a way to implement non-stationary value function approximations.

3.1 Approximate Linear Programming

Let $\boldsymbol{\alpha}$ be a vector of non-negative weights corresponding to each state $(\mathbf{b},\mathbf{d}) \in \mathcal{S}$, which are called state-relevance weights in the literature (de Farias and Van Roy 2003). The dynamic program (1) can be equivalently formulated as the following linear program:

$$\begin{aligned} \mathbf{P}^{\boldsymbol{\alpha}}: \quad z^{\boldsymbol{\alpha}} = \min_{\vartheta(\cdot)} & \sum_{(\mathbf{b},\mathbf{d})\in\mathcal{S}} \alpha(\mathbf{b},\mathbf{d}) \vartheta(\mathbf{b},\mathbf{d}) \\ \text{s.t.} \quad & \vartheta(\mathbf{b},\mathbf{d}) \ge r(\mathbf{a}) + \beta \sum_{(\mathbf{b}',\mathbf{d}')\in\mathcal{S}} p(\mathbf{b}',\mathbf{d}'|\mathbf{b},\mathbf{d},\mathbf{a}) \vartheta(\mathbf{b}',\mathbf{d}'), \qquad \forall (\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}. \end{aligned}$$

In the formulation above, the decision variables are $\vartheta(\mathbf{b},\mathbf{d})$ for all $(\mathbf{b},\mathbf{d})\in\mathcal{S}$. Without loss of generality, we can assume that the state-relevance weights $\boldsymbol{\alpha}$ sum to one. As a result, we can interpret these weights as a distribution over the initial states. Thus, the optimal objective value of $\mathbf{P}^{\boldsymbol{\alpha}}$ is the weighted value function; that is $z^{\boldsymbol{\alpha}}=\sum_{(\mathbf{b},\mathbf{d})\in\mathcal{S}}\alpha(\mathbf{b},\mathbf{d})J(\mathbf{b},\mathbf{d})$.

The linear program $\mathbf{P}^{\boldsymbol{\alpha}}$ suffers from the same curse of dimensionality as the original dynamic programming formulation. It is intractable for moderate M and N due to the potentially huge number of decision variables and constraints. To achieve tractability, it is common to resort to approximately solving the linear program by imposing a restriction on the value function. This can be done by representing the value function $\vartheta(\mathbf{b},\mathbf{d})$ using a collection of weighted basis functions. Consider a set of basis functions $\phi_k:\mathcal{S}\to\mathbb{R}$ for $k\in\mathcal{K}$, where $\mathcal{K}$ is some index set, and take

$$\vartheta(\mathbf{b},\mathbf{d})\approx\theta+\sum_{k\in\mathcal{K}}V_k\phi_k(\mathbf{b},\mathbf{d}),\quad\forall(\mathbf{b},\mathbf{d})\in\mathcal{S},\tag{2}$$

where V_k is a parameter that weighs basis function $\phi_k(\cdot)$, and θ is a constant offset.

Substituting (2) into $\mathbf{P}^{\boldsymbol{\alpha}}$ yields a linear program over the parameters θ and V_k, which is the so-called approximate linear program (ALP). The optimal objective value of the ALP provides an upper bound on the optimal objective value of $\mathbf{P}^{\boldsymbol{\alpha}}$, because the value function approximation restricts the feasible solutions of $\mathbf{P}^{\boldsymbol{\alpha}}$. Compared to $\mathbf{P}^{\boldsymbol{\alpha}}$, the number of decision variables in the ALP is substantially smaller when a moderate number of basis functions are used, though the number of constraints is still exponential in both the number of customer classes and the number of days in the booking horizon. This suggests that the dual of the ALP can be solved using column generation methods (Desrosiers and Lübbecke 2005), though alternative approaches such as constraint sampling (de Farias and Van Roy 2004) have also been proposed.

Within the context of infinite-horizon dynamic programming problems, the ALP approach outlined above has received considerable attention (e.g., de Farias and Van Roy 2003, Adelman and Mersereau 2008, Patrick et al. 2008, Diamant et al. 2018). However, it is important to note that this approach implies an a-priori restriction to *stationary* value function approximations. This restriction appears natural, given the optimality of stationary value functions in the original dynamic programming formulation. Yet, it is not clear that the restriction to stationary value functions is necessary or desirable when approximating the dynamic program. We will show below that removing the stationarity restriction can improve the strength of the approximation.

3.2 Non-Stationary Value Function Approximations

To apply non-stationary value function approximations, we first extend $\mathbf{P}^{\boldsymbol{\alpha}}$ to an infinite-dimensional linear programming formulation that corresponds to the dynamic programming recursion, which is given by

$$\mathbf{P}_{\infty}^{\boldsymbol{\alpha}}: \; z_{\infty}^{\boldsymbol{\alpha}} = \min_{\{\vartheta_t(\cdot)\}_{\forall t \in \mathbb{Z}_+}} \sum_{(\mathbf{b},\mathbf{d}) \in \mathcal{S}} \alpha(\mathbf{b},\mathbf{d}) \vartheta_0(\mathbf{b},\mathbf{d})$$
$$\text{s.t.} \;\; \vartheta_t(\mathbf{b},\mathbf{d}) \geq \; r(\mathbf{a}) + \beta \sum_{(\mathbf{b}',\mathbf{d}') \in \mathcal{S}} p(\mathbf{b}',\mathbf{d}'|\mathbf{b},\mathbf{d},\mathbf{a}) \vartheta_{t+1}(\mathbf{b}',\mathbf{d}'), \qquad \forall (\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}, t \in \mathbb{Z}_+.$$

Observe that $\vartheta_t(\mathbf{b},\mathbf{d})$ will be time-independent at optimality in this linear program, as it is the optimal value function for a stationary infinite-horizon discounted dynamic program. Therefore, $\vartheta_t(\mathbf{b},\mathbf{d}) = \vartheta(\mathbf{b},\mathbf{d})$ for all $(\mathbf{b},\mathbf{d}) \in \mathcal{S}$ and $t \in \mathbb{Z}_+$, and $\mathbf{P}_{\infty}^{\boldsymbol{\alpha}}$ and $\mathbf{P}^{\boldsymbol{\alpha}}$ are equivalent in that $z_{\infty}^{\boldsymbol{\alpha}} = z^{\boldsymbol{\alpha}}$.

To streamline our exposition, we further make two modifications to this formulation. First, we consider a version of $\mathbf{P}_{\infty}^{\boldsymbol{\alpha}}$ where the weights $\boldsymbol{\alpha}$ are concentrated on a given initial state $(\mathbf{b}_0,\mathbf{d}_0)$. This restriction is natural in the finite-horizon approximation we introduce later. In addition, we scale the decision variables $\vartheta_t(\cdot)$ by a factor β^{-t} for each t to simplify notation. The formulation after applying the two modifications is given by

$$\mathbf{P}_{\infty}(\mathbf{b}_0,\mathbf{d}_0): \; z_{\infty}(\mathbf{b}_0,\mathbf{d}_0) = \min_{\{\vartheta_t(\cdot)\}_{\forall t \in \mathbb{Z}_+}} \vartheta_0(\mathbf{b}_0,\mathbf{d}_0)$$
$$\text{s.t.} \;\; \vartheta_t(\mathbf{b},\mathbf{d}) \geq \beta^t r(\mathbf{a}) + \sum_{(\mathbf{b}',\mathbf{d}') \in \mathcal{S}} p(\mathbf{b}',\mathbf{d}'|\mathbf{b},\mathbf{d},\mathbf{a}) \vartheta_{t+1}(\mathbf{b}',\mathbf{d}'),$$
$$\forall (\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}, t \in \mathbb{Z}_+.$$

The natural dual (Romeijn et al. 1992) of $\mathbf{P}_{\infty}(\mathbf{b}_0,\mathbf{d}_0)$ is given by

$$\mathbf{D}_{\infty}(\mathbf{b}_0,\mathbf{d}_0): \max_{\{\pi_t(\cdot)\}_{\forall t \in \mathbb{Z}_+}} \sum_{t=0}^{\infty} \sum_{(\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}} \beta^t r(\mathbf{a}) \pi_t(\mathbf{b},\mathbf{d},\mathbf{a})$$
$$\text{s.t.} \sum_{\mathbf{a} \in \mathcal{A}(\mathbf{b},\mathbf{d})} \pi_t(\mathbf{b},\mathbf{d},\mathbf{a}) \; = \begin{cases} \mathbb{1}_{\{(\mathbf{b},\mathbf{d}) = (\mathbf{b}_0,\mathbf{d}_0)\}}, & \text{if } t = 0, \\ \displaystyle\sum_{(\mathbf{b}',\mathbf{d}',\mathbf{a}') \in \mathcal{X}} p(\mathbf{b},\mathbf{d}|\mathbf{b}',\mathbf{d}',\mathbf{a}') \pi_{t-1}(\mathbf{b}',\mathbf{d}',\mathbf{a}'), & \text{if } t > 0, \end{cases}$$
$$\forall (\mathbf{b},\mathbf{d}) \in \mathcal{S}, t \in \mathbb{Z}_+,$$
$$\boldsymbol{\pi}_t \; \geq \mathbf{0}, \qquad \forall t \in \mathbb{Z}_+.$$

In general, infinite-dimensional linear programming problems can pose a number of technical challenges; the objective function is not guaranteed to converge, and strong or even weak duality may fail to hold. However, it is possible to verify that weak duality, complementary slackness, and strong duality hold in our setting by viewing $\mathbf{P}_{\infty}(\mathbf{b}_0,\mathbf{d}_0)$ as a special case of the infinite linear programming formulation proposed for a class of nonstationary infinite-horizon Markov decision processes in Ghate and Smith (2013).

Using the same set of basis functions ϕ_k over the index set $\mathcal{K}$ as in (2), we can construct a time-dependent value function approximation

$$\vartheta_t(\mathbf{b},\mathbf{d}) \approx \theta_t + \sum_{k \in \mathcal{K}} V_{t,k} \phi_k(\mathbf{b},\mathbf{d}), \quad \forall (\mathbf{b},\mathbf{d}) \in \mathcal{S}, t \in \mathbb{Z}_+. \tag{3}$$

Plugging (3) into $\mathbf{P}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ yields the ALP

$$
\begin{aligned}
&\mathbf{ALP}_\infty(\mathbf{b}_0,\mathbf{d}_0):\\
&\quad \tilde{z}_\infty(\mathbf{b}_0,\mathbf{d}_0) = \min_{\theta,V}\ \theta_0 + \sum_{k\in\mathcal{K}} V_{0,k}\phi_k(\mathbf{b}_0,\mathbf{d}_0)\\
&\qquad\qquad \text{s.t. } \theta_t - \theta_{t+1} + \sum_{k\in\mathcal{K}}\left(V_{t,k}\phi_k(\mathbf{b},\mathbf{d}) - \sum_{(\mathbf{b}',\mathbf{d}')\in\mathcal{S}} p(\mathbf{b}',\mathbf{d}'|\mathbf{b},\mathbf{d},\mathbf{a})V_{t+1,k}\phi_k(\mathbf{b}',\mathbf{d}')\right) \geq \beta^t r(\mathbf{a}),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}, t\in\mathbb{Z}_+.
\end{aligned}
$$

We will refer to $\tilde{z}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ as the *infinite-horizon approximation bound*. The corresponding dual problem is

$$
\begin{aligned}
&\mathbf{ALD}_\infty(\mathbf{b}_0,\mathbf{d}_0):\\
&\max_{\{\pi_t(\cdot)\}\forall t\in\mathbb{Z}_+} \sum_{t=0}^{\infty}\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \beta^t r(\mathbf{a})\pi_t(\mathbf{b},\mathbf{d},\mathbf{a})\\
&\text{s.t.} \quad \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \pi_t(\mathbf{b},\mathbf{d},\mathbf{a}) = \begin{cases} 1, & \text{if } t=0,\\ \displaystyle\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \pi_{t-1}(\mathbf{b},\mathbf{d},\mathbf{a}), & \text{if } t>0, \end{cases} \qquad \forall t\in\mathbb{Z}_+,\\
&\qquad \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \phi_k(\mathbf{b},\mathbf{d})\pi_t(\mathbf{b},\mathbf{d},\mathbf{a}) = \begin{cases} \phi_k(\mathbf{b}_0,\mathbf{d}_0), & \text{if } t=0,\\ \displaystyle\sum_{\substack{(\mathbf{b},\mathbf{d})\in\mathcal{S},\\ (\mathbf{b}',\mathbf{d}',\mathbf{a}')\in\mathcal{X}}} \phi_k(\mathbf{b},\mathbf{d})p(\mathbf{b},\mathbf{d}|\mathbf{b}',\mathbf{d}',\mathbf{a}')\pi_{t-1}(\mathbf{b}',\mathbf{d}',\mathbf{a}'), & \text{if } t>0, \end{cases}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall k\in\mathcal{K}, t\in\mathbb{Z}_+,\\
&\qquad \boldsymbol{\pi}_t \geq \mathbf{0} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall t\in\mathbb{Z}_+.
\end{aligned}
$$

The decision variables $\pi_t(\mathbf{b},\mathbf{d},\mathbf{a})$ in $\mathbf{ALD}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ can be interpreted as approximate state-action probabilities (for further discussion, see Adelman and Mersereau 2008). Thus, the term $\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}}\phi_k(\mathbf{b},\mathbf{d})\pi_t(\mathbf{b},\mathbf{d},\mathbf{a})$ can be interpreted as the expected value of the basis function at the beginning of period t, and the constraints in $\mathbf{ALD}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ require that the expectations of the basis functions with respect to the state distribution are maintained over time.

It is worth noting that applying functional approximations to $\mathbf{P}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ where the weights are concentrated on a given initial state $(\mathbf{b}_0,\mathbf{d}_0)$ can lead to tighter bounds relative to applying functional approximations to $\mathbf{P}^{\boldsymbol{\alpha}}_\infty$ directly. Let $\tilde{z}^{\boldsymbol{\alpha}}_\infty$ be the optimal objective value of the infinite-horizon ALP formulation that applies (3) to $\mathbf{P}^{\boldsymbol{\alpha}}_\infty$, which we do not explicitly state for brevity. A standard result in stochastic programming implies that $\tilde{z}^{\boldsymbol{\alpha}}_\infty \geq \mathbb{E}_{\boldsymbol{\alpha}}[\tilde{z}_\infty(\mathbf{b},\mathbf{d})]$, where the expectation can be obtained using a sample average approximation (Shapiro 2003). This suggests that solving $\mathbf{ALP}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ or $\mathbf{ALD}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ for randomly sampled initial states from the distribution $\boldsymbol{\alpha}$ can produce a tighter bound, although this sample average bound is subject to statistical error and the inequality is not guaranteed to hold with a finite sample.

3.3 The Finite-Horizon Approximation

Solving the ALPs from the non-stationary value function approximations remains challenging. While the value function approximation does reduce the number of decision variables per time period, the ALPs still have an infinite number of variables and constraints. The predominant approach to solving infinite dimensional linear programs is based on using finite-horizon approximations that truncate the formulation beyond finitely many variables and constraints; see, e.g., Grinold (1977), Bean and Smith (1984), and Schochetman and Smith (1992).

