A strong Lagrangian relaxation for general discrete-choice network revenue management

Sumit Kunnumkal[1] · Kalyan Talluri[2]

Abstract

Discrete-choice network revenue management (DC-NRM) captures both customer behavior and the resource-usage interaction of products, and is appropriate for airline and hotel revenue management, dynamic sales of bundles in advertising, and dynamic assortment optimization in retail. The state-space of the DC-NRM stochastic dynamic program explodes and approximation methods such as the choice deterministic linear program, the affine, and the piecewise-linear approximations have been proposed to approximate it in practice. The affine relaxation (and thereby, its generalization, the piecewise-linear approximation) is intractable even for the simplest choice models such as the multinomial logit (MNL) choice model with a single segment. In this paper we propose a new Lagrangian relaxation method for DC-NRM based on an extended set of multipliers. An attractive feature of our method is that the number of constraints in our formulation scales linearly with the resource capacities. While the number of constraints in our formulation is an order of magnitude smaller that the piecewise-linear approximation (polynomial vs exponential), it obtains a bound that is as tight as the piecewise-linear bound. If we assume that the consideration sets of the different customer segments are small in size—a reasonable modeling tradeoff in many practical applications—our method is an indirect way to obtain the piecewise-linear approximation on large problems effectively. Our results are not specific to a particular functional form (such as MNL), but hold for any discrete-choice model of demand. We show by numerical experiments that our Lagrangian relaxation method can provide substantial improvements over existing benchmark methods, both in terms of tighter upper bounds, as well as revenues from policies based on the relaxation.

Keywords Dynamic programming approximations · Transportation · Revenue management · Choice models

✉ Kalyan Talluri
kalyan.talluri@imperial.ac.uk

Extended author information available on the last page of the article

1 Introduction and literature review

In industries such as hotels, advertising, and airlines, the products consume one multiple resources (for example, multi-night stays, bundles of advertising time-slots, multi-leg itineraries) and each product has a specific price, set based on the price sensitivity of the customer segment it is aimed at, the market conditions, and the product characteristics. The resources are perishable; for instance, for an airline, empty at the moment of departure get no revenue, so the inventory "perishes" at departure. The firm's revenue management function is to decide, at every point in time during the sale period, what products to open for sale—the tradeoff being selling too much at too low a price early and running out of capacity, or, rejecting too many low-valuation customers and ending up with unsold inventory. Network revenue management (NRM) is control based on the demands for the entire network. Chapter 3 of Talluri and van Ryzin [23] contains the necessary background on NRM.

NRM incorporating more realistic models of customer behavior, as customers choosing from a set of offered products, was initiated in Talluri and van Ryzin [22] for the single-resource problem. Discrete-choice models are parsimonious as they model the probability of purchase as a function of product characteristics, such as price and restrictions, reducing the number of parameters to estimate. The subsequent optimization problem can be formulated as a stochastic dynamic program that we call discrete-choice NRM (DC-NRM) dynamic program.

The DC-NRM dynamic program is computationally intractable and hence many approximation methods have been proposed, starting with Gallego et al. [7] and Liu and van Ryzin [15] who formulate the choice deterministic linear program (*CDLP*). Zhang and Adelman [26] propose an affine approximation to the value function, while Meissner and Strauss [16] propose a piecewise-linear approximation. Kunnumkal and Topaloglu [14] use Lagrangian relaxation ideas to come up with a separable approximation. All of the above mentioned approximation methods obtain upper bounds on the value function, with the piecewise-linear approximation obtaining the tightest upper bound.

Unfortunately, most of these approximation methods themselves are intractable, even for simple choice models. Liu and van Ryzin [15] show that *CDLP* is tractable for the multinomial logit (MNL) model provided the subset of products of interest to the different customer segments (consideration sets) are disjoint. However, *CDLP* is NP-complete when the segment consideration sets overlap even under the MNL model with two customer segments (Bront et al. [5], Rusmevichientong et al. [18]). Kunnumkal and Talluri [11] show that the affine approximation of Zhang and Adelman [26] is NP-hard for the MNL model with even a single segment, which implies a similar hardness result for the piecewise-linear approximation as well. These negative results show us the limits of tractability even for the simplest choice model such as the MNL model.

One approach to obtain a tractable model is to approximate the underlying discrete choice model with a simplified model that limits the number of choices. For example, Chen and de Mello [6] consider a buy-up/buy-down model which is

tractable and can be solved efficiently. In this paper, we take a different approach to tractability; we work with the underlying discrete-choice model but come up with a tractable approximation to the value function of the DC-NRM dynamic program. In particular, we propose a new Lagrangian relaxation method for DC-NRM based on an extended set of multipliers. The number of constraints in our formulation scales linearly with the resource capacities, making it a tractable alternative for large networks. On the other hand, we show that our extended Lagrangian relaxation obtains a bound that is as tight as the piecewise-linear bound. The biggest practical impact of our work is in showing that the complexity of the Lagrangian relaxation method (in terms of the number of constraints in the linear programming formulation of the problem) scales linearly with the resource capacities, while that of the piecewise-linear approximation is exponential—yet they arrive at the same value function approximation. If we assume that the consideration sets of the different customer segments are small in size—a reasonable modeling tradeoff in many practical applications as we argue next—our method is able to solve relatively large problems, which would have been impossible if one were to try solving the piecewise-linear approximation directly.

Since we work with general discrete-choice models, all the negative complexity results for the MNL model carry over for the Lagrangian relaxation as well as we show that our Lagrangian relaxation is equivalent to the piecewise-linear relaxation. The great advantage however is that the number of constraints in the Lagrangian relaxation depends directly on the size of the consideration sets of the different segments. These fortunately are quite small in practice: in the airline setting for instance, a segment's consideration set consists of choices (on one airline) for travel between an origin and destination, and typically there are only a few alternatives on a given date (Talluri [20]). For hotels, as the product consists of a multi-night stay and most customers arrive with a fixed duration of stay in mind, the consideration set consists of the types of rooms and products, usually not a very large number.

Research in the marketing area also gives evidence supporting that customers have relatively small consideration sets: Hauser and Wernerfelt [9] report average consideration set sizes of 3 brands for deodorants, 4 brands for shampoos, 2.2 brands for air fresheners, 4 brands for laundry detergents and 4 brands for coffees. (Note that the study is for brands rather than choices of sizes or colors.) Another line of marketing research finds great value in deliberately limiting customer choices to a small number (Iyengar and Lepper [10]). Assuming small consideration set sizes, Talluri [21] and Meissner et al. [17] study tractable approximations to *CDLP* for DC-NRM. While our segment-based relaxation has the same underlying motivation, the development is quite different as we work with the Lagrangian relaxation of the DC-NRM dynamic program.

Our final contribution in this paper is showing by numerical experiments that our Lagrangian relaxation method can provide substantial improvements over existing methods.

Our work builds on previous research on Lagrangian relaxations for NRM. Topaloglu [24] was the first paper to propose a time-based Lagrangian relaxation approach for the perfect segmentation case (also called the independent-class assumption). The Lagrangian relaxation of Topaloglu [24] associates Lagrange

multipliers with each product in each time period, and Kunnumkal and Talluri [13] show that it obtains an upper bound that coincides with the piecewise-linear bound. In contrast, for DC-NRM, we give an example which illustrates that the approach proposed by Topaloglu [24] can be weaker than the piecewise-linear approximation. This motivates our approach that associates Lagrange multipliers with every offer set, and obtains a bound that is as tight as the piecewise-linear bound. We note that the Lagrangian relaxation we propose is different from that in Topaloglu [24] and consequently has different structural properties. Moreover, an appealing feature of our method is that it extends very naturally to the case where there are multiple customer segments each interested in a small subset of the products. We build on our Lagrangian relaxation idea and propose a segment-based approximation that is a tractable alternative when the considerations sets of the different segments are small in size.

In parallel to a working version of this paper, Vossen and Zhang [25] study the properties of approximate linear programming methods for NRM and show that the affine and piecewise-linear formulations can be significantly reduced in size. Their reduced formulation of the piecewise-linear program for DC-NRM can be shown to be equivalent to our Lagrangian relaxation method (Sect. 5.4 in Vossen and Zhang [25]). We note however that our starting point is a Lagrangian relaxation of the DC-NRM dynamic program and our proof technique is quite different. Moreover, we show that it is not possible to recover the piecewise-linear approximation upper bound for DC-NRM by using product-specific multipliers as in the perfect segmentation case. Consequently, the expanded form, using offer-set specific multipliers is the best we can hope for. Finally, even though the number o of our Lagrangian relaxation (and the reduced piecewise-linear approximation) is linear in the number of resources, it is still exponential in the number of products. We propose a new segment-based Lagrangian relaxation that remains tractable provided the consideration sets of the different customer segments are small in size.