We follow the approach in Grinold (1977), and start by defining a *planning horizon* $T \geq 0$. We also introduce the set notations $\mathcal{T} = \{0, \dots, T\}$ and $\mathcal{T}^- = \{0, \dots, T-1\}$. For any $t \geq T$, take $\theta_t = \beta^{t-T}\theta_T$ and $V_{t,k} = \beta^{t-T} V_{T,k}$ for all $k \in \mathcal{K}$; i.e., for all $t > T$, θ_t and V_t equal θ_T and V_T, respectively, before scaling. Plugging this into $\mathbf{ALP}_\infty(\mathbf{b}_0, \mathbf{d}_0)$ yields the linear program

$$\mathbf{ALP}_T(\mathbf{b}_0, \mathbf{d}_0):$$

$$\tilde{z}_T(\mathbf{b}_0, \mathbf{d}_0) = \min_{\theta, V} \; \theta_0 + \sum_{k \in \mathcal{K}} V_{0,k} \phi_k(\mathbf{b}_0, \mathbf{d}_0)$$

$$\text{s.t. } \theta_t - \theta_{t+1} + \sum_{k \in \mathcal{K}} \left(V_{t,k}\phi_k(\mathbf{b}, \mathbf{d}) - \sum_{(\mathbf{b}', \mathbf{d}') \in \mathcal{S}} p(\mathbf{b}', \mathbf{d}' | \mathbf{b}, \mathbf{d}, \mathbf{a}) V_{t+1,k} \phi_k(\mathbf{b}', \mathbf{d}') \right) \geq \beta^t r(\mathbf{a}),$$

$$\forall (\mathbf{b}, \mathbf{d}, \mathbf{a}) \in \mathcal{X}, t \in \mathcal{T}^-,$$

$$(1-\beta)\theta_T + \sum_{k \in \mathcal{K}} \left(V_{T,k}\phi_k(\mathbf{b}, \mathbf{d}) - \beta \sum_{(\mathbf{b}', \mathbf{d}') \in \mathcal{S}} p(\mathbf{b}', \mathbf{d}' | \mathbf{b}, \mathbf{d}, \mathbf{a}) V_{T,k} \phi_k(\mathbf{b}', \mathbf{d}') \right) \geq \beta^T r(\mathbf{a}),$$

$$\forall (\mathbf{b}, \mathbf{d}, \mathbf{a}) \in \mathcal{X}.$$

The corresponding dual is

$$\mathbf{ALD}_T(\mathbf{b}_0, \mathbf{d}_0):$$

$$\max_{\{\pi_t(\cdot)\}_{\forall t \in \mathbb{Z}_+}} \sum_{\substack{t \in \mathcal{T}, \\ (\mathbf{b}, \mathbf{d}, \mathbf{a}) \in \mathcal{X}}} \beta^t r(\mathbf{a}) \pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a})$$

$$\text{s.t. } (1 - \mathbb{1}_{\{t=T\}}\beta) \sum_{(\mathbf{b}, \mathbf{d}, \mathbf{a}) \in \mathcal{X}} \pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a}) = \begin{cases} 1, & \text{if } t = 0, \\ \displaystyle\sum_{(\mathbf{b}, \mathbf{d}, \mathbf{a}) \in \mathcal{X}} \pi_{t-1}(\mathbf{b}, \mathbf{d}, \mathbf{a}), & \text{if } t > 0, \end{cases} \qquad \forall t \in \mathcal{T},$$

$$\sum_{(\mathbf{b}, \mathbf{d}, \mathbf{a}) \in \mathcal{X}} \phi_k(\mathbf{b}, \mathbf{d}) \pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a}) - \mathbb{1}_{\{t=T\}}\beta \sum_{\substack{(\mathbf{b}, \mathbf{d}) \in \mathcal{S}, \\ (\mathbf{b}', \mathbf{d}', \mathbf{a}') \in \mathcal{X}}} \phi_k(\mathbf{b}, \mathbf{d}) p(\mathbf{b}, \mathbf{d} | \mathbf{b}', \mathbf{d}', \mathbf{a}') \pi_t(\mathbf{b}', \mathbf{d}', \mathbf{a}')$$

$$= \begin{cases} \phi_k(\mathbf{b}_0, \mathbf{d}_0), & \text{if } t = 0, \\ \displaystyle\sum_{\substack{(\mathbf{b},\mathbf{d})\in\mathcal{S}, \\ (\mathbf{b}',\mathbf{d}',\mathbf{a}')\in\mathcal{X}}} \phi_k(\mathbf{b},\mathbf{d}) p(\mathbf{b},\mathbf{d}|\mathbf{b}',\mathbf{d}',\mathbf{a}')\pi_{t-1}(\mathbf{b}',\mathbf{d}',\mathbf{a}'), & \text{if } t > 0, \end{cases} \qquad \forall k \in \mathcal{K}, t \in \mathcal{T},$$

$$\boldsymbol{\pi}_t \geq \mathbf{0}, \qquad \forall t \in \mathcal{T}.$$

We refer to $\mathbf{ALP}_T(\mathbf{b}_0, \mathbf{d}_0)$ as the *finite-horizon ALP*. Note that the decision variables $\pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a})$ in $\mathbf{ALD}_T(\mathbf{b}_0, \mathbf{d}_0)$ can again be interpreted as approximate state-action probabilities for period $t \in \mathcal{T}^-$; however, the decision variables $\pi_T(\mathbf{b}, \mathbf{d}, \mathbf{a})$ have an interpretation as the total discounted time (from period T) spent in state $(\mathbf{b}, \mathbf{d})$ taking action $\mathbf{a}$.

Observe that, when $T = 0$, $\mathbf{ALP}_T(\mathbf{b}_0, \mathbf{d}_0)$ is equivalent to the stationary approximation that results from plugging the value function approximation (2) into $\mathbf{P}^{\boldsymbol{\alpha}}$. Grinold (1977) shows that the finite horizon approximation converges as the planning horizon T increases, that is,

$$\lim_{T\to\infty} \tilde{z}_T(\mathbf{b}_0, \mathbf{d}_0) = \tilde{z}_\infty(\mathbf{b}_0, \mathbf{d}_0).$$

Moreover, Grinold also shows that this convergence is monotone, that is, $\tilde{z}_{T+1}(\mathbf{b}_0, \mathbf{d}_0) \leq \tilde{z}_T(\mathbf{b}_0, \mathbf{d}_0)$ for all $T \geq 0$. Given that $\tilde{z}_\infty(\mathbf{b}_0, \mathbf{d}_0)$ provides an upper bound on the stochastic dynamic program, that is, $\tilde{z}_\infty(\mathbf{b}_0, \mathbf{d}_0) \geq z_\infty(\mathbf{b}_0, \mathbf{d}_0)$, the finite-horizon approximation will therefore produce an upper bound for any planning horizon T. We summarize these results in the following proposition.

Proposition 1 **(Grinold, 1977)**. *For any initial state $(\mathbf{b}_0, \mathbf{d}_0) \in \mathcal{S}$, $\tilde{z}_T(\mathbf{b}_0, \mathbf{d}_0)$ decreases monotonically in T and converges to a limit $\tilde{z}_\infty(\mathbf{b}_0, \mathbf{d}_0)$, which is an upper bound on $z_\infty(\mathbf{b}_0, \mathbf{d}_0)$.*

Solving the finite-horizon ALPs can still pose a challenge. A common approach is to solve the dual problem $\mathbf{ALD}_T(\mathbf{b}_0, \mathbf{d}_0)$ by a column generation procedure, which uses only a small subset of the decision variables and adds more variables only when needed. This, however, can present a computational burden even for stationary approximations (i.e., when $T = 0$). In the next section, we address this issue and construct a compact formulation that is equivalent to the finite-horizon ALP when considering the affine value function approximation.

4. Affine Finite-Horizon Approximations

In this section, we consider the affine value function approximation

$$\vartheta_t(\mathbf{b}, \mathbf{d}) \approx \theta_t + \sum_{m\in\mathcal{M}^-} V_{t,m} b_m + \sum_{n\in\mathcal{N}} W_{t,n} d_n, \qquad \forall (\mathbf{b}, \mathbf{d}) \in \mathcal{S}, t \in \mathcal{T}. \tag{4}$$

Here, $V_{t,m}$ represents the total discounted value of a booking on day m, and $W_{t,n}$ represents the value of an additional class-n customer. As before, θ_t represents an adjusting constant.

Using (4) as the functional approximation in $\mathbf{ALP}_T(\mathbf{b}_0, \mathbf{d}_0)$, we obtain

$$
\begin{aligned}
&\mathbf{AFP}_T(\mathbf{b}_0, \mathbf{d}_0):\\
&\quad \tilde{z}_T(\mathbf{b}_0, \mathbf{d}_0) = \min_{\theta, V, W}\ \theta_0 + \sum_{m\in\mathcal{M}^-} b_{0,m}V_{0,m} + \sum_{n\in\mathcal{N}} d_{0,n}W_{0,n}\\
&\qquad\qquad \text{s.t.}\ \ \theta_t - \theta_{t+1} + \sum_{m\in\mathcal{M}^-}\left(b_m V_{t,m} - (b_{m+1} + \sum_{n\in\mathcal{N}} a_{m+1,n})V_{t+1,m}\right)\\
&\qquad\qquad\qquad + \sum_{n\in\mathcal{N}}\left(d_n W_{t,n} - \bar{d}_n W_{t+1,n}\right) \geq \beta^t r(\mathbf{a}),\\
&\qquad\qquad\qquad\qquad \forall (\mathbf{b}, \mathbf{d}, \mathbf{a}) \in \mathcal{X}, t \in \mathcal{T}^-,\\
&\qquad\qquad (1-\beta)\theta_T + \sum_{m\in\mathcal{M}^-}\left(b_m - \beta(b_{m+1} + \sum_{n\in\mathcal{N}} a_{m+1,n})\right)V_{T,m}\\
&\qquad\qquad\qquad + \sum_{n\in\mathcal{N}}\left(d_n - \beta\bar{d}_n\right)W_{T,n} \geq \beta^T r(\mathbf{a}),\\
&\qquad\qquad\qquad\qquad \forall (\mathbf{b}, \mathbf{d}, \mathbf{a}) \in \mathcal{X}.
\end{aligned}
$$

Here, $\bar{d}_n$ is the expected number of arrivals for class-n customers during a period. The corresponding dual problem is

$\mathbf{AFD}_T(\mathbf{b}_0, \mathbf{d}_0):$

$$
\max_{\{\pi_t(\cdot)\}_{\forall t\in\mathbb{Z}_+}} \sum_{\substack{t\in\mathcal{T},\\ (\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}}} \beta^t r(\mathbf{a})\pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a})
$$

$$
\text{s.t.}\quad (1 - \mathbb{1}_{\{t=T\}}\beta)\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a}) = \begin{cases} 1, & \text{if } t = 0,\\ \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \pi_{t-1}(\mathbf{b}, \mathbf{d}, \mathbf{a}), & \text{if } t > 0, \end{cases} \qquad \forall t \in \mathcal{T}, \quad (5)
$$

$$
\begin{aligned}
&\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}}\left(b_m - \mathbb{1}_{\{t=T\}}\beta(b_{m+1} + \sum_{n\in\mathcal{N}} a_{m+1,n})\right)\pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a})\\
&\qquad = \begin{cases} b_{0,m}, & \text{if } t = 0,\\ \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}}\left(b_{m+1} + \sum_{n\in\mathcal{N}} a_{m+1,n}\right)\pi_{t-1}(\mathbf{b}, \mathbf{d}, \mathbf{a}), & \text{if } t > 0, \end{cases}\\
&\qquad\qquad \forall m \in \mathcal{M}^-, t \in \mathcal{T}, \quad (6)
\end{aligned}
$$

$$
\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}}\left(d_n - \mathbb{1}_{\{t=T\}}\beta\bar{d}_n\right)\pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a}) = \begin{cases} d_{0,n}, & \text{if } t = 0,\\ \bar{d}_n\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \pi_{t-1}(\mathbf{b}, \mathbf{d}, \mathbf{a}), & \text{if } t > 0, \end{cases} \qquad \forall t \in \mathcal{T}, n \in \mathcal{N}, \quad (7)
$$

$$
\boldsymbol{\pi}_t \geq \mathbf{0}, \qquad \forall t \in \mathcal{T}.
$$

Observe that constraint (5) can be simplified to

$$
\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a}) = \frac{1}{1 - \mathbb{1}_{\{t=T\}}\beta}, \qquad \forall t \in \mathcal{T}. \quad (8)
$$

It follows that constraint (7) simplifies to

$$\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} d_n \pi_t(\mathbf{b},\mathbf{d},\mathbf{a}) = \begin{cases} d_{0,n} + \mathbb{1}_{\{t=T\}} \frac{\beta \bar{d}_n}{1-\beta}, & \text{if } t=0, \\ \frac{\bar{d}_n}{1-\mathbb{1}_{\{t=T\}}\beta}, & \text{if } t>0, \end{cases} \quad \forall n \in \mathcal{N}, t \in \mathcal{T}. \tag{9}$$

The resulting ALP has $(T+1)\times(N+M)$ constraints, though its number of decision variables grows exponentially in both M and N. As noted before, it is possible to use column generation procedure to solve the resulting problems. However, this might still require significant computational effort, especially when the planning horizon T is large. We address this issue in the following proposition, which shows that the ALP reduces to a much smaller linear program.

Proposition 2. *The ALP dual* $\boldsymbol{AFD}_T(\mathbf{b}_0,\mathbf{d}_0)$ *is equivalent to the formulation*

$$\widehat{\boldsymbol{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0): \max_{\hat{\mathbf{a}},\hat{\mathbf{b}}} \sum_{\substack{t\in\mathcal{T},\\ m\in\mathcal{M}, n\in\mathcal{N}}} \beta^t v_{m,n} \hat{a}_{t,m,n}$$

$$\text{s.t.} \quad \hat{b}_{t,m} - \mathbb{1}_{\{t=T\}}\beta\left(\hat{b}_{t,m+1} + \sum_{n\in\mathcal{N}} \hat{a}_{t,m+1,n}\right) = \begin{cases} b_{0,m}, & \textit{if } t=0, \\ \hat{b}_{t-1,m+1} + \sum_{n\in\mathcal{N}} \hat{a}_{t-1,m+1,n}, & \textit{if } t>0, \end{cases} \quad \forall m\in\mathcal{M}^-, t\in\mathcal{T}, \tag{10}$$

$$\sum_{n\in\mathcal{N}} \hat{a}_{t,m,n} + \hat{b}_{t,m} \le \frac{C}{1-\mathbb{1}_{\{t=T\}}\beta}, \quad \forall m\in\mathcal{M}, t\in\mathcal{T}, \tag{11}$$

$$\sum_{m\in\mathcal{M}} \hat{a}_{t,m,n} \le \begin{cases} d_{0,n} + \mathbb{1}_{\{t=T\}} \frac{\beta \bar{d}_n}{1-\beta}, & \textit{if } t=0, \\ \frac{\bar{d}_n}{1-\mathbb{1}_{\{t=T\}}\beta}, & \textit{if } t>0, \end{cases} \quad \forall n\in\mathcal{N}, t\in\mathcal{T}, \tag{12}$$

$$\hat{\mathbf{a}} \ge \mathbf{0},$$

with $\hat{b}_{t,M} = 0$ *for all* t.

The proof of Proposition 2 relies on the structure of the subproblems when solving $\mathbf{AFD}_T(\mathbf{b}_0,\mathbf{d}_0)$ with a column generation method. If these column generation subproblems can be formulated as compact linear programming problems, the ALPs admit a compact linear programming formulation as well; refer to Vossen and Zhang (2015) for additional discussion.

Following Vossen and Zhang (2015), we call $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ the reduced formulation, which has a natural interpretation as a deterministic approximation to the original dynamic programming formulation. For periods $t\in\mathcal{T}^-$, constraints (11) and (12) allocate the expected demand of each customer class to the available capacity on each day in the booking horizon, where variable $\hat{b}_{t,m}$ is the expected booking level, and $\hat{a}_{t,m,n}$ is the allocation of expected demand; for period T, these constraints allocate the discounted total expected demand to the discounted expected total capacity, where $\hat{b}_{T,m}$ is the discounted expected booking level, and $\hat{a}_{T,m,n}$ is the allocation of discounted expected demand. Constraint (10) is a flow balance constraint that maintains the expected number of bookings $\hat{\mathbf{b}}$ over time. The formulation $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ is also closely related to the deterministic linear programming formulation (3)-(6) in Erdelyi and Topaloglu (2010); the main difference is that they consider a finite-horizon dynamic program and do not account for the truncation after period T.

5. Bounds and Convergence of the Finite-Horizon Approximation

Proposition 1 shows that the finite-horizon approximation produces an upper bound that improves as the planning horizon T increases and converges to the infinite-horizon approximation bound. For computational purposes, it is useful to know how far the bound is from the infinite-horizon approximation bound for different values of T. One way to do this is to construct lower bounds on the infinite-horizon approximation bound, given the solution to the finite-horizon approximation. In this section, we propose such lower bounds. Section 5.1 discusses a naïve lower bound. The gap between the finite-horizon approximation and the naïve lower bound vanishes as T increases, establishing the convergence of the finite-horizon approximation to the infinite-horizon approximation bound. Section 5.2 further proposes an improved lower bound, which is often much tighter than the naïve lower bound, allowing us to establish convergence of the finite-horizon affine approximation for much smaller planning horizons T.

5.1 A Naïve Lower Bound

To establish an initial lower bound, we first define $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0, \mathbf{d}_0)$ to be the infinite-dimensional reduced formulation that corresponds to the infinite horizon approximation $\widehat{\mathbf{ALD}}_\infty(\mathbf{b}_0, \mathbf{d}_0)$. We comment here that $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0, \mathbf{d}_0)$ is given by $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ for $T = \infty$. We choose not to explicitly state the formulation to avoid repetition. Next, we construct a feasible solution $(\mathbf{a}, \mathbf{b})$ to $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0, \mathbf{d}_0)$ based on an optimal solution $(\mathbf{a}^*, \mathbf{b}^*)$ for the dual problem $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$. To that end, we define

$$a_{t,m,n} = \begin{cases} a^*_{t,m,n} & \text{if } t < T, \\ 0 & \text{if } t \geq T, \end{cases} \qquad \forall t, m \in \mathcal{M}, n \in \mathcal{N},$$

$$b_{t,m} = \begin{cases} b^*_{t,m}, & \text{if } t < T, \\ b^*_{T-1,m+t-T+1}, & \text{if } t \geq T, m + t - T \leq M - 1, \\ 0, & \text{otherwise}, \end{cases} \qquad \forall t, m \in \mathcal{M}.$$

The resulting solution satisfies the constraints in $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0,\mathbf{d}_0)$, and has a solution value equal to $\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)-\beta^T\sum_{m\in\mathcal{M},n\in\mathcal{N}} v_{m,n}a^*_{T,m,n}$. Now, let $\overline{z}$ be some bound on the revenue that can be obtained in the final period of the time horizon. For example, we can define

$$\begin{aligned}
\overline{z} = \max_{\mathbf{a}} & \sum_{m\in\mathcal{M},n\in\mathcal{N}} v_{m,n}a_{m,n} \\
\text{s.t.} & \sum_{n\in\mathcal{N}} a_{m,n} \le \frac{1}{1-\beta}C, \quad \forall m\in\mathcal{M}, \\
& \sum_{m\in\mathcal{M}} a_{m,n} \le \frac{1}{1-\beta}\bar{d}_n, \quad \forall n\in\mathcal{N}, \\
& \mathbf{a} \ge \mathbf{0}.
\end{aligned}$$

From Proposition 1, we also know that the objective value $\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)$ is monotone decreasing in T for any initial state $(\mathbf{b}_0,\mathbf{d}_0)$, and that $\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)$ converges as T goes to infinity. As a result, we can establish both an upper and a lower bound on the infinite-horizon approximation, that is, $\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)\ge\tilde{z}_\infty(\mathbf{b}_0,\mathbf{d}_0)\ge\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)-\beta^T\overline{z}$. Due to discounting, the difference between the upper and lower bounds vanishes when T is large, and it follows that $\tilde{z}_\infty(\mathbf{b}_0,\mathbf{d}_0)=\lim_{T\to\infty}\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)$.

5.2 An Improved Lower Bound

In numerical tests, the bounds introduced in Section 5.1 require a large planning horizon T to achieve convergence to the infinite-horizon approximation bound. In this section, we propose an alternative lower bound. As in Section 5.1, we construct a solution $(\mathbf{a},\mathbf{b})$ to the infinite-horizon approximation $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ based on an optimal solution $(\mathbf{a}^*,\mathbf{b}^*)$ for the dual problem $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ for a given $T>0$. Let

$$a_{t,m,n} = \begin{cases} a^*_{t,m,n}, & \text{if } t<T, \\ (1-\beta)a^*_{T,m,n}, & \text{otherwise,} \end{cases} \qquad \forall m\in\mathcal{M}, n\in\mathcal{N}, t\in\mathbb{Z}_+, \tag{13}$$

$$b_{t,m} = \begin{cases} b^*_{t,m}, & \text{if } t<T, \\ (1-\beta)b^*_{T,m}, & \text{otherwise,} \end{cases} \qquad \forall m\in\mathcal{M}, t\in\mathbb{Z}_+. \tag{14}$$

The objective value of the solution $(\mathbf{a},\mathbf{b})$ to $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ equals $\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)$. This solution provides a lower bound to $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0,\mathbf{d}_0)$ if $(\mathbf{a},\mathbf{b})$ is feasible since the objective function maximizes revenue. However, $(\mathbf{a},\mathbf{b})$ defined in (13)–(14) may not be feasible for $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0,\mathbf{d}_0)$. While $(\mathbf{a},\mathbf{b})$ satisfies the constraints corresponding to (11)–(12) by construction, the corresponding flow balance constraints (10) may not be satisfied. Specifically, substituting $(\mathbf{a},\mathbf{b})$ into the corresponding flow balance constraint leads to

$$(1-\beta)b^*_{T,m} = b^*_{T-1,m+1}+\sum_{n\in\mathcal{N}} a^*_{T-1,m+1,n}, \qquad \forall m\in\mathcal{M}^-, \tag{15}$$

$$\beta b^*_{T,m} = \beta \left(b^*_{T,m+1} + \sum_{n\in\mathcal{N}} a^*_{T,m+1,n} \right), \qquad \forall m \in \mathcal{M}^-. \tag{16}$$

Here, we choose not to write down the constraint for $t < T$ since $(\mathbf{a}_t, \mathbf{b}_t) = (\mathbf{a}^*_t, \mathbf{b}^*_t)$ for $t < T$. The condition (15) results from the substitution when $t = T$ and can be interpreted as a flow balance constraint, while condition (16) results from the substitution when $t > T$ and can be interpreted as an equilibrium condition. The solution $(\mathbf{a}, \mathbf{b})$ constructed using (13)–(14) is feasible for $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0, \mathbf{d}_0)$ only if conditions (15)–(16) are satisfied.

However, this also suggests that it is possible to obtain a lower bound to $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0, \mathbf{d}_0)$ by constructing an *auxiliary* linear program that incorporates (15) and (16) as constraints. To illustrate this, we first note that, when $t = T$, constraint (10) in $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ equals

$$b^*_{T,m} - \beta \left[b^*_{T,m+1} + \sum_{n\in\mathcal{N}} a^*_{T,m+1,n} \right] = b^*_{T-1,m+1} + \sum_{n\in\mathcal{N}} a^*_{T-1,m+1,n}, \qquad \forall m \in \mathcal{M}^-. \tag{17}$$

We can construct an auxiliary linear program $\underline{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ by replacing constraints (17) with constraints (15)–(16) in $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$; intuitively, this corresponds to a disaggregation of constraints (17).

However, as indicated earlier, $(\mathbf{a}, \mathbf{b})$ constructed using (13)–(14) is feasible for $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0, \mathbf{d}_0)$ only if conditions (15)–(16) are satisfied; as a result, the auxiliary linear program might not be feasible when the planning horizon T is small: the equilibrium condition (16) requires that $b_{T,m} \geq b_{T,m+1}$ for all $m \in \mathcal{M}^-$, which may not be achievable when starting from an initial state where $b_{0,m} < b_{0,m+1}$ for some $m \in \mathcal{M}^-$. To account for this, we relax the auxiliary linear program and penalize potential constraint violations. Specifically, we relax the auxiliary program by allowing bookings that are present initially to be discarded. This yields the following formulation:

$\underline{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$:

$$\underline{z}_T(\mathbf{b}_0, \mathbf{d}_0) = \max_{\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{q}}} \sum_{\substack{t\in\mathcal{T},\\ m\in\mathcal{M}, n\in\mathcal{N}}} \beta^t v_{m,n} \hat{a}_{t,m,n} - \sum_{m\in\mathcal{M}^-} \rho_m \hat{q}_m$$

$$\text{s.t.} \quad (1 - \mathbb{1}_{\{t=T\}}\beta)\hat{b}_{t,m} = \begin{cases} b_{0,m} - \hat{q}_m, & \text{if } t = 0, \\ \hat{b}_{t-1,m+1} + \sum_{n\in\mathcal{N}} \hat{a}_{t-1,m+1,n}, & \text{if } t > 0, \end{cases} \quad \forall m \in \mathcal{M}^-, t \in \mathcal{T}, \tag{18}$$

$$\beta \hat{b}_{T,m} = \beta \left[\hat{b}_{T,m+1} + \sum_{n\in\mathcal{N}} \hat{a}_{T,m+1,n} \right], \qquad \forall m \in \mathcal{M}^-. \tag{19}$$

$$\sum_{n\in\mathcal{N}} \hat{a}_{t,m,n} + \hat{b}_{t,m} \leq \frac{C}{1 - \mathbb{1}_{\{t=T\}}\beta}, \qquad \forall m \in \mathcal{M}, t \in \mathcal{T}, \tag{20}$$

$$\sum_{m\in\mathcal{M}} \hat{a}_{t,m,n} \leq \begin{cases} d_{0,n}, & \text{if } t = 0, \\ \frac{\bar{d}_n}{1 - \mathbb{1}_{\{t=T\}}\beta}, & \text{if } t > 0, \end{cases} \qquad \forall n \in \mathcal{N}, t \in \mathcal{T}, \tag{21}$$

$$\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{q}} \ \geq \mathbf{0}.$$

In the above, we assume $\hat{b}_{t,M} = 0$ for all $t \in \mathcal{T}$. Of course, this requires penalty coefficients ρ_m that are "large enough" to ensure that $\underline{z}_T(\mathbf{b}_0, \mathbf{d}_0)$ provides a lower bound on the infinite-horizon approximation. The following proposition establishes such penalty coefficients.

Proposition 3. *Suppose $T > 0$, and let*

$$\rho_m = \max_{\substack{k \in \{1,\dots,m\},\\ n \in \mathcal{N}}} \beta^{m-k} v_{k,n}, \qquad \forall m \in \mathcal{M}^-. \tag{22}$$

Then,

(i) $\underline{z}_T(\mathbf{b}_0, \mathbf{d}_0) \leq \underline{z}_{T+1}(\mathbf{b}_0, \mathbf{d}_0)$, *and*

(ii) There exists an optimal solution $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*, \hat{\mathbf{q}}^*)$ *to* $\underline{\boldsymbol{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ *such that* $\hat{q}^*_m = 0$ *for all* $m \in \mathcal{M}^-$ *when* $T \geq M - 1$.

When $\hat{q}^*_m = 0$ for all $m \in \mathcal{M}^-$, the corresponding infinite horizon solution obtained using (13) and (14) will also be feasible for $\widehat{\mathbf{AFD}}_\infty(\mathbf{b}_0, \mathbf{d}_0)$. In that case, $\underline{z}_T(\mathbf{b}_0, \mathbf{d}_0)$ bounds the infinite-horizon approximation from below and we have

$$\tilde{z}_T(\mathbf{b}_0, \mathbf{d}_0) \geq \tilde{z}_\infty(\mathbf{b}_0, \mathbf{d}_0) \geq \underline{z}_T(\mathbf{b}_0, \mathbf{d}_0).$$

Because $\underline{z}_T(\mathbf{b}_0, \mathbf{d}_0)$ is monotone in T, $\underline{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ will therefore provide a lower bound for all $T > 0$.

We establish convergence in the following proposition. Proposition 4 implies that the difference between the upper and lower bounds on the infinite horizon approximation values (provided by the reduced formulation $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ and the auxiliary linear program $\underline{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$, respectively) vanishes as T increases.

Proposition 4. *Suppose $T > 0$, and let $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*)$ be an optimal solution for the dual problem* $\widehat{\boldsymbol{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$. *Then,*

$$\tilde{z}_T(\mathbf{b}_0, \mathbf{d}_0) - \underline{z}_T(\mathbf{b}_0, \mathbf{d}_0) \leq \sum_{m \in \mathcal{M}^-} \epsilon_{T,m} \left(\hat{b}^*_{T-1,m+1} + \sum_{n \in \mathcal{N}} \hat{a}^*_{T-1,m+1,n} \right) + \sum_{m \in \mathcal{M}^-} \gamma_{T,m} \left(\sum_{n \in \mathcal{N}} \hat{a}^*_{T,m+1,n} \right),$$

where

$$\epsilon_{T,m} = \begin{cases} \max \left(\rho_{T+m}, \max\limits_{\substack{k=1,\dots,\min(T,M-m),\\ n \in \mathcal{N}}} \beta^{T-k} v_{m+k,n} \right), & \textit{if } T + m \leq M - 1, \\ \max\limits_{\substack{k=1,\dots,\min(T,M-m),\\ n \in \mathcal{N}}} \beta^{T-k} v_{m+k,n}, & \textit{o.w.}, \end{cases} \qquad \forall m \in \mathcal{M}^-,$$

and $\gamma_{T,m}$ is defined recursively as

$$\gamma_{T,m} = \begin{cases} \max\limits_{n \in \mathcal{N}} \beta^T v_{M,n}, & \textit{if } m = M - 1, \\ \max \left(\max\limits_{n \in \mathcal{N}} \beta^T v_{m+1,n}, (1-\beta)\epsilon_{T,m+1} + \gamma_{T,m+1} \right), & \textit{o.w.}, \end{cases} \qquad \forall m \in \mathcal{M}^-.$$

The proof of Proposition 4 relies on the fact that the reduced formulation $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ can be obtained by aggregating constraints (15) and (16) in the auxiliary linear program $\underline{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$. Then, the inequality in Proposition 4 follows as an *a posteriori* bound on the error induced by aggregating these constraints. Such bounds have been studied extensively, and we refer to Zipkin (1980), Mendelssohn (1980), and Shetty and Taylor (1987) for details on this approach.