To summarize, we make the following research contributions in this paper:

1. We propose a new Lagrangian relaxation method for DC-NRM based on an extended set of Lagrange multipliers.
2. The number of constraints in our Lagrangian relaxation (in the linear programming formulation of the problem) scales linearly with the resource capacities making it a tractable alternative for large networks. Another appealing feature of our approach is that it decomposes the network problem into a number of single resource problems, which can be potentially solved in a distributed fashion.
3. We show that our Lagrangian relaxation obtains an upper bound on the value function as tight as the piecewise-linear bound. The number of constraints in the Lagrangian relaxation scales linearly with the resource capacities, while the piecewise-linear approximation is exponential, hence leading to a substantial reduction in running time in practice. This result also implies that the Lagrangian relaxation bound is stronger than the affine approximation bound (hence also *CDLP*) that was not known previously. It should be noted that our results are not specific to a particular functional form (such as MNL), but hold at a full level of generality, for any discrete-choice model of demand.

4. We show how our ideas lead to a tractable method for DC-NRM (for any discrete-choice model) when the consideration sets of the different segments are small in size.

5. Our numerical experiments indicate that our Lagrangian relaxation can obtain significantly tighter upper bounds than the benchmark methods including the *CDLP* and the affine approximation. We also extract policies from our formulation and show by numerical experiments that it can lead to notable improvements in revenue over existing methods.

The remainder of the paper is organized as follows: In Sect. 2 we describe the DC-NRM model, the notation, and the stochastic dynamic programming formulation of the problem. Section 3 reviews the value function approximation methods in the literature including *CDLP*, the affine and the piecewise-linear approximation methods. In Sect. 4 we describe the Lagrangian relaxation approach and show that it obtains a bound that is as tight as the piecewise-linear bound; we defer the formal proof to Sect. 5. Section 6 describes the segment-based Lagrangian relaxation that is tractable under the assumption of small consideration sets. In Sect. 7 we perform numerical experiments to compare the performance of the Lagrangian relaxation approach with benchmark solution methods.

2 Problem formulation

A product is a specification of a price and the set of resources that it consumes. For example, for an airline a product is an itinerary-fare class combination, where an itinerary is a combination of flight legs making up a passenger's journey; for a hotel, a product is a multi-night stay for a particular room type at a certain price point.

Our model is discrete-time with τ intervals, indexed by t. The booking horizon begins at time $t = 1$ and ends at $t = \tau$; all the resources perish instantaneously at time $\tau + 1$. We make the standard assumption that the time intervals are fine enough so that the probability of more than one customer arriving in any single time period is negligible. We index resources by i, products by j, and time periods by t. We refer to the set of all resources as $\mathcal{I}$ and the set of all products as $\mathcal{J}$. We let f_j denote the revenue associated with product j and $\mathcal{I}_j \subset \mathcal{I}$ the subset of resources it uses. Similarly, we let $\mathcal{J}_i \subset \mathcal{J}$ denote the subset of products that use resource i.

We use superscripts on vectors to index the vectors and subscripts to indicate their components. For example, we write the resource capacity vector associated with time period t as $\boldsymbol{r}^t$ and use r_i^t to represent the capacity on resource i in time period t. Therefore, $\boldsymbol{r}^1 = \left[r_i^1 \right]$ represents the initial capacities on the resources and $\boldsymbol{r}^t = \left[r_i^t \right]$ denotes the remaining capacities on the resources at the beginning of time period t. The remaining capacity r_i^t takes values in the set $\mathcal{R}_i = \left\{ 0, \dots, r_i^1 \right\}$ and $\mathcal{R} = \prod_i \mathcal{R}_i$ represents the state space.

2.1 Demand model

In each time period the firm offers a subset S of its products for sale, called the *offer set*. A customer arrives with probability α and given an offer set S, an arriving customer purchases a product j in the set S or decides not to purchase. The no-purchase option is indexed by 0 and is always present for the customer. We let $P_j(S)$ denote the probability that the firm sells product j given that a customer arrives and the offer set is S. Clearly, $P_j(S) = 0$ if $j \notin S$. The probability of no sale given a customer arrival is $P_0(S) = 1 - \sum_{j \in S} P_j(S)$. We assume that the choice probabilities are given by an oracle, as the model represents a general discrete-choice model; they could conceivably be calculated by a simple formula as in the case of the multinomial logit (MNL) model (Ben-Akiva and Lerman [3]).

2.2 DC-NRM dynamic program

The dynamic program (DP) to determine optimal controls is as follows: Let $V_t(\mathbf{r}^t)$ denote the maximum expected revenue to go, given remaining capacity $\mathbf{r}^t$ at the beginning of period t. Then $V_t(\mathbf{r}^t)$ must satisfy the Bellman equation

$$
V_t(\mathbf{r}^t) = \max_{S \subset \mathcal{Q}(\mathbf{r}^t)} \left\{ \sum_{j \in S} \alpha P_j(S) \left[f_j + V_{t+1}\left(\mathbf{r}^t - \sum_{i \in \mathcal{I}_j} \mathbf{e}^i\right) \right] + \left[\alpha P_0(S) + 1 - \alpha\right] V_{t+1}\left(\mathbf{r}^t\right) \right\},
$$
(1)

where $\mathbf{e}^i$ is a vector with a 1 in the ith position and 0 elsewhere and

$$
\mathcal{Q}(\mathbf{r}) = \left\{ j \mid \mathbb{1}_{[j \in \mathcal{J}_i]} \leq r_i, \ \forall i \right\}
$$

represents the set of products that can be offered given the capacity vector $\mathbf{r}$. We use $\mathbb{1}_{[\cdot]}$ as the indicator function, 1 if true and 0 if false. The boundary conditions are $V_{\tau+1}(\mathbf{r}) = V_t(\mathbf{0}) = 0$ for all $\mathbf{r}$ and for all t, where $\mathbf{0}$ is a vector of all zeroes. $V^{DP} = V_1(\mathbf{r}^1)$ denotes the optimal expected total revenue over the booking horizon, given the initial capacity vector $\mathbf{r}^1$.

For brevity of notation, we assume that $\alpha = 1$ in the remaining part of the paper. We note that this is without loss of generality since this is equivalent to letting $\tilde{P}_j(S) = \alpha P_j(S)$ and $\tilde{P}_0(S) = \alpha P_0(S) + 1 - \alpha$, and working with the choice probabilities $\left\{ \tilde{P}_j(S) \mid \forall j, S \right\}$.

2.3 Linear programming formulation of the DC-NRM dynamic program

The value functions can, alternatively, be obtained by solving a linear program, following Schweitzer and Seidmann [19]. The linear programming formulation of the DC-NRM dynamic program has a decision variable $V_t(\mathbf{r})$ for each state vector $\mathbf{r}$ in each period t and is as follows:

$$V^{DP} = \min_V V_1(\boldsymbol{r}^1)$$

s.t.

$$(DP)\, V_t(\boldsymbol{r}) \geq \sum_j P_j(S)\left[f_j + V_{t+1}\left(\boldsymbol{r} - \sum_{i \in \mathcal{I}_j} \boldsymbol{e}^i\right) - V_{t+1}(\boldsymbol{r})\right] + V_{t+1}(\boldsymbol{r}) \qquad (2)$$

$$\forall \boldsymbol{r} \in \mathcal{R}, S \subset \mathcal{Q}(\boldsymbol{r}), t$$

$$V_t(\boldsymbol{r}) \geq 0 \quad \forall \boldsymbol{r} \in \mathcal{R}, t$$

with the boundary conditions that $V_{\tau+1}(\boldsymbol{r}) = V_t(\boldsymbol{0}) = 0$ for all $\boldsymbol{r}$ and for all t. Both the dynamic program (1) and linear program DP are computationally intractable, but the linear program DP turns out to be useful in developing value function approximation methods.

3 Value function approximation methods

In this section, we describe three methods in the literature to approximate the DC-NRM dynamic program value function. We begin with the choice-based deterministic linear program and then outline the affine and the piecewise-linear approximations.

3.1 Choice-based deterministic linear program (*CDLP*)

The choice-based deterministic linear program *CDLP*), proposed in Gallego et al. [7] and Liu and van Ryzin [15] is given by

$$V^{CDLP} = \max_\rho \sum_t \sum_S [\sum_{j \in S} P_j(S)f_j]\rho_{S,t}$$

s.t.

$$(CDLP) \sum_t \sum_S \left[\sum_{j \in \mathcal{J}_i} P_j(S)\right]\rho_{S,t} \leq r_i^1 \quad \forall i \qquad (3)$$

$$\sum_S \rho_{S,t} = 1 \quad \forall t$$

$$\rho_{S,t} \geq 0 \quad \forall t, S.$$

In the above linear program, we interpret the decision variable $\rho_{S,t}$ as the frequency with which set S is offered at time period t. The objective function captures the total expected revenue, while the first set of constraints ensure that the total expected

capacity consumed on each resource does not exceed its available capacity. The second set of constraints ensures that the total frequencies add up to 1 at each time period.

Liu and van Ryzin [15] show that the optimal objective function value of *CDLP* gives an upper bound on the optimal expected revenue. That is, $V_1(\mathbf{r}^1) \leq V^{CDLP}$. Since *CDLP* has an exponential number of decision variables it has to be solved using column generation. Liu and van Ryzin [15] show that column generation is efficient for the MNL model with a single segment. However, the column generation procedure is intractable in general, NP-complete for MNL with just two segments (Bront et al. [5] and Rusmevichientong et al. [18]).

3.2 Affine approximation

The affine approximation, proposed by Zhang and Adelman [26], replaces the value function $V_t(\mathbf{r})$ with the affine function $\theta_t + \sum_i V_{i,t} r_i$ in the linear program *DP* to obtain the following linear program:

$$V^{AF} = \min_{\theta, V} \theta_1 + \sum_i V_{i,1} r_i^1$$

$$\text{s.t.}$$

$$(AF) \theta_t + \sum_i V_{i,t} r_i \geq \sum_j P_j(S) \Big[f_j - \sum_{i \in \mathcal{I}_j} V_{i,t+1} \Big] + \theta_{t+1} + \sum_i V_{i,t+1} r_i$$

$$\forall \mathbf{r} \in \mathcal{R}, S \subset \mathcal{Q}(\mathbf{r}), t$$

$$\theta_t, V_{i,t} \geq 0 \forall i, t.$$

Zhang and Adelman [26] show that the optimal objective function value of *AF* gives an upper bound on the optimal expected revenue and this bound is tighter than the *CDLP* bound. While the number of decision variables in *AF* is manageable, the number of constraints is exponential both in the number of resources as well as the products. Vossen and Zhang [25] show that *AF* has a reduced formulation where the number of constraints is exponential only in the number of products. Still, the problem has to be solved by constraint generation and the separation problem is intractable even for the MNL choice model with a single segment (Kunnumkal and Talluri [11]).

3.3 Piecewise-linear approximation

Meissner and Strauss [16] propose the piecewise-linear approximation, which approximates the value function $V_t(\mathbf{r})$ with $\sum_i v_{i,t}(r_i)$ in *DP*. The resulting linear program is

$$V^{PL} = \min_v \sum_i v_{i,1}(r_i^1)$$

s.t.

$$(PL) \sum_i v_{i,t}(r_i) \geq \sum_j P_j(S)\Big[f_j + \sum_{i \in \mathcal{I}_j} \big(v_{i,t+1}(r_i - 1) - v_{i,t+1}(r_i)\big)\Big] + \sum_i v_{i,t+1}(r_i)$$

$$\forall r \in \mathcal{R}, S \subset Q(r), t$$

$$v_{i,\tau+1}(\cdot) = 0 \quad \forall i, t.$$

$$(4)$$

Since the piecewise-linear approximation uses a more refined approximation architecture than the affine approximation, it is natural to expect that it obtains a tighter bound on the value function. Indeed, Meissner and Strauss [16] show $V^{DP} \leq V^{PL} \leq V^{AF} \leq V^{CDLP}$. However, as with *CDLP* and *AF*, tractability remains as much, if not more, of an issue for *PL* as well. Note that linear program *PL* has $\mathcal{O}(2^{|\mathcal{J}|}\tau \prod_i r_i^1)$ constraints, which is exponential both in the number of products and the number of resources. Moreover, the separation problem of this linear program is NP-complete, even for MNL with a single segment; see Kunnumkal and Talluri [11].

4 Lagrangian relaxation using offer-set specific multipliers

In this section, we present our approximation method to obtain an upper bound on the value function of the DC-NRM dynamic program. We first describe our approach and then show that it obtains an upper bound that is as tight as that obtained by *PL*.

We begin by introducing some notation. For an offer set *S*, we let

$$\mathcal{I}_S = \big\{ i \in \mathcal{I} \mid \exists j \in S \text{ with } j \in \mathcal{J}_i \big\},$$

be the set of resources used by the products in *S*. We follow the convention that the empty set does not use any resource. We also let $C_i = \big\{S \mid i \in \mathcal{I}_S\big\}$ and note that $i \in \mathcal{I}_S$ if and only if $S \in C_i$. Therefore, C_i can be roughly interpreted as the collection of offer sets that use resource *i*. Optimality equation (1) can be equivalently written, in an expanded form, as

$$V_t(r) = \max_h \sum_S h_{\phi,S,t}\left\{\sum_{j \in S} P_j(S)\Big[f_j + V_{t+1}\big(r - \sum_{i \in \mathcal{I}_j} e^i\big) - V_{t+1}(r)\Big]\right\} + V_{t+1}(r)$$

s.t.

$$h_{i,S,t} = h_{\phi,S,t} \; \forall S, i \in \mathcal{I}_S$$

$$(5)$$

$$h_{i,S,t} \leq r_i \; \forall i, S \in C_i$$

$$(6)$$

$$\sum_S h_{\phi,S,t} \le 1 \tag{7}$$

$$\sum_{S \in C_i} h_{i,S,t} \le 1 \, \forall i \tag{8}$$

$$h_{\phi,S,t}, h_{i,S,t} \ge 0 \, \forall S, i \in \mathcal{I}_S \tag{9}$$

where the decision variable $h_{\phi,S,t}$ can be interpreted as the frequency with which set S is offered at time period t. In the expanded formulation, we also have resource level decision variables $h_{i,S,t}$, which can be interpreted as the frequency with which set S is offered on resource i at time period t. Constraints (5) ensure that for each set S, the frequencies are identical across the resources used by it, while constraints (6) ensure that a set is offered only if there is sufficient capacity on each resource that it uses. Constraints (7)-(9) ensure that the frequencies are nonnegative and add up to at most 1. We note that although the expanded formulation has a number of redundant decision variables and constraints, they turn out to be useful in the development of our approximation method below.

The expanded formulation still has a high dimensional state space and remains intractable. Our approximation method reduces the dimensionality by solving a number of single dimensional dynamic programs, one for each resource. We relax constraints (5) and associate Lagrange multipliers $\lambda_{i,S,t}$ with them. This decouples the decisions across the different resources and makes the optimization problem separable across the resources. Indeed, it can be shown by induction that the value function approximation for the relaxed problem can be obtained by solving a number of single dimensional dynamic programs; one for each resource and an auxiliary dynamic program. In particular, we solve the optimality equation

$$
\begin{aligned}
\vartheta_{i,t}^{\lambda}(r_i) = \max_{h_i} \sum_{S \in C_i} & \left\{ h_{i,S,t}\left(\lambda_{i,S,t} + \sum_{j \in \mathcal{J}_i} P_j(S)\left[\vartheta_{i,t+1}^{\lambda}(r_i - 1) - \vartheta_{i,t+1}^{\lambda}(r_i) \right] \right) \right\} \\
& + \vartheta_{i,t+1}^{\lambda}(r_i)
\end{aligned}
\tag{10}
$$

$$\text{s.t.}$$
$$(6), (8), h_{i,S,t} \ge 0 \, \forall S \in C_i$$

for resource i with the boundary condition that $\vartheta_{i,\tau+1}^{\lambda}(\cdot) = 0$, where the superscript emphasizes the dependence of the value function approximation on the Lagrange multipliers. Collecting the terms involving $h_{\phi,S,t}$ in the objective function of the relaxed problem, we solve the auxiliary dynamic program

$$
\vartheta_{\phi,t}^{\lambda} = \max_{h_\phi} \sum_S \left\{ h_{\phi,S,t}\Big[\sum_{j \in S} P_j(S)f_j - \sum_{i \in \mathcal{I}_S} \lambda_{i,S,t} \Big] \right\} + \vartheta_{\phi,t+1}^{\lambda}
\tag{11}
$$

$$\text{s.t.}$$
$$(7), h_{\phi,S,t} \ge 0 \, \forall S$$

with the boundary condition that $\vartheta^\lambda_{\phi,\tau+1} = 0$. We use

$$V^\lambda_t(\boldsymbol{r}) = \sum_i \vartheta^\lambda_{i,t}(r_i) + \vartheta^\lambda_{\phi,t} \tag{12}$$

as our value function approximation corresponding to the Lagrangian relaxation.

Lemma 1 below shows that the Lagrangian relaxation using offer-set specific multipliers (*LRo*) obtains an upper bound on the optimal expected revenue and that this bound is potentially weaker than the piecewise-linear bound.

Lemma 1
$$V^{DP} \le V^{PL} \le V^\lambda_1(\boldsymbol{r}^1).$$

Proof Appendix. □

We obtain the tightest upper bound on the optimal expected revenue by solving

$$V^{LRo} = \min_\lambda V^\lambda_1(\boldsymbol{r}^1).$$

The following proposition gives the subgradients of $\vartheta^\lambda_{i,t}(r_i)$ and $\vartheta^\lambda_{\phi,t}$ (thereby also showing they are convex functions of λ). It follows that $V^\lambda_t(\boldsymbol{r})$ is also a convex function of λ and hence the above optimization problem can be efficiently solved as a convex program, for example by using subgradient search (Bertsekas [4], Chapter 7). That is, for a given set of Lagrange multipliers λ, we can solve the single-dimensional dynamic programs (10) and (11) to obtain $\vartheta^\lambda_{i,t}(\cdot)$ and $\vartheta^\lambda_{\phi,t}$, respectively. We can then use Proposition 1 below to obtain a subgradient of $V^\lambda_1(\boldsymbol{r}^1)$ and update the Lagrange multipliers.