We note that constraints (10)–(11) in the reduced formulation imply that $\hat{b}^*_{t-1,m+1} + \sum_{n\in\mathcal{N}} \hat{a}^*_{t-1,m+1,n} \leq C$ and $\sum_{n\in\mathcal{N}} \hat{a}_{T,m+1,n} \leq \frac{C}{1-\beta}$ for all $m \in \mathcal{M}^-$ in Proposition 4. As a result, we can also use Proposition 4 and the penalty coefficients defined therein to construct an *a priori* bound on the gap that solely depends on an instance's input parameters. In principle, this a priori bound could be used to determine an appropriate horizon length T. However, because the auxiliary linear program $\underline{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ requires little computational overhead, an alternative approach is to gradually increase the horizon length T until the observed gap after solving both $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ and $\underline{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ is sufficiently small.

6. Heuristic Policies

In this section, we introduce heuristic policies that can be derived from the finite-horizon approximations. The relevant primal-dual pair considered is $\widehat{\mathbf{AFP}}_T(\mathbf{b}_0, \mathbf{d}_0)$–$\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$, where the primal formulation $\widehat{\mathbf{AFP}}_T(\mathbf{b}_0, \mathbf{d}_0)$ is not explicitly stated. Section 6.1 discusses the one-step greedy policies based on the solution of $\widehat{\mathbf{AFP}}_T(\mathbf{b}_0, \mathbf{d}_0)$ and Section 6.2 lays out probabilistic allocation policies based on the solution of $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$. As a benchmark, we also consider a myopic policy that disregards future impacts and selects an allocation that maximizes immediate revenues.

6.1 Primal Policies

Let $(\mathbf{V}^*, \mathbf{W}^*, \theta^*)$ be an optimal solution to $\widehat{\mathbf{AFP}}_T(\mathbf{b}_0, \mathbf{d}_0)$, which can either be obtained from $\widehat{\mathbf{AFP}}_T(\mathbf{b}_0, \mathbf{d}_0)$ or derived from the dual solution to $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$.

For a stationary approximation ($T = 0$), $V^*_{0,m}$ can be interpreted as the total discounted value (cost) of a booking on day m and $W^*_{0,n}$ as the value of an additional class-n customer. A one-step greedy policy based on these weights can be derived by substituting the resulting approximation (4) into the right hand side of (1) for each state $(\mathbf{b}, \mathbf{d})$, resulting in the optimization problem

$$\max_{\mathbf{a}\in\mathcal{A}(\mathbf{b},\mathbf{d})} r(\mathbf{a}) + \beta \left[\sum_{m\in\mathcal{M}^-} \left(b_{m+1} - \sum_{n\in\mathcal{N}} a_{m+1,n} \right) V^*_{0,m} + \sum_{n\in\mathcal{N}} \bar{d}_n W^*_{0,n} + \theta^* - 0 \right].$$

Omitting the constant terms in the objective function, the problem can be more explicitly stated as

$$\max_{\mathbf{a}} \sum_{m\in\mathcal{M}, n\in\mathcal{N}} (v_{m,n} - \beta V^*_{0,m-1}) a_{m,n}$$

$$\begin{aligned} \text{s.t.} \sum_{n\in\mathcal{N}} a_{m,n} &\leq C - b_m, \quad \forall m \in \mathcal{M}, \\ \sum_{m\in\mathcal{M}} a_{m,n} &\leq d_n, \qquad\quad \forall n \in \mathcal{N}, \\ \mathbf{a} &\geq \mathbf{0}. \end{aligned}$$

The optimization problem above is a transportation problem. As a result, the optimal solution is integral, even though the problem is expressed as a linear program with no explicit integrality constraints.

When $T > 0$, we instead base our value function approximations on the values of the decision variables $V^*_{1,m}$, which can be interpreted as the marginal value (cost) of a booking on day 1 at the end of the first period. Incorporating the scaling factor that we applied to construct non-stationary value function approximations in Section 3.2, the resulting optimization problem will be

$$\max_{\mathbf{a}\in\mathcal{A}(\mathbf{b},\mathbf{d})} r(\mathbf{a}) + \beta \left[\sum_{m\in\mathcal{M}^-} \left(b_{m+1} - \sum_{n\in\mathcal{N}} a_{m+1,n} \right) V^*_{1,m} + \sum_{n\in\mathcal{N}} \bar{d}_n W^*_{1,n} + \theta^* - 1 \right].$$

As before, this problem can be expressed as a linear program.

The one-step greedy policy can be applied both with and without resolving. With resolving, the ALP is resolved whenever the state is updated. Thus, the value function estimates that are used in the one-step greedy policy may also change after every update. Alternatively, we can also solve the reduced formulation once, using the expected number of bookings and demand based on the initial state distribution $\boldsymbol{\alpha}$ as the initial state. In this case, the value function estimates that are used in the one-step greedy policy remains constant.

6.2 Dual Policies

The finite horizon approximation also enables an alternative approach to constructing heuristic policies, based on the optimal solution to $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$. To illustrate this approach, suppose that $(\mathbf{a}^*, \mathbf{b}^*)$ is an optimal solution for $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ with $T > 0$. We focus on the first period and observe that

$$\sum_{n\in\mathcal{N}} a^*_{0,m,n} \leq C - b_{0,m}, \quad \forall m \in \mathcal{M}, \tag{23}$$

$$\sum_{m\in\mathcal{M}} a^*_{0,m,n} \leq d_{0,n}, \qquad \forall n \in \mathcal{N}. \tag{24}$$

It is important to note that $\mathbf{a}^*_0$ might be fractional, due to the linkages presented by constraints (10) in $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$. As a result, we cannot construct a policy that simply selects the allocation $\mathbf{a}^*_0$. However, the following proposition shows that $\mathbf{a}^*_0$ can be expressed as a convex combination of integer allocations that satisfy constraints (23)–(24).

Proposition 5. *Let $\tilde{\mathcal{A}}(\mathbf{b},\mathbf{d})$ be the real-valued superset of $\mathcal{A}(\mathbf{b},\mathbf{d})$ such that*

$$\tilde{\mathcal{A}}(\mathbf{b},\mathbf{d}) = \left\{ \mathbf{a} \in \mathbb{R}_+^{M\times N} : \sum_{m\in\mathcal{M}} a_{m,n} \le d_n,\ \forall n \in \mathcal{N},\ \sum_{n\in\mathcal{N}} a_{m,n} \le C - b_m,\quad \forall m \in \mathcal{M} \right\}.$$

Suppose $\mathbf{a}_0^* \in \tilde{\mathcal{A}}(\mathbf{b},\mathbf{d})$. Then $\mathbf{a}_0^*$ can be written as a convex combination of integer allocations in $\mathcal{A}(\mathbf{b},\mathbf{d})$. That is, there exist some positive integer K that is polynomially bounded by M and N and integer allocations $\mathbf{a}^k \in \mathcal{A}(\mathbf{b},\mathbf{d})$ for $k=1,\ldots,K$ such that

$$\mathbf{a}_0^* = \sum_{k=1}^{K} \lambda_k \mathbf{a}^k,$$

where $\lambda_k \ge 0$ for $k=1,\ldots,K$ and $\sum_{k=1}^{K} \lambda_k = 1$.

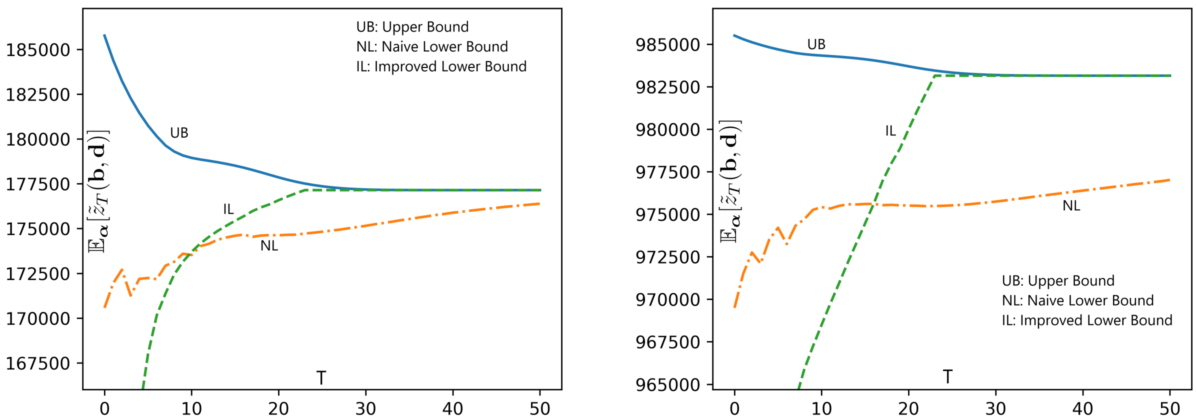
The proof of Proposition 5 relies on Birkhoff's algorithm to generate the integer allocations, and shows that the number of such allocations is polynomially bounded.

For our purposes, this result immediately suggests a probabilistic allocation policy: given an optimal solution to $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$, we generate a convex combination of integer allocations $\mathbf{a}^k$ from the first period allocation $\mathbf{a}_0^*$ and select allocation $\mathbf{a}_k$ with probability λ_k. Observe that this policy requires that we resolve $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ every time the state is updated.

7. Numerical Study

We conduct a numerical study to evaluate the bounds produced by our finite horizon approximation and to investigate policy performance. We consider both the one-step greedy policy (with and without resolving) and the probabilistic allocation policy introduced in Section 6.2. We also use the myopic policy as a benchmark.

Our test instances are similar to those in Section 6 of Patrick et al. (2008), although we consider neither demand carryover nor overtime capacity. These test instances consider an outpatient clinic with total capacity of 60 scans per day. We take the booking horizon length $M=30$, and there are three priority classes ($N=3$). In the base case, the daily demands for the three classes follow Poisson distributions with means of 30, 18, and 12, respectively. We define the system load as the ratio between the expected total demand and the regular daily capacity, which is given by $\rho = \sum_i \lambda_i / C$. Following this definition, the load in the base case is exactly 1. In our numerical experiments, we vary the load from 0.6 to 2 with stepsize 0.2 by scaling the mean demand in the base case by the load ρ. Rewards are determined by a combination of base rewards and target dates. The base rewards f_n equal 200, 100, and 50 for classes 1, 2, and 3, respectively, and we use target dates t_n of 7 days, 14 days, and 21 days, respectively. Following Patrick et al. (2008), the reward takes the form $v_{m,n} = \beta^{\max(0,m-t_n)} f_n$. Here, β is the per period discount factor and takes



(a) $\beta = 0.95, \rho = 1.4$

(b) $\beta = 0.99, \rho = 1.4$

Figure 1 Upper and Lower Bounds for Different Horizon Length T

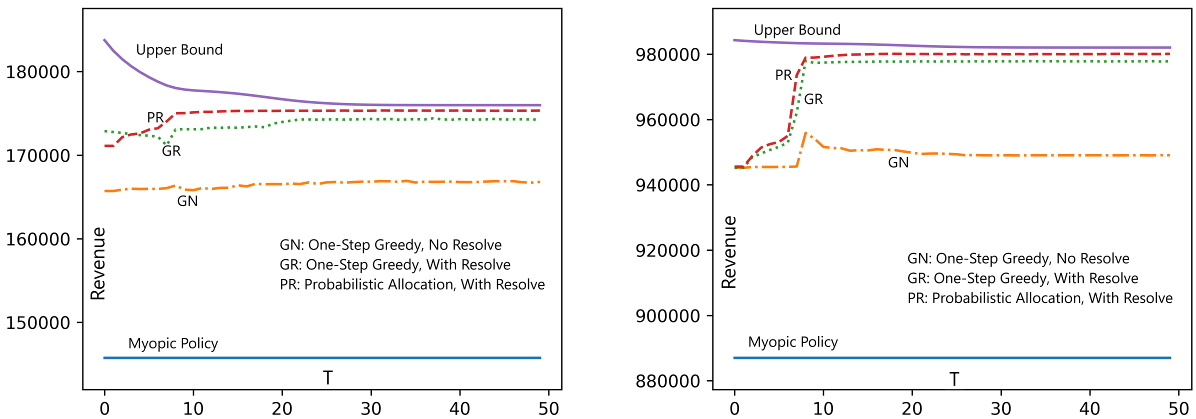
the values 0.95 and 0.99 in our numerical study. All experiments were run on a machine with an Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz and 64 GB RAM. All algorithms were implemented in `Python 3.6.5` and the linear programs were solved using `Gurobi 8.0.1`.

7.1 Approximation Bounds

As a first step, we evaluate the bounds produced by our finite horizon approximation. The questions we consider are (1) how does the length of the planning horizon impact the upper bound from the finite horizon approximations and (2) how fast do the finite horizon approximations converge to the infinite-horizon approximation bounds.

We evaluate the upper bounds obtained by solving $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ for different values of T (recall that the stationary affine approximation has $T = 0$). We determine these bounds by the sample average over 100 randomly generated initial states. The number of bookings in the initial state are sampled from the uniform distribution, while the demand is sampled from the corresponding Poisson distributions. For each T, we also infer lower bounds on the infinite horizon approximation using the two approaches introduced in Section 5. The computational times to solve the reduced formulations is negligible even for large T (e.g., less than 0.2 seconds when $T = 50$); thus, we do not report computational times for individual instances.

Figure 1 shows the results for two representative instances; they are representative of the overall patterns that emerged in our experiments. In particular, the finite horizon approximations can provide a substantial improvement in the upper bounds relative to the stationary approximation (corresponding to $T = 0$). The magnitude of improvement depends on the discount factor, and is



(a) $\beta = 0.95, \rho = 1.4$ (b) $\beta = 0.99, \rho = 1.4$

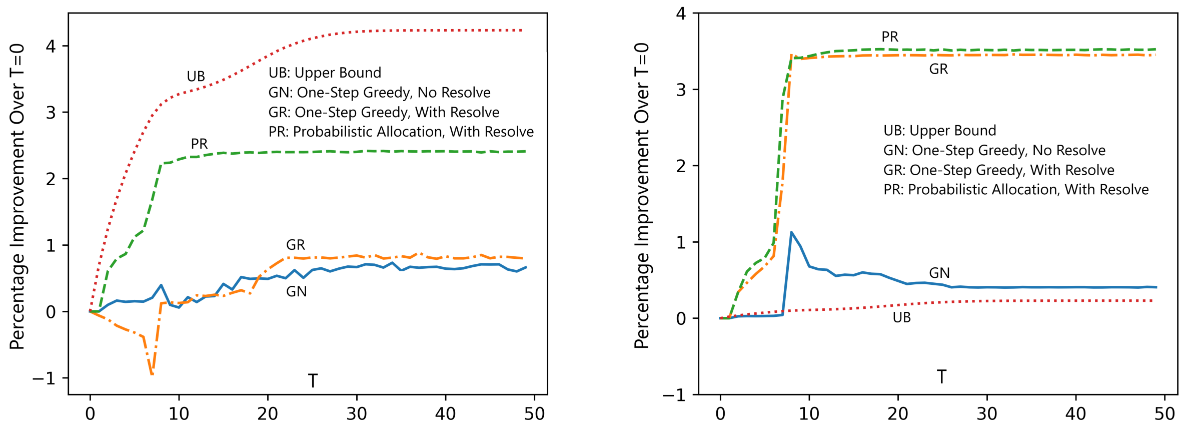
Figure 2 Performance of Heuristic Policies for Different Horizon Length T

larger for smaller β. In addition, these improvements can be obtained for moderate values of T. When the discount factor β is 0.95, convergence to within 1% and 0.01% of the infinite horizon approximation bound requires average T values of 19.7 and 30.4, respectively. Moreover, with a discount factor of 0.99, these numbers change to 16.0 and 27.1, respectively. Therefore, the lower and upper bounds from the finite horizon approximation appear to converge to the infinite horizon approximation bound rather quickly. We emphasize that the improved lower bounds obtained by solving the auxiliary linear program are often substantially stronger than the naïve lower bounds for moderate values of T and are therefore critical in establishing convergence. For example, when $\beta = 0.99$ the naïve lower bounds can still be quite far from the upper bounds even for $T = 300$. Using naïve lower bounds to establish convergence would have caused substantial, but unnecessary, computational overhead. Note that we choose not to plot the improved lower bounds for small T since they are quite loose in that region.