We introduce some notation for this purpose. Let $h^\lambda_{i,S,t}(r_i)$ denote an optimal solution to problem (10) where the arguments emphasize the dependence of the optimal solution on the Lagrange multipliers and the remaining capacity on the resource. We define $h^\lambda_{\phi,S,t}$ in a similar manner for (11). Also, let $X^\lambda_{i,t}$ denote the random variable which represents the remaining capacity on resource i at time period t when we offer sets according to the optimal solution to problem (10). We have the following result.

Proposition 1 *Let λ and $\hat\lambda$ be two sets of Lagrange multipliers. Then,*
1. $\vartheta^{\hat\lambda}_{i,t}(r_i) \ge \vartheta^\lambda_{i,t}(r_i) + \sum_{k=t}^\tau \sum_{S \in C_i} \left\{ \sum_{x=0}^{r_i} \Pr\{X^\lambda_{i,k} = x \mid X^\lambda_{i,t} = r_i\} h^\lambda_{i,S,t}(x) \right\} [\hat\lambda_{i,S,k} - \lambda_{i,S,k}].$
2. $\vartheta^{\hat\lambda}_{\phi,t} \ge \vartheta^\lambda_{\phi,t} + \sum_{k=t}^\tau \sum_i \sum_{S \in C_i} h^\lambda_{\phi,S,k} [\hat\lambda_{\phi,S,k} - \lambda_{\phi,S,k}].$

Proof Appendix. □

We note that besides showing that $V^\lambda_1(\boldsymbol{r})$ is a convex function of λ, Proposition 1 also gives an explicit expression for its subgradient. This allows us to use subgradient search to find the optimal set of Lagrange multipliers. Proposition 2 below shows that by doing this, we in fact obtain the piecewise-linear bound.

Proposition 2
$$V^{PL} = V^{LRo}.$$

We defer the proof of Proposition 2 to the next section. Here we note some of its implications. Proposition 2 together with the results in Meissner and Strauss [16] implies that $V^{LRo} \leq V^{CDLP}$ and $V^{LRo} \leq V^{AF}$. Therefore, *LRo* obtains an upper bound that is tighter than both *CDLP* and *AF*. It is also quite surprising that the *LRo* bound is as tight as *PL* since the complexity (the number of constraints in the linear programming formulation) of *LRo* is linear in the resource capacities ($\sum_i r_i^1$), while that of *PL* is exponential ($\prod_i r_i^1$). For typical DC-NRM instances, the number of constraints in *LRo* can be orders of magnitude smaller than that of *PL*. Moreover, since *LRo* decomposes the network problem into a number of single resource problems that can be solved in parallel, it may also be more suitable for distributed computing techniques.

In closing this section, we note that while Lagrangian relaxation ideas have been applied to NRM previously, there are some important differences between our approach and previous proposals. Under the independent-class assumption, Topaloglu [24] proposes relaxing the constraints that the same acceptance decisions be made for each product across the resources that it uses. The resulting Lagrangian relaxation associates a multiplier $\lambda_{i,j,t}$ for each product j and each resource $i \in \mathcal{J}_i$. Kunnumkal and Talluri [13] show that the Lagrangian relaxation with product-specific multipliers (*LRp*) turns out be equivalent to *PL* in the independent-class setting. We give an example in the Appendix which illustrates that the same result fails to hold for DC-NRM. That is, for DC-NRM, *LRp* can be weaker than *PL* (and by Proposition 2, weaker than *LRo* as well). We also note that while *LRp* has a smaller number of multipliers than *LRo*, solving the resource level problem (27) in *LRp* is still intractable for a general discrete-choice model. Consequently, it is not clear whether *LRp* provides any significant computational benefits over *LRo*.

The optimality equation (11) is not redundant for DC-NRM as we show in an example in the Appendix which illustrates that the optimal multipliers in *LRo* can be such that $\sum_{i \in \mathcal{I}_S} \lambda_{i,S,t} < \sum_{j \in S} P_j(S) f_j$. This is in contrast to the perfect segmentation case where it is known that the optimal Lagrange multipliers can be interpreted as the pro-rated fare of a product on each resource that it uses, and consequently $v_{\phi,t}^\lambda = 0$ for all t.

5 Tightness of the Lagrangian bound

In this section we formally show that the upper bound obtained by *LRo* is as strong as the *PL* bound. We begin with some preliminary results in Sect. 5.1 before proceeding to the proof of Proposition 2 in Sect. 5.2.

5.1 Preliminaries

We note that optimality equation (10) can be equivalently written as

$$\vartheta_{i,t}^{\lambda}(r_i) = \max_{S \subset Q_i(r_i)} \left\{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} + \sum_{j \in \mathcal{J}_i} P_j(S) \left[\vartheta_{i,t+1}^{\lambda}(r_i - 1) - \vartheta_{i,t+1}^{\lambda}(r_i) \right] \right.$$
$$\left. + \vartheta_{i,t+1}^{\lambda}(r_i) \right\} \tag{13}$$

where $Q_i(r_i) = \left\{ j \mid \mathbb{1}_{[j \in \mathcal{J}_i]} \leq r_i \right\}$ denotes the set of products that can be offered considering the available capacity on resource i. On the other hand, letting $R(S) = \sum_{j \in S} P_j(S) f_j$ denote the expected revenue from offering set S and $\lambda_{\phi,S,t} = R(S) - \sum_{i \in \mathcal{I}_S} \lambda_{i,S,t}$, we can write optimality equation (11) equivalently as

$$\vartheta_{\phi,t}^{\lambda} = \max_{S \subset \mathcal{J}} \left\{ \lambda_{\phi,S,t} \right\} + \vartheta_{\phi,t+1}^{\lambda}. \tag{14}$$

If we let $\lambda = \{ \lambda_{\phi,S,t}, \lambda_{i,S,t} \mid \forall t, S, i \in \mathcal{I}_S \}$ and

$$\Lambda = \left\{ \lambda \mid \lambda_{\phi,S,t} + \sum_{i \in \mathcal{I}_S} \lambda_{i,S,t} = R(S), \forall S \subset \mathcal{J}, t \right\}, \tag{15}$$

then we can write $V^{LRo} = \min_{\lambda \in \Lambda} V_1^{\lambda}(r^1)$. The above minimization problem can be represented as the linear program

$$V^{LRo} = \min_{\lambda, v} \sum_i v_{i,1}(r_i^1) + v_{\phi,1}$$

$$\text{s.t.}$$

$$(LRo) v_{i,t}(r_i) \geq \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} + \sum_{j \in \mathcal{J}_i} P_j(S) \left[v_{i,t+1}(r_i - 1) - v_{i,t+1}(r_i) \right]$$

$$+ v_{i,t+1}(r_i) \quad \forall t, i, r_i \in \mathcal{R}_i, S \subset Q_i(r_i)$$

$$v_{\phi,t} \geq \lambda_{\phi,S,t} + v_{\phi,t+1} \quad \forall t, S$$

$$\sum_{i \in \mathcal{I}_S} \lambda_{i,S,t} + \lambda_{\phi,S,t} = R(S) \quad \forall t, S$$

$$v_{\phi,\tau+1} = 0, v_{i,\tau+1}(\cdot) = 0 \quad \forall i.$$

Linear program *LRo* turns out to be useful when we make the connection between the *LRo* and *PL* bounds.

5.2 Proof of Proposition 2

We give a formal proof of Proposition 2. Lemma 1 implies that $V^{PL} \leq \min_{\lambda \in \Lambda} V_1^{\lambda}(r^1) = V^{LRo}$. So it only remains to show the more difficult part: $V^{PL} \geq V^{LRo}$. Let $\psi_{i,t+1}(r_i) = v_{i,t+1}(r_i) - v_{i,t+1}(r_i - 1)$ and $\Delta_{i,t}(r_i) = v_{i,t+1}(r_i) - v_{i,t}(r_i)$. Note that $\psi_{i,t+1}(\cdot)$ can be interpreted as the marginal value of capacity, while $\Delta_{i,t}(\cdot)$ can be interpreted as the marginal value of time. The proof proceeds by considering the *PL*

separation problem: For each period t, given values of $\{v_{i,t}(r_i) \mid \forall t, i, r_i \in \mathcal{R}_i\}$, to check if

$$
\begin{aligned}
\Phi_t(v) = \max_{r \in \mathcal{R}, S \subset \mathcal{Q}(r)} \sum_{j \in S} P_j(S) \left[f_j + \sum_{i \in \mathcal{I}_j} [v_{i,t+1}(r_i - 1) - v_{i,t+1}(r_i)] \right] \\
+ \sum_i \left[v_{i,t+1}(r_i) - v_{i,t}(r_i) \right]
\end{aligned}
\tag{16}
$$

is less than or equal to zero. If yes, constraint (4) is satisfied at time period t for all r and $S \subset \mathcal{Q}(r)$. Otherwise, we find a violated constraint and add it to the linear program.