7.2 Heuristic Policy Performance

We evaluate the heuristic policies introduced in Section 6 for different horizon lengths T via simulation over 100 randomly generated initial states. Four policies are evaluated: one-step greedy with and without resolving, probabilistic allocation, and myopic. We collect the total discounted revenue for each simulation (starting on the first day), and terminate each simulation only when the number of days k is such that $\beta^k \leq 10^{-6}$. We remark that the myopic policy does not depend on T and is shown as a benchmark.

Representative results are shown in Figure 2. We observe that all of the approximation-based policies significantly outperform the myopic policy. Furthermore, resolving after each state update



(a) $\beta = 0.95, \rho = 1.4$

(b) $\beta = 0.99, \rho = 1.4$

Figure 3 Percentage Improvements over Stationary Affine ($T = 0$) for Different Horizon Length T

leads to substantial improvement in policy performance. The probabilistic allocation policy is the best overall and outperforms the one step greedy policies. Moreover, the results again indicate that policies derived from the finite horizon approximation tend to improve as the horizon length T increases.

The use of finite horizon approximations not only provides tighter bounds but also leads to performance improvements in the heuristic policies. Figure 3 shows the percentage improvements of the non-stationary finite-horizon approximation relative to those from the stationary approximation ($T = 0$). From Figure 3(a), when $\beta = 0.95$ and $\rho = 1.4$, the non-stationary finite-horizon approximation can improve the upper bound by 4.23% and the policy performance by up to 2.4%; together, these improvements reduce the optimality gap from 6.88% to 0.37%. In Figure 3(b), with $\beta = 0.99$ and $\rho = 1.4$, the non-stationary finite-horizon approximation can improve the upper bound by 0.23% and the policy performance by up to 3.52%; together, these improvements reduce the optimality gap from 3.93% to 0.19%. We also compare the performance of the affine finite-horizon approximation with the performance of a stronger stationary approximation structure, the separable piecewise linear approximation. It is known that the separable piecewise linear approximation can yield better bounds and potentially stronger heuristic policies than affine approximation; see, e.g., Meissner and Strauss (2012). It is unclear whether separable piecewise linear approximation is stronger than affine finite-horizon approximation. In Appendix A, we conduct a numerical study on randomly generated problem instances to compare the separable piecewise linear approximation and the affine finite-horizon approximation. Even though the separable piecewise linear approximation can produce tighter bounds for very small instances, we show that this advantage disappears for moderately-sized problems and that the affine finite-horizon approximation usually produces tighter bounds. Furthermore, solving the separable piecewise linear approximation can take hours for relatively small problem instances, while the affine finite-horizon approximation solves in a fraction of a second. Therefore, the affine finite-horizon approximation is a strong alternative to the separable piecewise linear approximation. Note that solving the ALPs from the separable piecewise linear approximation is challenging since no compact reformulations, like the one we established for the affine finite-horizon approximation, are known. Of course, one can also consider finite-horizon separable piecewise linear approximations. However, such an approximation is not likely to be tractable.

8. Concluding Remarks

Approximate linear programs have been widely used to approximately solve stochastic dynamic programs that suffer from the curse of dimensionality. Due to the canonical results establishing the optimality of stationary value functions and policies for infinite-horizon dynamic programs, existing research has predominantly focused on approximation architectures that are stationary over time. We consider finite-horizon approximations where the parameters are time-dependent within a pre-determined time horizon and are stationary afterwards. Such finite-horizon approximations are commonly used to solve infinite-horizon linear programming problems, but have not been considered in the ALP literature.

We apply this approach, together with an affine approximation architecture, to a rolling-horizon capacity allocation problem and obtain three main results. First, non-stationary approximations can substantially improve upper bounds on the optimal revenue. Second, these upper bounds are monotonically decreasing as the horizon length increases, and converge to the upper bound from the infinite-horizon approximation. Finally, the improvement does not come at the expense of tractability, as the resulting ALPs admit compact representations and can be solved efficiently. They also produce strong heuristic policies.

Although some of the techniques employed in this paper are specific to our setting, we believe that our approach to constructing finite-horizon approximations can be generalized to a broad class of infinite-horizon MDPs. The application of a non-stationary functional approximation to the linear programming formulation of an infinite-horizon MDP provides a general framework; hence, convergence results analogous to Proposition 1 may also hold in more general settings. Although the bounding techniques introduced in Section 5 are specific to our setting, their underlying ideas rely on generic concepts from linear programming aggregation that are applicable to a broad class of linear programming problems (Zipkin 1980, Shetty and Taylor 1987). In general, however, the derivation of equivalent reduced formulations similar to the formulation introduced in Section 4 depends on the structure of the underlying applications and may therefore not be attainable in general. Nevertheless, the recent work of Ke et al. (2021) shows that even when equivalence cannot be established, compact reformulations can be constructed to yield strong bounds and policies. Therefore, the framework in our paper might be applied even if the setting does not admit *equivalent* reduced formulations.

The rolling-horizon capacity allocation problem naturally admits a number of additional considerations, ranging from incorporating demand carryover and/or overtime capacity as discussed in the Patrick et al. (2008) to settings where multiple resources become available in each time period and/or situations where customers might cancel bookings. It would be interesting to consider such characteristics and evaluate their impact on the performance of the resulting finite-horizon approximations.

Relative to a stationary separable piecewise linear approximation, we observed that the finite-horizon approximations are not only more efficient but also more effective as they can achieve better bounds on the optimal revenue. It would be interesting to further investigate stronger approximation structures, both in terms of their use in conjunction with finite-horizon approximations and in their performance relative to finite-horizon approximations that rely on more basic approximation structures. Selecting appropriate basis functions for an approximation architecture can be a challenging task that is often problem specific. However, finite-horizon approximations may effectively provide a principled approach to adding such basis functions that can boost performance of weaker approximation architectures with little additional computational overhead. We believe it would be worthwhile to explore this trade-off in other settings as well.

References

Adelman, D. 2007. Dynamic bid-prices in revenue management. *Operations Research* **55**(4) 647–661.

Adelman, D., A. J. Mersereau. 2008. Relaxations of weakly coupled stochastic dynamic programs. *Operations Research* **56**(3) 712–727.

Bean, James C., Robert L. Smith. 1984. Conditions for the existence of planning horizons. *Mathematics of Operations Research* **9**(3) 391–401.

Bertsekas, D. P. 2001. *Dynamic Programming and Optimal Control*, vol. II. Athena Scientific.

Bertsimas, Dimitris, Velibor V Mišić. 2016. Decomposable markov decision processes: A fluid optimization approach. *Operations Research* **64**(6) 1537–1555.

Brown, David B, Martin B Haugh. 2017. Information relaxation bounds for infinite horizon markov decision processes. *Operations Research* **65**(5) 1355–1379.

Brualdi, Richard A. 1982. Notes on the Birkhoff algorithm for doubly stochastic matrices. *Canadian Mathematical Bulletin* **25**(2) 191–198.

Brumelle, S. L., J. I. McGill. 1993. Airline seat allocation with multiple nested fare classes. *Operations Research* **41**(1) 127–137.

Chand, Suresh, Vernon Ning Hsu, Suresh Sethi. 2002. Forecast, solution, and rolling horizons in operations management problems: a classified bibliography. *Manufacturing & Service Operations Management* **4**(1) 25–43.

de Farias, D. P., B. Van Roy. 2003. The linear programming approach to approximate dynamic programming. *Operations Research* **51**(6) 850–865.

de Farias, D.P., B. Van Roy. 2004. On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research* **29**(3) 462–478.

Desrosiers, Jacques, Marco E Lübbecke. 2005. A primer in column generation. *Column Generation*. Springer US, 1–32.

Diamant, Adam, Joseph Milner, Fayez Quereshy. 2018. Dynamic patient scheduling for multi-appointment health care programs. *Production and Operations Management* **27**(1) 58–79.

Erdelyi, Alexander, Huseyin Topaloglu. 2010. Approximate dynamic programming for dynamic capacity allocation with multiple priority levels. *IIE Transactions* **43**(2) 129–142.

Farias, Vivek F, Benjamin Van Roy. 2007. An approximate dynamic programming approach to network revenue management. *Preprint* .

Ghate, A., Robert L. Smith. 2013. A linear programming approach to nonstationary infinite-horizon markov decision processes. *Operations Research* **61**(2) 413–425.

Ghate, Archis. 2010. Infinite horizon problems. *Wiley Encyclopedia of Operations Research and Management Science* .

Grinold, Richard C. 1977. Finite horizon approximations of infinite horizon linear programs. *Mathematical Programming* **12**(1) 1–17.

Ke, Jiannan, Dan Zhang, Huan Zheng. 2021. Compact reformulations of approximate linear programs for finite-horizon markov decision processes. Working paper, University of Colorado Boulder.

Liu, Nan, Serhan Ziya, Vidyadhar G Kulkarni. 2010. Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing & Service Operations Management* **12**(2) 347–364.

Marcus, M, R Ree. 1959. Notes on the Birkhoff algorithm for doubly stochastic matrices. *Quarterly Journal of Mathematics* **10**(1) 296–302.

Mayne, David Q. 2014. Model predictive control: Recent developments and future promise. *Automatica* **50**(12) 2967–2986.

Meissner, J., A. K. Strauss. 2012. Network revenue management with inventory-sensitive bid prices and customer choice. *European Journal of Operational Research* **216**(2) 459–468.

Mendelssohn, Roy. 1980. Improved bounds for aggregated linear programs. *Operations Research* **28**(6) 1450–1453.

Ovacik, Irfan M, Reha Uzsoy. 1995. Rolling horizon procedures for dynamic parallel machine scheduling with sequence-dependent setup times. *International Journal of Production Research* **33**(11) 3173–3192.

Patrick, J., M. Puterman, M. Queyranne. 2008. Dynamic multi-priority patient scheduling for a diagnostic resource. *Operations Research* **56**(6) 1507–1525.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, New York.

Reindorp, Matthew, Michael Fu. 2011. Dynamic lead time promising. *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)t* 176–183.

Robinson, Lawrence W. 1995. Optimal and approximate control policies for airline booking with sequential nonmonotonic fare classes. *Operations Research* **43** 252–263.

Rohleder, Thomas R, Kenneth J Klassen. 2002. Rolling horizon appointment scheduling: a simulation study. *Health Care Management Science* **5**(3) 201–209.

Romeijn, H. Edwin, Robert Smith, James Bean. 1992. Duality in infinite deminensional linear programming. *Mathematical Programming* **53**(1–3) 79–97.

Schochetman, Irwin E., Robert L. Smith. 1992. Finite dimensional approximation in infinite dimensional mathematical programming. *Mathematical Programming* **54**(1-3) 307–333.

Schweitzer, P. J., A. Seidmann. 1985. Generalized polynomial approximations in Markovian decision processes. *Journal of Mathematical Analysis and Applications* **110**(2) 568–582.

Sethi, Suresh, Gerhard Sorger. 1991. A theory of rolling horizon decision making. *Annals of Operations Research* **29**(1) 387–415.

Shapiro, A. 2003. Monte Carlo sampling methods. A. Ruszczyński, A. Shapiro, eds., *Handbook of Stochastic Optimization*. Elsevier Science Publishers B.V., Amsterdam, Netherlands.

Shetty, C M, Richard W Taylor. 1987. Improved bounds for aggregated linear programs. *Computers and Operations Research* **14**(5) 385–393.

Sridharan, V, William L Berry, V Udayabhanu. 1987. Freezing the master production schedule under rolling planning horizons. *Management Science* **33**(9) 1137–1149.

Tong, Chaoxu, Huseyin Topaloglu. 2013. On the approximate linear programming approach for network revenue management problems. *INFORMS Journal on Computing* **26**(1) 121–134.

Vossen, T., D. Zhang. 2015. Reductions of approximate linear programs for network revenue management. *Operations Research* **63**(6) 1352–1371.

Wolsey, L.A. 1998. *Integer Programming*. Wiley, New York.

Ye, Fan, Helin Zhu, Enlu Zhou. 2018. Weakly coupled dynamic program: Information and lagrangian relaxations. *IEEE Transactions on Automatic Control* **63**(3) 698–713. doi:10.1109/TAC.2017.2731817.

Zhang, D., D. Adelman. 2009. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science* **43**(3) 381–394.

Zipkin, Paul H. 1980. Bounds for row-aggregation in linear programming. *Operations Research* **28**(4) 903–916.

Appendix A: Comparison with Separable Piecewise Linear Approximations

We compare the performance of the affine finite-horizon approximation with the widely used separable piecewise linear approximation, which is given by

$$\vartheta(\mathbf{b},\mathbf{d}) \approx \theta + \sum_{m\in\mathcal{M}^-}\sum_{j=1}^{C} V_{m,j}\mathbb{1}\{b_m \geq j\} + \sum_{n\in\mathcal{N}}\sum_{k=1}^{D}\mathbb{1}\{d_n \geq k\}W_{n,k}. \tag{25}$$

Note that (25) is a stationary approximation, as is typical in the literature for infinite horizon problems. Clearly, the separable piecewise linear approximation provides a more flexible approximation architecture than the (stationary) affine approximation. Our focus here is to compare the stationary separable piecewise linear approximation with the affine finite-horizon approximation. Our numerical study shows that the affine finite-horizon approximation significantly outperforms the separable piecewise linear approximation, in that it tends to produce stronger bounds (except for very small problem instances). Moreover, the separable piecewise linear approximation takes hours to solve to optimality whereas the affine finite-horizon approximation requires only a fraction of a second.

Substituting (25) into $\mathbf{P}^{\boldsymbol{\alpha}}$ yields

$$\begin{aligned}
&\mathbf{PLP}(\mathbf{b}_0,\mathbf{d}_0):\\
&\min_{\theta,\mathbf{V},\mathbf{W}} \theta + \sum_{m\in\mathcal{M}^-}\sum_{j=1}^{b_{0,m}} V_{m,j} + \sum_{n\in\mathcal{N}}\sum_{k=1}^{d_{0,n}} W_{n,k}\\
&\quad \text{s.t. } (1-\beta)\theta + \sum_{m\in\mathcal{M}^-}\left(\sum_{j=1}^{b_m} V_{m,j} - \beta\sum_{j=1}^{s_{m+1}(\mathbf{b},\mathbf{a})} V_{m,j}\right) + \sum_{n\in\mathcal{N}}\left(\sum_{k=1}^{d_n} W_{n,k} - \beta\sum_{k=1}^{D} p\{d'_n \geq k\}W_{n,k}\right) \geq r(\mathbf{a}),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X},
\end{aligned}$$

with $s_m(\mathbf{b},\mathbf{a}) = b_m + \sum_{n\in\mathcal{N}} a_{m,n}$. Its corresponding dual equals

$$\begin{aligned}
&\mathbf{PLD}(\mathbf{b}_0,\mathbf{d}_0):\\
&\max_{\pi} \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} r(\mathbf{a})\pi(\mathbf{b},\mathbf{d},\mathbf{a})\\
&\text{s.t.} \quad \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \pi(\mathbf{b},\mathbf{d},\mathbf{a}) = \frac{1}{1-\beta},\\
&\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \left(\mathbb{1}\{b_m \geq j\} - \beta\mathbb{1}\{s_{m+1}(\mathbf{b},\mathbf{a}) \geq j\}\right)\pi(\mathbf{b},\mathbf{d},\mathbf{a}) = \begin{cases}1, & \text{if } j \leq b_{0,m},\\ 0 & \text{o.w.,}\end{cases} \qquad \forall m\in\mathcal{M}^-, j\in\mathcal{C},\\
&\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \left(\mathbb{1}\{d_n \geq k\} - \beta p\{d'_n \geq k\}\right)\pi(\mathbf{b},\mathbf{d},\mathbf{a}) = \begin{cases}1, & \text{if } k \leq d_{0,n},\\ 0 & \text{o.w.,}\end{cases} \qquad \forall n\in\mathcal{N},\\
&\boldsymbol{\pi} \geq \mathbf{0}.
\end{aligned}$$

Column Generation

The dual problem $\mathbf{PLD}(\mathbf{b}_0, \mathbf{d}_0)$ has relatively few constraints, though the number of decision variables is exponential in both the number of customer classes and the length of the booking horizon. This enables a column generation procedure (Desrosiers and Lübbecke 2005), which uses only a small subset of the decision variables and adds more variables when needed. This approach has received considerable attention in the literature (Adelman 2007, Zhang and Adelman 2009, Meissner and Strauss 2012). The procedure maintains a *restricted master problem* ($\mathbf{RMP}$), which works with a small subset of the columns in the master problem with $\mathbf{PLD}(\mathbf{b}_0, \mathbf{d}_0)$. Let $(\mathbf{V}, \mathbf{W}, \theta)$ be the variables in the primal problem $\mathbf{PLP}(\mathbf{b}_0, \mathbf{d}_0)$ corresponding to in ($\mathbf{RMP}$). A column generation subproblem is used to determine the columns with the highest reduced cost. For the separable piecewise linear approximation, the column generation subproblem is given by

$$
\begin{aligned}
\varphi = \max_{\mathbf{b},\mathbf{d},\mathbf{a}} r(\mathbf{a}) - (1-\beta)\theta - \sum_{m\in\mathcal{M}^-}\sum_{j=1}^{C} V_{m,j}\left[\mathbb{1}\{b_m \geq j\} - \beta\mathbb{1}\{b_{m+1} + \sum_{n\in\mathcal{N}} a_{m+1,n} \geq j\}\right] \\
- \sum_{n\in\mathcal{N}}\sum_{k=1}^{D} W_{n,k}\left[\mathbb{1}\{d_n \geq k\} - \beta p\{d'_n \geq k\}\right] \\
\text{s.t.} \quad \sum_{m\in\mathcal{M}} a_{m,n} &\leq d_n, \quad \forall n \in \mathcal{N}, \\
\sum_{n\in\mathcal{N}} a_{m,n} + b_m &\leq C, \quad \forall m \in \mathcal{M}, \\
d_n &\leq D, \quad \forall n \in \mathcal{N}, \\
\mathbf{b} \in \mathbb{Z}_+^M, \mathbf{d} \in \mathbb{Z}_+^N, \mathbf{a} \in \mathbb{Z}_+^{M\times N}.
\end{aligned}
$$

If the reduced cost $\varphi \leq 0$, the solution to ($\mathbf{RMP}$) is optimal to $\mathbf{PLD}(\mathbf{b}_0, \mathbf{d}_0)$ as well. Otherwise, we can add columns with positive reduced costs to ($\mathbf{RMP}$) based on the solution to the column generation subproblems, and continue by re-optimizing ($\mathbf{RMP}$). Observe that the objective function of the column generation subproblem is nonlinear; auxiliary binary variables are needed to reformulate the problem as a mixed integer linear program.

PROPOSITION 6. *The column generation subproblem is equivalent to the following mixed integer linear program:*

$$
\begin{aligned}
\varphi = \max_{\mathbf{b},\mathbf{d},\mathbf{a},\mathbf{y},\mathbf{s},\mathbf{q}} \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}} v_{m,n}a_{m,n} - (1-\beta)\theta - \sum_{m\in\mathcal{M}^-}\sum_{j=1}^{C} V_{m,j}\left[y_{m,j} - \beta s_{m,j}\right] \\
- \sum_{n\in\mathcal{N}}\sum_{k=1}^{D} W_{n,k}\left[q_{n,k} - \beta p\{d'_n \geq k\}\right] \\
\text{s.t.} \quad \sum_{m\in\mathcal{M}} a_{m,n} &\leq d_n, & \forall n \in \mathcal{N}, \\
\sum_{n\in\mathcal{N}} a_{m,n} + b_m &\leq C, & \forall m \in \mathcal{M},
\end{aligned}
$$

$$
\begin{aligned}
b_m &= \sum_{j=1}^{C} y_{m,j}, && \forall m \in \mathcal{M}, \\
b_{m+1} + \sum_{n \in \mathcal{N}} a_{m+1,n} &= \sum_{j=1}^{C} s_{m,j}, && \forall m \in \mathcal{M}^-, \\
d_n &= \sum_{k=1}^{D} q_{n,k}, && \forall n \in \mathcal{N}, \\
y_{m,j+1} &\le y_{m,j}, && \forall m \in \mathcal{M}, j = 1, ..., C-1, , \\
s_{m,j+1} &\le s_{m,j}, && \forall m \in \mathcal{M}^-, j = 1, ..., C-1, \\
q_{n,k+1} &\le q_{n,k}, && \forall n \in \mathcal{N}, k = 1, ..., D-1, \\
&\mathbf{b} \in \mathbb{Z}_+^M, \mathbf{d} \in \mathbb{Z}_+^N, \mathbf{a} \in \mathbb{Z}_+^{M \times N}, \\
&\mathbf{y} \in \{0,1\}^{M \times C}, \mathbf{s} \in \{0,1\}^{M-1 \times C}, \mathbf{q} \in \{0,1\}^{N \times D}.
\end{aligned}
$$

■

In the formulation above, the objective function is linearized by introducing binary variables $\mathbf{y}, \mathbf{s}$, and $\mathbf{q}$. By imposing the requirement that $y_{m,j}$ is decreasing in j, we have a one-to-one correspondence between the vector $\mathbf{y}_m$ and b_m. We impose similar monotonicity conditions on $\mathbf{s}$ and $\mathbf{q}$ to ensure a one-to-one mapping with the corresponding variables. It is important to note that while the column generation subproblem for the affine approximation can be formulated as a linear programming problem (which in turn allows us derive a compact formulation for the overall problem), we were unable to establish a similar result for the separable piecewise linear approximation

We also note that the objective function value $\tilde{z}$ of an intermediate solution to (**RMP**) provides a lower bound on the optimal objective function value of $\mathbf{PLD}(\mathbf{b}_0, \mathbf{d}_0)$, which we denote by $\tilde{z}^*(\mathbf{b}_0, \mathbf{d}_0)$. Observe also that $\tilde{z}^*(\mathbf{b}_0, \mathbf{d}_0)$ is itself an upper bound of the total expected reward of the original MDP. However, it is well-known that $\tilde{z} + \frac{1}{1-\beta}\varphi$ provides an upper bound on $\tilde{z}^*(\mathbf{b}_0, \mathbf{d}_0)$. As a result, we obtain an upper bound of the original MDP after every iteration of the column generation procedure.

Numerical Results

We compare the performance of the separable piecewise linear approximation and the affine finite-horizon approximation on randomly generated instances. We consider two sets of problem instances. These instances are relatively small. Due to the slow convergence of the column generation procedures, it is intractable to test larger instances. In fact, the smaller instances we tested already take hours for the column generation procedures to converge. On the other hand, the affine finite-horizon approximation requires less than a second for all instances. For the first set of small test instances,

Table 1 Bounds for problem instances with $M = 3$, $N = 2$, and $C = 3$. The last column reports the number of instances (out of 100) for which the SPL bound is tighter than the FH50 bound

β	ρ	FH-LB	FH1		FH50		SPL		#
			value	gap (%)	value	gap (%)	value	gap (%)	
0.95	0.6	245.7	260.8	6.1	257.9	5.0	257.7	4.9	72
0.95	1.0	346.3	423.9	22.4	421.9	21.8	412.5	19.1	89
0.95	1.5	409.8	451.4	10.1	448.1	9.3	445.4	8.7	75
0.95	3.0	461.2	472.9	2.5	471.8	2.3	469.0	1.7	85
0.99	0.6	1246.0	1303.1	4.6	1300.2	4.3	1287.8	3.4	93
0.99	1.0	1745.8	2161.1	23.8	2158.8	23.6	2090.4	19.7	99
0.99	1.5	2071.4	2274.7	9.8	2271.4	9.6	2242.3	8.2	92
0.99	3.0	2321.7	2380.5	2.5	2379.4	2.5	2358.6	1.6	94

Table 2 Bounds for problem instances with $M = 7$, $N = 3$, and $C = 10$. The last column reports the number of instances (out of 10) for which the SPL bound is tighter than the FH50 bound

β	ρ	FH-LB	FH1		FH50		SPL		#
			value	gap (%)	value	gap (%)	value	gap (%)	
0.95	0.6	1027.0	1033.2	0.6	1022.3	-0.4	1033.0	0.6	2
0.95	1.0	1552.2	1668.1	7.5	1660.5	7.0	1666.6	7.4	0
0.95	1.5	1774.2	1840.5	3.7	1815.7	2.3	1837.0	3.5	0
0.95	3.0	1923.7	1957.3	1.7	1944.2	1.1	1946.2	1.2	5
0.99	0.6	5109.7	5105.9	-0.1	5094.6	-0.3	5105.7	-0.1	2
0.99	1.0	7827.3	8454.4	8.0	8443.2	7.9	8442.6	7.9	2
0.99	1.5	8751.1	8995.8	2.8	8974.0	2.5	8975.3	2.6	3
0.99	3.0	9349.1	9451.7	1.1	9440.1	1.0	9427.0	0.8	5

the separable piecewise linear approximation produces better bounds than the affine ALP. However, for a second set of somewhat larger instances, the affine finite-horizon approximation produces better bounds for the majority of problem instances. Thus, it appears that the affine finite-horizon approximation significantly outperforms the separable piecewise linear approximation overall.

The first set of small problem instances have a capacity of 3 slots per day. We use a booking horizon $M = 3$, and there are two customer classes ($N = 2$). Daily demand rates and reward parameters are randomly generated. We vary the discount factor β and system load ρ. For each parameter configuration, 100 instances are randomly generated. We calculate the stationary affine approximation bound (FH1), the affine finite-horizon approximation upper bound with $T = 50$ (FH50), and the separable piecewise linear approximation bound (SPL) for each instance. We also simulate the heuristic policy derived from affine finite-horizon approximation policy, which gives us a lower bound (FH-LB). We report the gaps between the upper bounds and FH-LB as a conservative estimate on the optimality gaps. Each column generation subproblem requires approximately 0.5 seconds, while the finite-horizon approximation solved within 0.01 seconds. We report the average value across 100 instances for each configuration in Table 1.

In Table 1, the separable piecewise linear approximation results in tighter bounds than the affine ALP (FH1), which is to be expected. Moreover, for the majority of instances the separable piecewise linear approximation bound is better than the affine finite-horizon approximation bound. Due to the small size of the instances, the column generation subproblems that arise solving the separable piecewise linear approximations are reasonably fast, even though the affine finite-horizon approximation is significantly faster.

We next consider slightly larger instances with a booking horizon $M = 7$, $N = 3$ customer classes, and a daily capacity $C = 10$. We again vary the discount factor and load factor, and conduct numerical experiments on randomly generated instances. For each discount factor and load factor combination, 10 instances are generated. We report the average value across the 10 instances in Table 2. In Table 2, the quality of affine finite-horizon approximation bound is better than that of the separable piecewise linear approximation bound for the majority of instances. However, the affine finite-horizon approximation can be solved within 0.05 second for each test instance, whereas the column generation procedure that is use to solve the separable piecewise linear approximations takes 7.8 hours to converge on average. Note that the gaps reported are negative in a few cases. This is due to the fact that the lower bounds are obtained through simulation and therefore subject to statistical error.

Appendix B: Technical Proofs

Proof of Proposition 2

The proof proceeds in two steps. First, we show that $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ is a relaxation of $\mathbf{AFD}_T(\mathbf{b}_0, \mathbf{d}_0)$, in that for any solution to $\mathbf{AFD}_T(\mathbf{b}_0, \mathbf{d}_0)$ we can construct a solution to $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ with the same objective value by variable aggregation. This implies that $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ produces a lower bound for $\mathbf{AFD}_T(\mathbf{b}_0, \mathbf{d}_0)$. Second, we establishing the reverse direction by showing that $\mathbf{AFD}_T(\mathbf{b}_0, \mathbf{d}_0)$ can be interpreted as a Dantzig-Wolfe reformulation of $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$.

Step 1: Consider a feasible solution $\{\boldsymbol{\pi}_t\}_{\forall t \in \mathcal{T}}$ to $\mathbf{AFD}_T(\mathbf{b}_0, \mathbf{d}_0)$ and define $\hat{b}_{t,m} = \sum_{(\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}} b_m \pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a})$ and $\hat{a}_{t,m,n} = \sum_{(\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}} a_{m,n} \pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a})$ for all $m \in \mathcal{M}^-$, $n \in \mathcal{N}$, and $t \in \mathcal{T}$. By construction, the solution $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$ satisfies constraint (10) in $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ and have the same objective function value. We also have

$$\sum_{n \in \mathcal{N}} \hat{a}_{t,m,n} + \hat{b}_{t,m} = \sum_{(\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}} \pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a}) \left(\sum_{n \in \mathcal{N}} a_{m,n} + b_m \right) \leq \sum_{(\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}} \pi_t(\mathbf{b}, \mathbf{d}, \mathbf{a}) C = \frac{C}{1 - \mathbb{1}_{\{t=T\}} \beta},$$

where the inequality follows from the definition of $\mathcal{A}(\mathbf{b},\mathbf{d})$; thus, the solution also satisfies constraint (11). Similarly, we have

$$\sum_{m\in\mathcal{M}} \hat{a}_{t,m,n} = \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \pi_t(\mathbf{b},\mathbf{d},\mathbf{a}) \left(\sum_{m\in\mathcal{M}} a_{m,n}\right) \le \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}} \pi_t(\mathbf{b},\mathbf{d},\mathbf{a}) d_n,$$

where the inequality again follows from the definition of $\mathcal{A}(\mathbf{b},\mathbf{d})$. Together with constraint (9), this shows that the solution satisfies constraint (12).