Equation (16) can be written as

$$
\Phi_t(v) = \max_{r \in \mathcal{R}, S \subset \mathcal{Q}(r)} \sum_j P_j(S) \left[f_j - \sum_{i \in \mathcal{I}_j} \psi_{i,t+1}(r_i) \right] + \sum_i \Delta_{i,t}(r_i).
$$

Using $R(S) = \sum_{j \in S} P_j(S) f_j$ and the facts that $P_j(S) = 0$ for $j \notin S$, and $\mathbb{1}_{[S \in \mathcal{C}_i]} = 1$ for $j \in S$ and $i \in \mathcal{I}_j$, we can write $\Phi_t(v)$ as

$$
\Phi_t(v) = \max_{r \in \mathcal{R}, S \subset \mathcal{Q}(r)} R(S) - \sum_{j \in S} P_j(S) \sum_{i \in \mathcal{I}_j} \mathbb{1}_{[S \in \mathcal{C}_i]} \psi_{i,t+1}(r_i) + \sum_i \Delta_{i,t}(r_i).
\tag{17}
$$

Next, consider the optimization problem

$$
\begin{aligned}
\Pi_t(v) = \min_{\lambda \in \Lambda} \Bigg\{ \sum_i \max_{r_i \in \mathcal{R}_i, S \subset \mathcal{Q}_i(r_i)} \left\{ \mathbb{1}_{[S \in \mathcal{C}_i]} \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\} \\
+ \max_{S \subset \mathcal{J}} \{ \lambda_{\phi,S,t} \} \Bigg\}.
\end{aligned}
$$

$\Pi_t(v)$ can be interpreted as a Lagrangian relaxation of $\Phi_t(v)$ which is obtained by relaxing the requirement that the same offer set decision be made for all the resources. This can be shown formally by following the same steps that lead to optimality equations (10) and (11), and we omit the details. Here we only note that equations (10) and (11) are equivalent to equations (13) and (14), respectively. Furthermore, note the similarity between the terms on the right hand side of the equation defining $\Pi_t(v)$ and the right hand sides of optimality equations (13) and (14).

We show below that $\Phi_t(v) = \Pi_t(v)$ and use this result to show that $V^{PL} \geq V^{LRo}$. We begin with some preliminary results. Lemma 2 shows that $\Pi_t(v)$ is an upper bound on $\Phi_t(v)$. This is intuitive if we interpret $\Pi_t(v)$ as being a relaxation of $\Phi_t(v)$. We defer the formal proof to the Appendix.

Lemma 2 $\Phi_t(v) \leq \Pi_t(v)$.

It remains to show that $\Phi_t(v) \geq \Pi_t(v)$. At a high level, this is because the optimal Lagrange multipliers are able to coordinate the offer set decisions across the different resources. As a result, we are able to construct an offer set that is feasible to the maximization problem on the right hand side of (17) from the optimal solution to $\Pi_t(v)$. The following lemma shows that we can restrict ourselves to sets in $C_i = \{S \mid i \in \mathcal{I}_S\}$ when solving the optimization problem for resource i.

Lemma 3

$$\max_{r_i \in \mathcal{R}_i, S \subset Q_i(r_i)} \left\{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\}$$

$$= \max \left\{ \Delta_{i,t}(0), \max_{r_i \in \{1,\dots,r_i^1\}, S \in C_i} \{ \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \} \right\}.$$

Proof Appendix. □

If we let $w_{i,t} = \max_{r_i \in \mathcal{R}_i, S \subset Q_i(r_i)} \left\{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\}$ and $w_{\phi,t} = \max_{S \subset \mathcal{J}} \{\lambda_{\phi,S,t}\}$, then Lemma 3 implies that we can write $\Pi_t(v)$ as the linear program

$$\Pi_t(v) = \min_{\lambda, w} w_{\phi,t} + \sum_i w_{i,t}$$

$$\text{s.t.} \tag{18}$$

$$(LP_{\Pi_t(v)}) w_{i,t} \geq \Delta_{i,t}(0) \quad \forall i$$

$$w_{i,t} \geq \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r) + \Delta_{i,t}(r) \quad \forall i, r \in \{1, \dots, r_i^1\}, S \in C_i \tag{19}$$

$$w_{\phi,t} \geq \lambda_{\phi,S,t} \quad \forall S \tag{20}$$

$$w_{\phi,t} \geq 0 \tag{21}$$

$$\lambda_{\phi,S,t} + \sum_{i \in \mathcal{I}_S} \lambda_{i,S,t} = R(S) \quad \forall S. \tag{22}$$

In the following, we analyze the structure of a particular optimal solution to $LP_{\Pi_t(v)}$ that allows us to construct a feasible solution to $\Phi_t(v)$, which in turn shows that $\Phi_t(v) \geq \Pi_t(v)$. We introduce some notation that will be useful for this purpose. Given a solution (λ, w) to $LP_{\Pi_t(v)}$, let

$$\xi_{i,S,t}(r) = w_{i,t} - \left[\lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r) + \Delta_{i,t}(r) \right]$$

denote the slack in constraint (19) for resource i, offer set $S \in C_i$ and $r \in \{1, \dots, r_i^1\}$. Let

$$B_{i,S}(\lambda, w) = \left\{ r \in \{1, \dots, r_i^1\} \mid \xi_{i,S,t}(r) = 0 \right\}$$

denote the set of capacity levels for which constraint (19) is binding for resource i and offer set S. We use the arguments (λ, w) emphasize the dependence of the set of binding constraints on the given solution. Observe that if $B_{i,S}(\lambda, w)$ is empty, then $w_{i,t} > \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S)\psi_{i,t+1}(r) + \Delta_{i,t}(r)$ for all $r \in \{1, \dots, r_{i,1}\}$.

From now on, we use a concept of optimal solutions with minimal set of binding constraints. The linear program $(LP_{\Pi_t(v)})$ has a finite optimal solution and possibly multiple ones. Naturally, any optimal solution has a set of binding constraints out of (18)–(22). Given any optimal solution, we can look for another optimal solution whose set of binding constraints is a strict subset of those of the previous optimal solution. If there is no such optimal solution, we consider that as having a minimal set of binding constraints. Lemmas 4-6 together imply that an optimal solution with a minimal set of binding constraints coordinates the offer set decision across the different resources. Letting $(\hat{\lambda}, \hat{w})$ be an optimal solution to $LP_{\Pi_t(v)}$ with a minimal number of binding constraints.

Lemma 4 *Fix a set S. Either*

    (i) *$B_{i,S}(\hat{\lambda}, \hat{w})$ is nonempty for all $i \in \mathcal{I}_S$ and $\hat{w}_{\phi,t} = \hat{\lambda}_{\phi,S,t}$, or*
    (ii) *$B_{i,S}(\lambda, \hat{w})$ is empty for all $i \in \mathcal{I}_S$ and $\hat{w}_{\phi,t} > \hat{\lambda}_{\phi,S,t}$.*

Proof Suppose that the statement of the lemma is false. First consider the case where $B_{i,S}(\hat{\lambda}, \hat{w})$ is nonempty but $B_{l,S}(\hat{\lambda}, \hat{w})$ is empty for $i, l \in \mathcal{I}_S$. Let $\epsilon = \min_{r \in \{1, \dots, r_l^1\}} \{\xi_{l,S,t}(r)\} > 0$. Let $(\widetilde{\lambda}, \widetilde{w})$ be given by $\widetilde{\lambda} = \hat{\lambda} - \delta e^{i,S,t} + \delta e^{l,S,t}$ and $\widetilde{w} = \hat{w}$ for some $\delta \in (0, \epsilon)$, where $e^{i,j,k}$ is a vector with a 1 in component (i, j, k) and zeroes everywhere else. Note that $(\widetilde{\lambda}, \widetilde{w})$ is identical to $(\hat{\lambda}, \hat{w})$ except that $\widetilde{\lambda}_{i,S,t} = \hat{\lambda}_{i,S,t} - \delta$ and $\widetilde{\lambda}_{l,S,t} = \hat{\lambda}_{l,S,t} + \delta$. We show that $(\widetilde{\lambda}, \widetilde{w})$ is an optimal solution with a strictly fewer number of binding constraints which gives us a contradiction.

Notice that we only need to check constraint (19) for resources i and l and offer set S as all other resources and offer sets continue to have the same λ's and w's as before. For resource i, since $B_{i,S}(\hat{\lambda}, \hat{w})$ is nonempty, there exists $r \in \{1, \dots, r_i^1\}$ such that $\hat{w}_{i,t} = \hat{\lambda}_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S)\psi_{i,t+1}(r) + \Delta_{i,t}(r)$. We have

$$\widetilde{w}_{i,t} = \hat{w}_{i,t} = \hat{\lambda}_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S)\psi_{i,t+1}(r) + \Delta_{i,t}(r) > \widetilde{\lambda}_{i,S,t}$$
$$- \sum_{j \in \mathcal{J}_i} P_j(S)\psi_{i,t+1}(r) + \Delta_{i,t}(r),$$

which means that the number of binding constraints (19) for resource i and offer set S decreases by at least one. For resource l and offer set S, $\widetilde{w}_{l,t} - [\widetilde{\lambda}_{l,S,t} - \sum_{j \in \mathcal{J}_l} P_j(S)\psi_{l,t+1}(r) + \Delta_{l,t}(r)] = \hat{w}_{l,t} - [\hat{\lambda}_{l,S,t} - \sum_{j \in \mathcal{J}_l} P_j(S)\psi_{l,t+1}(r) + \Delta_{l,t}(r)] - \delta > 0$, for all $r \in \{1, \dots, r_l^1\}$, where the inequality follows from the definition of δ. Therefore, all constraints (19) continue to be nonbinding for resource l and offer

set S. Overall, $(\widetilde{\lambda}, \widetilde{w})$ has strictly fewer binding constraints than $(\hat{\lambda}, \hat{w})$, which gives a contradiction.

The above arguments imply that either $B_{i,S}(\hat{\lambda}, \hat{w})$ is nonempty for all $i \in \mathcal{I}_S$ or $B_{i,S}(\hat{\lambda}, \hat{w})$ is empty for all $i \in \mathcal{I}_S$. Suppose the $B_{i,S}(\hat{\lambda}, \hat{w})$ is nonempty for all $i \in \mathcal{I}_S$ but $\hat{w}_{\phi,t} > \hat{\lambda}_{\phi,St}$. In this case, pick a resource $i \in \mathcal{I}_S$ and let $\widetilde{\lambda} = \hat{\lambda} - \delta e^{i,S,t} + \delta e^{\phi,S,t}$ and $\widetilde{w} = \hat{w}$, where $\delta \in (0, \epsilon)$ and $\epsilon = \hat{w}_{\phi,t} - \hat{\lambda}_{\phi,S,t}$. It can be verified that the number of binding constraints (19) for resource i and offer set S strictly decreases from $(\hat{\lambda}, \hat{w})$ to $(\widetilde{\lambda}, \widetilde{w})$, while the number of binding constraints (20) remains unchanged, leading to a contradiction. This proves part (i) of the lemma.

On the other hand, if $B_{i,S}(\hat{\lambda}, \hat{w})$ is empty for all $i \in \mathcal{I}_S$ but $\hat{w}_{\phi,t} = \hat{\lambda}_{\phi,S,t}$. Pick a resource $i \in \mathcal{I}_S$. Since $B_{i,S}(\hat{\lambda}, \hat{w})$ is empty, we have $\epsilon = \min_{r \in \{1,\dots,r_i^1\}}\{\xi_{i,S,t}(r)\} > 0$. Let $\widetilde{\lambda} = \hat{\lambda} + \delta e^{i,S,t} - \delta e^{\phi,S,t}$, where $\delta \in (0, \epsilon)$, and $\widetilde{w} = \hat{w}$. It can be verified that the number of binding constraints (19) for resource i and offer set S is exactly the same for $(\hat{\lambda}, \hat{w})$ and $(\widetilde{\lambda}, \widetilde{w})$, while the number of binding constraints (20) decreases by one $(\widetilde{w}_{\phi,t} = \hat{w}_{\phi,t} = \hat{\lambda}_{\phi,S,t} > \widetilde{\lambda}_{\phi,S,t})$. As a result, $(\widetilde{\lambda}, \widetilde{w})$ has strictly fewer binding constraints than $(\hat{\lambda}, \hat{w})$, which gives a contradiction. This proves part (ii) of the lemma. □

Let

$$\hat{C}_i = \{S \in C_i \mid B_{i,S}(\hat{\lambda}, \hat{w}) \text{ is nonempty}\},$$
$$\hat{C}_\phi = \{S \subset \mathcal{J} \mid \hat{w}_{\phi,t} = \hat{\lambda}_{\phi,S,t}\},$$

and

$$\hat{\mathcal{I}}^+ = \{i \mid \hat{w}_{i,t} > \Delta_{i,t}(0)\},$$

where the ˆ on the sets is to remind the reader of the dependence on $(\hat{\lambda}, \hat{w})$. Note that $\hat{C}_i$ can be interpreted as the collection of offer sets for which constraint (19) is binding for resource i. $\hat{C}_\phi$ has a similar interpretation, but for constraint (20).

Lemma 5 *Let $(\hat{\lambda}, \hat{w})$ be an optimal solution to $LP_{\Pi_t(v)}$ with a minimal number of binding constraints.*

(i) $\hat{C}_i$ *is nonempty for all $i \in \hat{\mathcal{I}}^+$.*

(ii) *If $S \in \hat{C}_i$, then $S \in \hat{C}_\phi$ (note that by definition the empty set does not consume any resources and so $\emptyset \notin \hat{C}_i$).*

Proof For part (i), for $i \in \hat{\mathcal{I}}^+$, $\hat{w}_{i,t} > \Delta_{i,t}(0)$. Since $(\hat{\lambda}, \hat{w})$ is optimal there exists $S \in C_i$ and $r \in \{1,\dots,r_i^1\}$ such that $\hat{w}_{i,t} = \hat{\lambda}_{i,S,t} - \sum_{j \in S} P_j(S)\psi_{i,t+1}(r) + \Delta_{i,t}(r) > \Delta_{i,t}(0)$ (otherwise, we can reduce $\hat{w}_{i,t}$ contradicting optimality). Therefore $B_{i,S}(\hat{\lambda}, \hat{w})$ is nonempty and so $S \in \hat{C}_i$ and $\hat{C}_i$ is nonempty.

For part (ii), $S \in \hat{C}_i$ implies that $i \in \mathcal{I}_S$. So we have a set S with $B_{i,S}(\hat{\lambda}, \hat{w})$ nonempty. By Lemma 4, $\hat{w}_{\phi,t} = \hat{\lambda}_{\phi,S,t}$ and so $S \in \hat{C}_\phi$. □

Lemma 6 shows that we can find an offer set for which constraint (19) is binding for all resources i. This turns out to be crucial for showing that $\Phi_t(v) \geq \Pi_t(v)$.

Lemma 6 *Let* $(\hat{\lambda}, \hat{w})$ *be an optimal solution to* $LP_{\Pi_t(v)}$ *with a minimal number of binding constraints. If* $\hat{\mathcal{I}}^+$ *is nonempty, then* $\cap_{i \in \hat{\mathcal{I}}^+} \hat{C}_i$ *is nonempty.*

Proof If $|\hat{\mathcal{I}}^+| = 1$, then the statement holds trivially by part (i) of Lemma 5. Consider the case $|\hat{\mathcal{I}}^+| > 1$. If $\cap_{i \in \hat{\mathcal{I}}^+} \hat{C}_i$ is empty, then this implies the following: Fix a resource $i \in \hat{\mathcal{I}}^+$. Part (i) of Lemma 5 implies that $\hat{C}_i$ is nonempty. Then for every $S \in \hat{C}_i$ there exists $l \in \hat{\mathcal{I}}^+$ such that $S \notin \hat{C}_l$. Note that since $l \in \hat{\mathcal{I}}^+$, $\hat{w}_{l,t} > \Delta_{l,t}(0)$.

So let $i \in \hat{\mathcal{I}}^+$, $\hat{S} \in \hat{C}_i$ and $l \in \hat{\mathcal{I}}^+$ with $\hat{S} \notin \hat{C}_l$. If $\hat{S} \notin \hat{C}_l$, there are two possibilities. First, $\hat{S} \in C_l$ but $B_{l,\hat{S}}(\hat{\lambda}, \hat{w})$ is empty. But since $\hat{S} \in \hat{C}_i$, $B_{i,\hat{S}}(\hat{\lambda}, \hat{w})$ is nonempty, which this contradicts part (i) of Lemma 4.

The other possibility is that $\hat{S} \notin C_l$. Let

$$
\epsilon = \min \left\{ \hat{w}_{l,t} - \Delta_{l,t}(0), \min_{S \in C_l \setminus \hat{C}_l, r \in \{1, \dots, r_l^1\}} \{\xi_{l,S,t}(r)\} \right\} > 0
$$

where we define the minimum over an empty set to be infinity (the second term could be empty). Let $\delta \in (0, \epsilon)$ and $(\widetilde{\lambda}, \widetilde{w})$ be given by $\widetilde{\lambda} = \hat{\lambda} - \delta \sum_{S \in \hat{C}_l} e^{l,S,t} + \delta \sum_{S \in \hat{C}_l} e^{\phi,S,t}$ and $\widetilde{w} = \hat{w} - \delta e^{l,t} + \delta e^{\phi,t}$. Therefore, we have $\widetilde{\lambda}_{l,S,t} = \hat{\lambda}_{l,S,t} - \delta$, $\widetilde{\lambda}_{\phi,S,t} = \hat{\lambda}_{\phi,S,t} + \delta$ for all $S \in \hat{C}_l$. Similarly, $\widetilde{w}_{l,t} = \hat{w}_{l,t} - \delta$ and $\widetilde{w}_{\phi,t} = \hat{w}_{\phi,t} + \delta$. All other terms remain the same.

We check that $(\widetilde{\lambda}, \widetilde{w})$ is feasible and look at the set of binding constraints associated with this solution. We look at the constraints in $LP_{\Pi_t(v)}$ one by one and compare the the number of binding constraints in $(\hat{\lambda}, \hat{w})$ with the number in $(\widetilde{\lambda}, \widetilde{w})$.