Step 2: To show the reverse direction, we slightly modify the formulation $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ by introducing decision variables $\hat{d}_{t,n}$ with $0 \le \hat{d}_{t,n} \le D$ for all $t\in\mathcal{T}$ and $n\in\mathcal{N}$, and splitting constraint (12) into

$$\sum_{m\in\mathcal{M}} \hat{a}_{t,m,n} \le \hat{d}_{t,n}, \qquad \forall n\in\mathcal{N}, t\in\mathcal{T}, \tag{26}$$

$$\hat{d}_{t,n} = \begin{cases} d_{0,n} + \mathbb{1}_{\{t=T\}} \frac{\beta \bar{d}_n}{1-\beta}, & \text{if } t=0, \\ \frac{\bar{d}_n}{1-\mathbb{1}_{\{t=T\}}\beta}, & \text{if } t>0, \end{cases} \qquad \forall n\in\mathcal{N}, t\in\mathcal{T}. \tag{27}$$

We also define the bounded polyhedron

$$\Omega = \left\{ \begin{array}{ll} (\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathbb{R}_+^M \times \mathbb{R}_+^N \times \mathbb{R}_+^{M\times N} : & \sum_{m\in\mathcal{M}} a_{m,n} \le d_n, \forall n\in\mathcal{N}, \\ & \sum_{n\in\mathcal{N}} a_{m,n} + b_m \le C, \forall m\in\mathcal{M}, \\ & d_n \le D, \forall n\in\mathcal{N} \end{array} \right\}.$$

Observe that Ω has integer extreme points, as its constraint matrix is totally unimodular (TU). To see that the constraint matrix is TU, recall a well-known sufficient condition for TU: a matrix A is TU if (i) $a_{ij}\in\{1,-1,0\}$ for all i,j; (ii) each column contains at most 2 nonzero coefficients; (iii) there exists a partition (M_1,M_2) of the set M of rows such that each column j containing two nonzero coefficients satisfies $\sum_{i\in M_1} a_{ij} - \sum_{i\in M2} a_{ij} = 0$, see Proposition 3.2 in (Wolsey 1998). We can verify that these conditions are satisfied by construction a partition where the constraints where the first and third constraint are in the first set and the second constraint is in the other set.

Let $\mathcal{X}^\Omega$ be the set of its extreme points. Note that $\mathcal{X}^\Omega \subseteq \mathcal{X}$ by construction.

Now, consider a feasible solution $\{(\mathbf{b}_t,\mathbf{d}_t,\mathbf{a}_t)\}_{t\in\mathcal{T}}$ to the modified reduced formulation $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$. For each $t\in\mathcal{T}^-$, we can express this solution as a convex combination of the extreme points in Ω, that is,

$$(\mathbf{b}_t,\mathbf{d}_t,\mathbf{a}_t) = \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}^\Omega} (\mathbf{b},\mathbf{d},\mathbf{a}) \lambda_t(\mathbf{b},\mathbf{d},\mathbf{a}),$$

such that $\lambda_t(\mathbf{b},\mathbf{d},\mathbf{a}) \geq 0$ for all $(\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}^\Omega$ and $\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}^\Omega} \lambda_t(\mathbf{b},\mathbf{d},\mathbf{a}) = 1$. When $t = T$, we scale the solution by a factor $(1-\beta)$ to ensure it is contained in Ω and obtain

$$(1-\beta)(\mathbf{b}_T,\mathbf{d}_T,\mathbf{a}_T) = \sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}^\Omega} (\mathbf{b},\mathbf{d},\mathbf{a})\lambda_T(\mathbf{b},\mathbf{d},\mathbf{a}),$$

with $\lambda_T(\mathbf{b},\mathbf{d},\mathbf{a}) \geq 0$ for all $(\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}^\Omega$ and $\sum_{(\mathbf{b},\mathbf{d},\mathbf{a})\in\mathcal{X}^\Omega} \lambda_T(\mathbf{b},\mathbf{d},\mathbf{a}) = 1$. Now, we can construct a solution to $\mathbf{AFD}_T(\mathbf{b}_0,\mathbf{d}_0)$ by defining

$$\pi_t(\mathbf{b},\mathbf{d},\mathbf{a}) \; = \begin{cases} \frac{\lambda_t(\mathbf{b},\mathbf{d},\mathbf{a})}{1-\mathbb{1}_{\{t=T\}}\beta}, & \text{if } (\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}^\Omega, \\ 0, & \text{otherwise}, \end{cases} \qquad \forall (\mathbf{b},\mathbf{d},\mathbf{a}) \in \mathcal{X}, t \in \mathcal{T}.$$

This solution satisfies the constraints in $\mathbf{AFD}_T(\mathbf{b}_0,\mathbf{d}_0)$ and has the same objective value as the corresponding solution to $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$.

Combining both steps completes the proof. ■

Proof of Proposition 3

We first establish the monotonicity of $\underline{z}_T(\mathbf{b}_0,\mathbf{d}_0)$ in Part (i).

Suppose $T > 0$, and let $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*, \hat{\mathbf{q}}^*)$ be an optimal solution to $\underline{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$. Given this solution, we construct a solution $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{q}})$ to $\underline{\mathbf{AFD}}_{T+1}(\mathbf{b}_0,\mathbf{d}_0)$ by defining

$$\tilde{a}_{t,m,n} = \begin{cases} \hat{a}^*_{t,m,n}, & \text{if } t < T, \\ (1-\beta)\hat{a}^*_{T,m,n}, & \text{if } t = T, \\ \hat{a}^*_{T,m,n}, & \text{if } t = T+1, \end{cases} \qquad \forall m \in \mathcal{M}, n \in \mathcal{N}, t \in \mathcal{T} \cup \{T+1\},$$

$$\tilde{b}_{t,m} = \begin{cases} \hat{b}^*_{t,m}, & \text{if } t < T, \\ (1-\beta)\hat{b}^*_{T,m}, & \text{if } t = T, \\ \hat{b}^*_{T,m}, & \text{if } t = T+1, \end{cases} \qquad \forall m \in \mathcal{M}^-, \in \mathcal{T} \cup \{T+1\},$$

$$\tilde{q}_m = \hat{q}^*_m, \qquad \forall m \in \mathcal{M}^-.$$

Observe that the solution value obtained with this construction equals $\underline{z}_T(\mathbf{b}_0,\mathbf{d}_0)$. Since $\underline{\mathbf{AFD}}_{T+1}(\mathbf{b}_0,\mathbf{d}_0)$ is a maximization problem, we only need to show that the solution $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{q}})$ satisfies the constraints in $\underline{\mathbf{AFD}}_{T+1}(\mathbf{b}_0,\mathbf{d}_0)$. Constraints (19), (20), and (21) are satisfied by construction, as is constraint (18) when $t \leq T$.

To show that constraint (18) is satisfied when $t = T+1$, consider any $m \in \mathcal{M}^-$. We have

$$(1-\beta)\tilde{b}_{T+1,m} - \tilde{b}_{T,m+1} - \sum_{n\in\mathcal{N}} \tilde{b}_{T,m+1,n} = (1-\beta)\left(\hat{b}^*_{T,m} - \hat{b}^*_{T,m+1} - \sum_{n\in\mathcal{N}} \hat{b}^*_{T,m+1,n}\right) = 0,$$

where the first equality follows by the definition of $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{q}})$ and the last equality follows because constraint (20) holds for any solution to $\underline{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$.

To establish Part (ii), let $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*, \hat{\mathbf{q}}^*)$ be an optimal solution to $\underline{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ when $T \geq M-1$, and suppose that $\hat{q}^*_{\bar{m}} > 0$ for some $\bar{m} \in \mathcal{M}^-$. We proceed to construct a solution to $\underline{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$ by modifying the values of $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*, \hat{\mathbf{q}}^*)$ such that $\hat{q}^*_{\bar{m}}$ will equal 0 while retaining feasibility. With $\rho_{\bar{m}}$ as defined in (22), the objective value of the modified solution will be at least as large as the objective value of $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*, \hat{\mathbf{q}}^*)$, establishing the desired result.

As a first step, we note that constraint (18) implies that any decrease in $\hat{q}^*_{\bar{m}}$ from a positive value to 0 leads to an increase in $\hat{b}^*_{0,\bar{m}}$ and potentially other $\hat{b}^*_{t',m'}$ such that $t'+m'=\bar{m}$. Specifically, according to constraint (18) we have

$$\begin{aligned}
\hat{b}^*_{0,\bar{m}} &= b_{0,\bar{m}} - \hat{q}^*_{\bar{m}}, \\
\hat{b}^*_{1,\bar{m}-1} &= \hat{b}^*_{0,\bar{m}} + \sum_{n\in\mathcal{N}} \hat{a}^*_{0,\bar{m},n} = \sum_{n\in\mathcal{N}} \hat{a}^*_{0,\bar{m},n} + b_{0,\bar{m}} - \hat{q}^*_{\bar{m}}, \\
\cdots \quad & \cdots \quad \cdots \\
\hat{b}^*_{\bar{m}-1,1} &= \sum_{t'=0}^{\bar{m}-2} \sum_{n\in\mathcal{N}} \hat{a}^*_{t',\bar{m}-t',n} + b_{0,\bar{m}} - \hat{q}^*_{\bar{m}}.
\end{aligned}$$

Plugging the above into (20), we then observe that all the following constraints must hold:

$$\begin{aligned}
\sum_{n\in\mathcal{N}} \hat{a}^*_{0,\bar{m},n} &\leq C - b_{0,\bar{m}} + \hat{q}^*_{\bar{m}}, \\
\sum_{t'=0}^{1} \sum_{n\in\mathcal{N}} \hat{a}^*_{t',\bar{m}-t',n} &\leq C - b_{0,\bar{m}} + \hat{q}^*_{\bar{m}}, \\
\cdots \quad & \cdots \quad \cdots \\
\sum_{t'=0}^{\bar{m}-1} \sum_{n\in\mathcal{N}} \hat{a}^*_{t',\bar{m}-t',n} &\leq C - b_{0,\bar{m}} + \hat{q}^*_{\bar{m}}.
\end{aligned} \tag{28}$$

To ensure that these constraints remain feasible when reducing $\hat{q}^*_{\bar{m}}$ to 0, some values $\hat{a}^*_{t',m',n}$ where $t'+m'=\bar{m}$ will also need to be decreased. Due to the non-negativity of $\hat{\mathbf{a}}^*$, it is sufficient to only consider the last inequality (28). In particular, if (28) becomes infeasible when decreasing $\hat{q}^*_{\bar{m}}$ to 0, feasibility can be restored by reducing $\hat{a}^*_{t',m',n}$ for some t' and m' such that $t'+m'=\bar{m}$. However, equation (22) defines $\rho_{\bar{m}}$ as the maximum of the objective function coefficients of the corresponding $\hat{a}^*_{t',m',n}$ variables, that is,

$$\rho_{\bar{m}} = \max_{\substack{t'\in\{1,\ldots,\bar{m}\},\\ n\in\mathcal{N}}} \beta^{\bar{m}-t'} v_{t',n}, \qquad \forall \bar{m} \in \mathcal{M}^-.$$

Therefore, we observe that such modifications do not lower the objective function value, which completes the proof. ■

Proof of Proposition 4

The proof proceeds in two steps. First, we bound the difference in the objective values between $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ and $\underline{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ as a function of optimal solutions to both $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ and the dual of $\underline{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$. In the second step, we bound the optimal values of the dual variables. We establish these results in the following two lemmas.

LEMMA 1. *Suppose $T>0$. Let $(\mathbf{a}^*,\mathbf{b}^*)$ be an optimal solution for the reduced formulation $\widehat{\boldsymbol{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$, and let (U^*,V^*,W^*,Z^*) be an optimal solution to the dual of $\underline{\boldsymbol{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$. Then,*

$$\begin{aligned}
&\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)-\underline{z}_T(\mathbf{b}_0,\mathbf{d}_0)\\
&\le \sum_{m\in\mathcal{M}^-}\left((1-\beta)V^*_{T,m}+\beta Z^*_m-\beta Z^*_{m-1}\right)b^*_{T,m}\\
&\quad-\sum_{m\in\mathcal{M}^-}V^*_{T,m}\left(\hat{b}^*_{T-1,m+1}+\sum_{n\in\mathcal{N}}\hat{a}^*_{T-1,m+1,n}\right)-\sum_{m\in\mathcal{M}^-}\beta Z^*_m\left(\sum_{n\in\mathcal{N}}\hat{a}^*_{T,m+1,n}\right),
\end{aligned}\tag{29}$$

where we assume $Z^*_0=0$.

<u>Proof of Lemma 1:</u> To establish a bound on the gap between the reduced formulation $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ and the auxiliary program $\underline{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$, we start by taking the Lagrangian relaxation of $\underline{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ with respect to constraint (18) with $t=T$ and constraint (19) to obtain $L_T(\mathbf{b}_0,\mathbf{d}_0,V_T,Z)=$

$$\begin{aligned}
\max_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\mathbf{q}}}\quad &\sum_{\substack{t\in\mathcal{T},\\ m\in\mathcal{M},n\in\mathcal{N}}}\beta^t v_{m,n}\hat{a}_{t,m,n}-\sum_{m\in\mathcal{M}^-}\rho_m\hat{q}_m\\
&-\sum_{m\in\mathcal{M}^-}V_{T,m}\left((1-\beta)\hat{b}_{T,m}-\hat{b}_{T-1,m+1}-\sum_{n\in\mathcal{N}}\hat{a}_{T-1,m+1,n}\right)\\
&-\sum_{m\in\mathcal{M}^-}\beta Z_m\left(\hat{b}_{T,m}-\hat{b}_{T,m+1}-\sum_{n\in\mathcal{N}}\hat{a}_{T,m+1,n}\right)\\
\text{s.t. }&(18),(20),(21),\\
&\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\mathbf{q}}\ \ge\ \mathbf{0}.
\end{aligned}$$

Rearranging terms and accounting for the fact that $b_{T,M}=0$, we obtain $L_T(\mathbf{b}_0,\mathbf{d}_0,V_T,Z)=$

$$\begin{aligned}
\max_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\mathbf{q}}}\quad &\sum_{\substack{t\in\mathcal{T},\\ m\in\mathcal{M},n\in\mathcal{N}}}\beta^t v_{m,n}\hat{a}_{t,m,n}-\sum_{m\in\mathcal{M}^-}\rho_m\hat{q}_m-\sum_{m\in\mathcal{M}^-}\left((1-\beta)V_{T,m}+\beta Z_m-\beta Z_{m-1}\right)\hat{b}_{T,m}\\
&+\sum_{m\in\mathcal{M}^-}V_{T,m}\left(\hat{b}_{T-1,m+1}+\sum_{n\in\mathcal{N}}\hat{a}_{T-1,m+1,n}\right)+\sum_{m\in\mathcal{M}^-}\beta Z_m\left(\sum_{n\in\mathcal{N}}\hat{a}_{T,m+1,n}\right)\\
\text{s.t. }&(18),(20),(21),
\end{aligned}$$

$$\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{q}} \;\geq\; \mathbf{0}.$$

Now, let (V_T^*, Z^*) be optimal Lagrangian multipliers and let $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*)$ be an optimal solution for the reduced problem $\widehat{\mathbf{AFD}}_T(\mathbf{b}_0, \mathbf{d}_0)$. Because $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*)$ satisfies constraints (18) (when $t<T$), (20) and (21) with $\hat{q}_m=0$, we have

$$\begin{aligned}
&\sum_{\substack{t\in\mathcal{T},\\ m\in\mathcal{M}, n\in\mathcal{N}}} \beta^t v_{m,n}\hat{a}^*_{t,m,n} - \sum_{m\in\mathcal{M}^-}\left((1-\beta)V^*_{T,m}+\beta Z^*_m-\beta Z^*_{m-1}\right)\hat{b}^*_{T,m}\\
&+\sum_{m\in\mathcal{M}^-} V^*_{T,m}\left(\hat{b}^*_{T-1,m+1}+\sum_{n\in\mathcal{N}}\hat{a}^*_{T-1,m+1,n}\right)+\sum_{m\in\mathcal{M}^-}\beta Z^*_m\left(\sum_{n\in\mathcal{N}}\hat{a}^*_{T,m+1,n}\right)\\
&\leq L_T(\mathbf{b}_0,\mathbf{d}_0,V_T^*,Z^*).
\end{aligned}$$

Since $\underline{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ is a linear program, it follows that $\underline{z}_T(\mathbf{b}_0,\mathbf{d}_0) = L_T(\mathbf{b}_0,\mathbf{d}_0,V_T^*,Z^*)$. Using $\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)=\sum_{t\in\mathcal{T},m\in\mathcal{M},n\in\mathcal{N}}\beta^t v_{m,n}a^*_{t,m,n}$ and taking the difference between $\tilde{z}_T(\mathbf{b}_0,\mathbf{d}_0)$ and $L_T(\mathbf{b}_0,\mathbf{d}_0,V_T^*,Z^*)$ leads to (29). ■

Lemma 2. *Suppose $T>0$. Then, there exists an optimal solution (U^*,V^*,W^*,Z^*) to the dual of* $\underline{\boldsymbol{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ *such that for all $m\in\mathcal{M}^-$*

(i) $(1-\beta)V^*_{T,m}+\beta Z^*_m-\beta Z^*_{m-1}\leq 0$,

(ii) $-V^*_{T,m}\leq\epsilon_{T,m}$, *and*

(iii) $-\beta Z^*_m\leq\gamma_{T,m}$,

where

$$\epsilon_{T,m}=\begin{cases}\max\left(\rho_{T+m},\max\limits_{\substack{k\in1,\ldots,\min(T,M-m),\\ n\in\mathcal{N}}}\beta^{T-k}v_{m+k,n}\right), & \textit{if } T+m\leq M-1;\\ \max\limits_{\substack{k\in1,\ldots,\min(T,M-m),\\ n\in\mathcal{N}}}\beta^{T-k}v_{m+k,n}, & \textit{o.w.},\end{cases}\qquad \forall m\in\mathcal{M}^-,$$

and $\gamma_{T,m}$ is defined recursively as

$$\gamma_{T,m}=\begin{cases}\max\limits_{n\in\mathcal{N}}\beta^T v_{M,n}, & \textit{if } m=M-1;\\ \max\left(\max\limits_{n\in\mathcal{N}}\beta^T v_{m+1,n},(1-\beta)\epsilon_{T,m+1}+\gamma_{T,m+1}\right), & \textit{o.w.},\end{cases}\qquad \forall m\in\mathcal{M}^-.$$

Proof of Lemma 2: As a first step, we state the dual of $\underline{\mathbf{AFD}}_T(\mathbf{b}_0,\mathbf{d}_0)$ as follows:

$$\min_{\mathbf{U},\mathbf{V},\mathbf{W},\mathbf{Z}}\ \sum_{m\in\mathcal{M}^-}b_{0,m}V_{0,m}+\sum_{t\in\mathcal{T},m\in\mathcal{M}}\frac{C}{1-\mathbb{1}_{\{t=T\}}\beta}U_{t,m}+\sum_{n\in\mathcal{N}}d_{0,n}W_{0,n}+\sum_{\substack{t\in\mathcal{T},n\in\mathcal{N}\\ t>0}}\frac{\bar{d}_n}{1-\mathbb{1}_{\{t=T\}}\beta}W_{t,n}$$

$$\begin{aligned}
\text{s.t.}\qquad && U_{t,m}-V_{t+1,m-1}+V_{t,m} &\geq 0, && \forall t\in\mathcal{T}^-, m\in\mathcal{M}^-, && (30)\\
&& U_{t,m}-V_{t+1,m-1}+W_{t,n} &\geq \beta^t v_{m,n}, && \forall t\in\mathcal{T}^-, m\in\mathcal{M}, n\in\mathcal{N}, && (31)\\
&& U_{T,m}+(1-\beta)V_{T,m}+\beta Z_m-\beta Z_{m-1} &\geq 0, && \forall m\in\mathcal{M}^-, && (32)
\end{aligned}$$

$$U_{T,m} - \beta Z_{m-1} + W_{T,n} \ \geq \beta^T v_{m,n}, \qquad \forall m \in \mathcal{M}, n \in \mathcal{N}, \qquad (33)$$

$$V_{0,m} \ \geq \ -\rho_m, \qquad \forall m \in \mathcal{M}, n \in \mathcal{N}, \qquad (34)$$

$$U_{t,m}, W_{t,n} \ \geq 0, \qquad \forall t \in \mathcal{T}, m \in \mathcal{M}, n \in \mathcal{N}.$$

Here, we assume $Z_{t,0} = 0$ and $V_{t,0} = 0$ for all $t \in \mathcal{T}$.

Consider an optimal solution (U^*, V^*, W^*, Z^*) to this dual formulation. To establish (i), suppose that for some $m \in \mathcal{M}^-$ we have

$$(1-\beta)V^*_{T,m} + \beta Z^*_m - \beta Z^*_{m-1} = \delta > 0.$$

Because $U_{T,m} \geq 0$, this implies that constraint (32) is non-binding for m. Therefore, constraint (32) also holds if $Z^*_{m'}$ is decreased by δ for all $m' \in \{m, \ldots, M-1\}$. Observe that such decreases do not impact the feasibility of constraint (33) and do not change the objective value.

To establish (ii), we first extend the definition of ϵ and let

$$\epsilon_{t,m} = \begin{cases} \rho_m, & \text{if } t = 0, \\ \max\limits_{n \in \mathcal{N}} \beta^{t-1} v_{M,n}, & \text{if } t > 0 \text{ and } m = M-1, \\ \max(\epsilon_{t-1,m+1}, \max\limits_{n \in \mathcal{N}} \beta^{t-1} v_{m+1,n}), & \text{o.w.,} \end{cases} \qquad \forall t \in \mathcal{T}, m \in \mathcal{M}^-.$$

Next, we show that $-V^*_{t,m} \leq \epsilon_{t,m}$ for all $t \in \mathcal{T}$ and $m \in \mathcal{M}^-$. Given constraint (34), this clearly holds if $t = 0$. Now, consider some $t > 0$ and suppose that for some $\delta > 0$

$$-V^*_{t,M-1} = \max_{n \in \mathcal{N}} \beta^{t-1} v_{M,n} + \delta.$$

Because $U_{t-1,M}$ and $W_{t,n}$ are non-negative for all $n \in \mathcal{N}$, constraint (31) is non-binding for t and all $n \in \mathcal{N}$ when $m = M$. Therefore, the constraint continues to hold if we increase $V^*_{t,M-1}$ by δ. Note that this does not impact the feasibility of constraint (32), and that $V^*_{t,M-1}$ does not occur in constraint (30). Since $t > 0$, this does not impact the objective value.

Similarly, consider some $t > 0$ and $m < M-1$ and suppose that

$$-V^*_{t,m} > \epsilon_{t-1,m+1} \text{ and } -V^*_{t,m} > \max_{n \in \mathcal{N}} \beta^{t-1} v_{m+1,n}.$$

By induction, we have $-V^*_{t-1,m+1} \leq \epsilon_{t-1,m+1}$ and therefore constraints (30) and (31) are non-binding for $t-1$ and $m+1$. Again therefore, we can increase $V^*_{t,m}$ until $-V^*_{t,m} = \epsilon_{t,m}$ without impacting the feasibility of constraints (30)–(32).

Finally, we establish the last condition by induction. First, suppose that

$$-\beta Z^*_{M-1} > \max_{n \in \mathcal{N}} \beta^T v_{M,n} = \gamma_{T,M-1}.$$

Then, constraint (33) is non-binding and Z^*_{M-1} can be increased until its value equals $\frac{\gamma_{T,M-1}}{\beta}$. Note that this does not impact the feasibility of constraint (32), because Z^*_{M-1} only occurs in the left hand side with a positive coefficient. Next, consider some $m < M-1$ given that we have established that $-\beta Z^*_{m+1} \leq \gamma_{T,m+1}$ and $-V^*_{T,m+1} \leq \epsilon_{T,m+1}$. Now, suppose that

$$-\beta Z^*_m > \max_{n\in\mathcal{N}} \beta^T v_{m+1,n} \text{ and } -\beta Z^*_m > (1-\beta)\epsilon_{T,m+1} + \gamma_{T,m+1}.$$

Then, constraints (32) and (33) are non-binding for $m+1$, and we can increase Z^*_m until $-\beta Z^*_m = \gamma_{T,m}$ without impacting feasibility. ■

Proposition 4 immediately follows from these two lemmas, by substituting the bounds obtained in Lemma 2 for the relevant terms in Lemma 1.

Proof of Proposition 5

Overall, the proof of Proposition 5 proceeds as follows. Given a fractional optimal allocation $\boldsymbol{a}^*$, we first calculate the "fractional part" of $\boldsymbol{a}^*$ and observe that this defines a fractional solution to a transportation problem. Next, we normalize the supply and demand to obtain a doubly stochastic matrix $\{\tilde{a}_{mi,nj}\}$. This allows us to apply Birkhoff's algorithm (also known as the Birkhoff–von Neumann decomposition), which takes a doubly stochastic matrix and returns a convex combination of permutation matrices. Finally, we reconstruct a feasible integer allocation for each permutation.

Specifically, given a (fractional) solution $\mathbf{a}^*_0 \in \mathcal{A}^*(\mathbf{b},\mathbf{d})$ we start by defining $c_m = C - b_m$ for all $m \in \mathcal{M}$. We capture the slack in the allocation by adding nodes using

$$\mathcal{M}_+ = \mathcal{M} \cup \{0\}, \qquad \mathcal{N}_+ = \mathcal{N} \cup \{0\}.$$

and defining $c_0 = \sum_{n\in\mathcal{N}} d_n$ and $d_0 = \sum_{m\in\mathcal{M}} c_m$. Using these definitions, we introduce the slack variables

$$\begin{aligned}
a^*_{0,m,0} &= c_m - \sum_{n\in\mathcal{N}} a^*_{0,m,n}, \quad \forall m \in \mathcal{M},\\
a^*_{0,0,n} &= d_n - \sum_{m\in\mathcal{M}} a^*_{0,m,n}, \quad \forall n \in \mathcal{N},\\
a^*_{0,0,0} &= \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}} a^*_{0,m,n}.
\end{aligned}$$

By construction, it follows that

$$\begin{aligned}
\sum_{n\in\mathcal{N}_+} a^*_{0,m,n} &= c_m, \quad \forall m \in \mathcal{M}_+,\\
\sum_{m\in\mathcal{M}_+} a^*_{0,m,n} &= d_n, \quad \forall n \in \mathcal{N}_+.
\end{aligned}$$

Next, we define

$$\begin{aligned} \hat{a}_{m,n} &= a^*_{0,m,n} - \lfloor a^*_{0,m,n} \rfloor, && \forall m \in \mathcal{M}_+, n \in \mathcal{N}_+, \\ \hat{c}_m &= c_m - \sum_{n \in \mathcal{N}_+} \lfloor a^*_{0,m,n} \rfloor, && \forall m \in \mathcal{M}_+, \\ \hat{d}_n &= d_n - \sum_{m \in \mathcal{M}_+} \lfloor a^*_{0,m,n} \rfloor, && \forall n \in \mathcal{N}_+. \end{aligned}$$

As a result, we have

$$\begin{aligned} \sum_{n \in \mathcal{N}_+} \hat{a}_{m,n} &= \hat{c}_m, && \forall m \in \mathcal{M}_+, \\ \sum_{m \in \mathcal{M}_+} \hat{a}_{m,n} &= \hat{d}_n, && \forall n \in \mathcal{N}_+. \end{aligned}$$

Observe that $\sum_{m \in \mathcal{M}_+} \hat{c}_m = \sum_{n \in \mathcal{N}_+} \hat{d}_n$ by construction, and that $\hat{c}_m \leq N+1$ for all $m \in \mathcal{M}_+$ and $\hat{d}_n \leq M+1$ for all $n \in \mathcal{N}_+$. Without loss of generality, we can also assume that $\hat{c}_m > 0$ for all $m \in \mathcal{M}_+$ and that $\hat{d}_n > 0$ for all $n \in \mathcal{N}_+$. In addition, we note that $0 \leq \hat{a}_{m,n} \leq 1$ for all m, n.

We proceed by splitting the supply and demand nodes into nodes with unit supply and demand. Specifically, we define $\tilde{\mathcal{M}} = \{(m,i) : m \in \mathcal{M}_+, i \in \{1, \ldots, \hat{c}_m\}\}$, and $\tilde{\mathcal{N}} = \{(n,j) : n \in \mathcal{N}_+, j \in \{1, \ldots, \hat{d}_n\}\}$. We also define

$$\tilde{a}_{mi,nj} = \frac{\hat{a}_{m,n}}{\hat{c}_m \hat{d}_n}.$$

Thus yields

$$\begin{aligned} \sum_{(n,j) \in \tilde{\mathcal{N}}} \tilde{a}_{mi,nj} &= 1, && \forall (m,i) \in \tilde{\mathcal{M}}, \\ \sum_{(m,i) \in \tilde{\mathcal{M}}} \tilde{a}_{mi,nj} &= 1, && \forall (n,j) \in \tilde{\mathcal{N}}. \end{aligned}$$

The entries $\tilde{a}_{mi,nj}$ therefore define a doubly stochastic matrix. Using Birkhoff's algorithm (Brualdi 1982), we can express this matrix as a convex combination of permutation matrices. Thus

$$\tilde{\mathbf{a}} = \sum_{k=1}^{K} \lambda_k \tilde{\mathbf{a}}^k,$$

where $\lambda_k \geq 0$ for $k = 1, \ldots, K$ and $\sum_{k=1}^{K} \lambda_k = 1$, and where $\tilde{\mathbf{a}}^k$ is a permutation matrix (i.e., a $(0,1)$ matrix whose rows and columns sum to one). It is known that $K \leq (|\tilde{\mathcal{M}}| - 1)^2 + 1$ (Marcus and Ree 1959); that is, the number of permutation matrices in the convex combination is polynomially bounded in the dimension of the doubly stochastic matrix. Because $\hat{c}_m \leq N+1$ for all $m \in \mathcal{M}_+$ and $\hat{d}_n \leq M+1$ for all $n \in \mathcal{N}_+$, it therefore follows that the number is polynomially bounded in the number of days in the booking horizon and customer classes as well.

Putting it all together, we have that

$$\mathbf{a}_0^* = \sum_{k=1}^{K} \lambda_k \mathbf{a}^k,$$

with

$$\mathbf{a}_{m,n}^k = \lfloor a_{0,m,n}^* \rfloor + \sum_{i=1}^{\hat{c}_m} \sum_{j=1}^{\hat{d}_n} \tilde{\mathbf{a}}_{mi,nj}^k \quad \forall m \in \mathcal{M}, n \in \mathcal{N}, k \in \{1, \ldots, K\}.$$

We can verify that $\mathbf{a}^k \in \mathcal{A}(\mathbf{b}, \mathbf{d})$ for all k. This completes the proof. ■