Constraints (21) *and* (22)*:* Since $\widetilde{w}_{\phi,t} > \hat{w}_{\phi,t}$, constraint (21) continues to hold for $(\widetilde{\lambda}, \widetilde{w})$ and the number of binding constraints do not increase. By construction $(\widetilde{\lambda}, \widetilde{w})$ satisfies constraint (22).

Constraints (18)*:* Note that we need to check constraints (18) and (19) only for resource l. For resource l, we have $\widetilde{w}_{l,t} = \hat{w}_{l,t} - \delta > \Delta_{l,t}(0)$ and so constraint (18) continues to be nonbinding.

Constraints (19)*:* For $S \in C_l \setminus \hat{C}_l$ and $r \in \{1, \dots, r_l^1\}$, we have $\widetilde{w}_{l,t} = \hat{w}_{l,t} - \delta > \hat{w}_{l,t} - \xi_{l,S,t}(r) = \hat{\lambda}_{l,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{l,t+1}(r) + \Delta_{l,t}(r) = \widetilde{\lambda}_{l,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{l,t+1}(r) + \Delta_{l,t}(r)$. Note that the last equality holds by definition of $\widetilde{\lambda}$. So constraint (19) remains nonbinding.

For $S \in \hat{C}_l$ and $r \in \{1, \dots, r_l^1\} \setminus B_{l,S}(\hat{\lambda}, \hat{w})$, $\widetilde{w}_{l,t} = \hat{w}_{l,t} - \delta > \hat{\lambda}_{l,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{l,t+1}(r) + \Delta_{l,t}(r) - \delta = \widetilde{\lambda}_{l,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{l,t+1}(r) + \Delta_{l,t}(r)$. Therefore, constraint (19) continues to be nonbinding. For $S \in \hat{C}_l$ and $r \in B_{l,S}(\hat{\lambda}, \hat{w})$, $\widetilde{w}_{l,t} = \hat{w}_{l,t} - \delta = \hat{\lambda}_{l,t} - \delta - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{l,t+1}(r) + \Delta_{l,t}(r) = \widetilde{\lambda}_{l,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{l,t+1}(r) + \Delta_{l,t}(r)$. So constraints (19) are binding for all such S and r. Note that these constraints, by definition, were also binding in $(\hat{\lambda}, \hat{w})$. So, $(\widetilde{\lambda}, \widetilde{w})$ satisfies constraints (18) and (19) for resource l and the number of binding constraints is exactly the same as in $(\hat{\lambda}, \hat{w})$.

Constraint (20)*:* For $S \in \hat{C}_l$, by definition $B_{l,S}(\hat{\lambda}, \hat{w})$ is nonempty. Part (i) of Lemma 4 implies that $\hat{w}_{\phi,t} = \hat{\lambda}_{\phi,S,t}$, which means that constraint (20) is binding. We have $\widetilde{w}_{\phi,t} = \hat{w}_{\phi,t} + \delta = \hat{\lambda}_{\phi,St} + \delta = \widetilde{\lambda}_{\phi,S,t}$ and so constraint (20) holds and continues to be binding. For $S \notin \hat{C}_l$, $\widetilde{\lambda}_{l,S,t} = \hat{\lambda}_{l,S,t}$. Therefore, $\widetilde{w}_{\phi,t} = \hat{w}_{\phi,t} + \delta \geq \widetilde{\lambda}_{l,S,t}$. So constraint (20) holds and the number of binding constraints do not increase.

Now we argue that the number of binding constraints (20) strictly decreases from $(\hat{\lambda}, \hat{w})$ to $(\widetilde{\lambda}, \widetilde{w})$. For the set $\hat{S}$, since $\hat{S} \in \hat{C}_i$, $B_{i,\hat{S}}(\hat{\lambda}, \hat{w})$ is nonempty. By, part (i) of Lemma 4, $\hat{w}_{\phi,t} = \hat{\lambda}_{\phi,\hat{S},t}$ and so the constraint is binding in $(\hat{\lambda}, \hat{w})$. But $\widetilde{w}_{\phi,t} = \hat{w}_{\phi,t} + \delta > \hat{\lambda}_{\phi,\hat{S},t} = \widetilde{\lambda}_{\phi,\hat{S},t}$ and the constraint is nonbinding in $(\widetilde{\lambda}, \widetilde{w})$. Overall, $(\widetilde{\lambda}, \widetilde{w})$ has strictly fewer number of binding constraints (20) and they are a subset of the set of binding constraints of $(\hat{\lambda}, \hat{w})$ contradicting minimality.

Since $\hat{w}_{\phi,t} + \sum_i \hat{w}_{i,t} = \widetilde{w}_{\phi,t} + \sum_i \widetilde{w}_{i,t}$, $(\widetilde{\lambda}, \widetilde{w})$ is optimal and this gives a contradiction. $\qquad\square$

We are now ready to show that $\Phi_t(v) \geq \Pi_t(v)$.

Proposition 3 $\Phi_t(v) \geq \Pi_t(v)$.

Proof Let $(\hat{\lambda}, \hat{w})$ be an optimal solution to $LP_{\Pi_t(v)}$ with a minimal number of binding constraints. We consider two cases.

Case 1: Suppose that $\hat{C}_i$ is empty for all i. This means that for all $S \in C_i$, $B_{i,S}(\hat{\lambda}, \hat{w})$ is empty. It follows that $\hat{w}_{i,t} = \Delta_{i,t}(0)$ for all i (otherwise we can reduce $\hat{w}_{i,t}$ contradicting optimality). Part (ii) of Lemma 4 implies that $\hat{w}_{\phi,t} > \hat{\lambda}_{\phi,S,t}$ for all S. It follows that $\hat{w}_{\phi,t} = 0$. Therefore, $\Pi_t(v) = \sum_i \Delta_{i,t}(0)$. Note that $r = 0$ and $S = \emptyset$ is feasible for $\Phi_t(v)$ and the objective function value associated with this solution is $\sum_i \Delta_{i,t}(0)$. Therefore $\Phi_t(v) \geq \sum_i \Delta_{i,t}(0) = \Pi_t(v)$.

Case 2: Suppose that $\hat{C}_i$ is nonempty for some i. We consider two subcases.

Case 2.a $\hat{I}^+$ is empty. We choose a resource l such that $\hat{C}_l$ is nonempty and choose $\hat{S} \in \hat{C}_l$ such that $B_{l,\hat{S}}(\hat{\lambda}, \hat{w})$ is nonempty. By part (i) of Lemma 4, $B_{i,\hat{S}}(\hat{\lambda}, \hat{w})$ is nonempty for all $i \in \mathcal{I}_{\hat{S}}$ and $\hat{w}_{\phi,t} = \hat{\lambda}_{\phi,\hat{S},t}$. So, for all $i \in \mathcal{I}_{\hat{S}}$, we have $\hat{w}_{i,t} = \hat{\lambda}_{i,\hat{S},t} - \sum_{j \in \mathcal{J}_i} P_j(\hat{S})\psi_{i,t+1}(\hat{r}_i) + \Delta_{i,t}(\hat{r}_i)$, where $\hat{r}_i \in B_{i,\hat{S}}(\hat{\lambda}, \hat{w})$. Note that $\hat{r}_i \geq 1$ for all $i \in \mathcal{I}_{\hat{S}}$. On the other hand, since $\hat{I}^+$ is empty. we have $\hat{w}_{i,t} = \Delta_{i,t}(0)$ for all i. In particular, $\hat{w}_{i,t} = \Delta_{i,t}(0)$ for all $i \notin \mathcal{I}_S$. Putting everything together,

$$
\begin{aligned}
\Pi_t(v) &= \hat{w}_{\phi,t} + \sum_{i \in \mathcal{I}_{\hat{S}}} \hat{w}_{i,t} + \sum_{i \notin \mathcal{I}_{\hat{S}}} \hat{w}_{i,t} \\
&= \hat{\lambda}_{\phi,\hat{S},t} + \sum_{i \in \mathcal{I}_{\hat{S}}} \left[\hat{\lambda}_{i,\hat{S},t} - \sum_{j \in \mathcal{J}_i} P_j(\hat{S})\psi_{i,t+1}(\hat{r}_i) + \Delta_{i,t}(\hat{r}_i) \right] + \sum_{i \notin \mathcal{I}_{\hat{S}}} \Delta_{i,t}(0) \\
&= R(\hat{S}) - \sum_{j \in S} \sum_{i \in \mathcal{I}_j} \mathbb{1}_{[\hat{S} \in C_i]} P_j(\hat{S})\psi_{i,t+1}(\hat{r}_i) + \sum_{i \in \mathcal{I}_{\hat{S}}} \Delta_{i,t}(\hat{r}_i) + \sum_{i \notin \mathcal{I}_{\hat{S}}} \Delta_{i,t}(0) \\
&\leq \Phi_t(v)
\end{aligned}
$$

where the last equality follows since $(\hat{\lambda}, \hat{w})$ satisfies constraint (22) and

$$
\begin{aligned}
\sum_{i \in \mathcal{I}_S} \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) &= \sum_{i} \sum_{j \in \mathcal{J}_i} \mathbb{1}_{[i \in \mathcal{I}_S]} P_j(S) \psi_{i,t+1}(r_i) \\
&= \sum_{j} \sum_{i \in \mathcal{I}_j} \mathbb{1}_{[i \in \mathcal{I}_S]} P_j(S) \psi_{i,t+1}(r_i) \\
&= \sum_{j \in S} \sum_{i \in \mathcal{I}_j} \mathbb{1}_{[S \in C_i]} P_j(S) \psi_{i,t+1}(r_i).
\end{aligned}
$$

The inequality follows from (17) by noting that $\hat{S}$ and $\boldsymbol{r}$, where $r_i = \hat{r}_i$ for $i \in \mathcal{I}_{\hat{S}}$ and $r_i = 0$ otherwise, is feasible to $\Phi_t(v)$. To see this, observe that for all $j \in \hat{S}$, if $\mathbb{1}_{[i \in \mathcal{I}_j]} = 1$, then $i \in \mathcal{I}_{\hat{S}}$ and so $r_i = \hat{r}_i \geq 1$. Therefore, for all $j \in \hat{S}$, $\mathbb{1}_{[i \in \mathcal{I}_j]} \leq r_i$ for all i and so $\hat{S} \subset \mathcal{Q}(\boldsymbol{r})$.

Case 2.b $\hat{\mathcal{I}}^+$ is nonempty. Let $\hat{S} = \cap_{i \in \hat{\mathcal{I}}^+} \hat{C}_i$, which by Lemma 6 is nonempty. Note that $\hat{S} \in \hat{C}_i \subset C_i$ for all $i \in \hat{\mathcal{I}}^+$. Now every $i \in \hat{\mathcal{I}}^+$ satisfies $\hat{w}_{i,t} > \Delta_{i,t}(0)$. Therefore if $i \notin \mathcal{I}_{\hat{S}}$ then $\hat{w}_{i,t} = \Delta_{i,t}(0)$. Since $\hat{S} \in \hat{C}_l$ for some l, $B_{l\hat{S}}(\hat{\lambda}, \hat{w})$ is nonempty. Part (i) of Lemma 4 implies that $B_{i,\hat{S}}(\hat{\lambda}, \hat{w})$ is nonempty for all $i \in \mathcal{I}_{\hat{S}}$ and that $\hat{w}_{\phi,t} = \hat{\lambda}_{\phi,\hat{S},t}$. Since $B_{i,\hat{S}}(\hat{\lambda}, \hat{w})$ is nonempty for $i \in \mathcal{I}_{\hat{S}}$, there exists $\hat{r}_i \in \{1, \ldots, r_i^1\}$ such that $\hat{w}_{i,t} = \hat{\lambda}_{i,\hat{S},t} - \sum_{j \in \mathcal{J}_i} P_j(\hat{S}) \psi_{i,t+1}(\hat{r}_i) + \Delta_{i,t}(\hat{r}_i)$ ($\hat{r}_i$ need not be unique, we can pick any $r \in B_{i,\hat{S}}(\hat{\lambda}, \hat{w})$). We have

$$
\begin{aligned}
\Pi_t(v) &= \hat{\lambda}_{\phi,\hat{S},t} + \sum_{i \in \mathcal{I}_{\hat{S}}} \left[\hat{\lambda}_{i,\hat{S},t} - \sum_{j \in \mathcal{J}_i} P_j(\hat{S}) \psi_{i,t+1}(\hat{r}_i) + \Delta_{i,t}(\hat{r}_i) \right] + \sum_{i \notin \mathcal{I}_{\hat{S}}} \Delta_{i,t}(0) \\
&= R(\hat{S}) - \sum_{j \in \hat{S}} P_j(\hat{S}) \sum_{i \in \mathcal{I}_j} \mathbb{1}_{[\hat{S} \in C_i]} \psi_{i,t+1}(\hat{r}_i) + \sum_{i \in \mathcal{I}_{\hat{S}}} \Delta_{i,t}(\hat{r}_i) + \sum_{i \notin \mathcal{I}_{\hat{S}}} \Delta_{i,t}(0) \\
&\leq \Phi_t(v)
\end{aligned}
$$

where the inequality follows from (17) by noting that $\hat{S}$ and $\boldsymbol{r}$ is feasible to $\Phi_t(v)$ where $r_i = \hat{r}_i \geq 1$ for $i \in \mathcal{I}_{\hat{S}}$ and $r_i = 0$ otherwise. $\qquad\square$

Lemma 2 and Proposition 3 together imply that $\Phi_t(v) = \Pi_t(v)$. We are now ready to show that $V^{PL} \geq V^{LRo}$. Note that we can write *PL* as $V^{PL} = \min_v \sum_i v_{i,1}(r_i^1)$ subject to $0 \geq \Phi_t(v)$ for all t with the boundary condition that $v_{i,\tau+1}(\cdot) = 0$. Letting $\underline{V} = \min_v \sum_i v_{i,1}(r_i^1) + \sum_t \Phi_t(v)$ subject to $0 \geq \Phi_t(v)$ for all t with the same boundary condition, it follows that $V^{PL} \geq \underline{V}$. Using the fact that $\Phi_t(v) = \Pi_t(v)$ together with the linear programming formulation $LP_{\Pi_t(v)}$ of $\Pi_t(v)$, we can write $\underline{V}$ as

$$\underline{V} = \min_{v} \quad \sum_i v_{i,1}(r_i^1) + \sum_t \left(w_{\phi,t} + \sum_i w_{i,t} \right)$$

s.t.

$$w_{i,t} + v_{i,t}(r_i)$$

$$\geq \max_{S \subset Q_i(r_i)} \left\{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} + \sum_{j \in \mathcal{J}_i} P_j(S) \left[v_{i,t+1}(r_i - 1) - v_{i,t+1}(r_i) \right] + v_{i,t+1}(r_i) \right\}$$

$$\forall t, r_i \in \mathcal{R}_i$$

$$w_{\phi,t} \geq \max_{S \subset \mathcal{J}} \left\{ \lambda_{\phi,S,t} \right\} \quad \forall t$$

$$w_{\phi,t} \geq 0 \quad \forall t$$

$$\lambda_{\phi,S,t} + \sum_{i \in \mathcal{I}_S} \lambda_{i,S,t} = R(S) \quad \forall t, S \subset \mathcal{J}$$

$$v_{i,\tau+1}(\cdot) = 0 \quad \forall i.$$

With the change of variables $\vartheta_{i,t}(r_i) = v_{i,t}(r_i) + \sum_{s=t}^{\tau} w_{i,s}$ and $\vartheta_{\phi,t} = \sum_{s=t}^{\tau} w_{\phi,s}$, we can write the above linear program as

$$\underline{V} = \min_{\vartheta} \quad \sum_i \vartheta_{i,1}(r_i^1) + \vartheta_{\phi 1}$$

s.t.

$$\vartheta_{i,t}(r_i) \geq \max_{S \subset Q_i(r_i)} \left\{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} + \sum_{j \in \mathcal{J}_i} P_j(S)[\vartheta_{i,t+1}(r_i - 1) - \vartheta_{i,t+1}(r_i)] + \vartheta_{i,t+1}(r_i) \right\}$$

$$\forall t, r_i \in \mathcal{R}_i$$

$$\vartheta_{\phi,t} \geq \max_{S \subset \mathcal{J}} \left\{ \lambda_{\phi,S,t} \right\} + \vartheta_{\phi,t+1} \quad \forall t$$

$$\lambda_{\phi,S,t} + \sum_{i \in \mathcal{I}_S} \lambda_{i,S,t} = R(S) \quad \forall t, S \subset \mathcal{J}$$

$$\vartheta_{i,\tau+1}(\cdot) = 0 \quad \forall i$$

which is exactly *LRo*. Therefore, $V^{PL} \geq V^{LRo}$. $\square$

6 Segment-based Lagrangian relaxation

Although the number of constraints in *LRo* scales linearly with the resource capacities, it is still exponential in the number of products. In this section, we present a tractable approximation that applies to settings where the total demand is comprised of multiple customer segments with small consideration sets. We first describe the demand model and then present the tractable variant of *LRo*, which we refer to as segment-based Lagrangian relaxation (*sLRo*).

6.1 Demand model with multiple customer segments

We consider the case where the total demand is comprised of demand from multiple customer segments. Modeling demand in this manner can be a way to capture heterogenous customer preferences. Moreover, each customer segment is interested only in a small subset of the products (its consideration set). Small consideration sets are a reasonable modeling tradeoff in many situations that also finds support in the empirical and behavioral literature. We let $\mathcal{G}$ denote the set of customer segments. Customer segment $g \in \mathcal{G}$ has a *consideration set* $\mathcal{J}^g \subset \mathcal{J}$ of products that it considers for purchase. A segment g customer is indifferent to a product outside its consideration set, in the sense that the customer's choice probabilities are not affected by products offered outside its consideration set. We assume that the consideration sets of the different customer segments are known to the firm by a previous process of estimation and analysis. We also assume that the consideration sets of the different segments are small enough for its power set to be enumerable.

In each period, we have exactly one customer arrival and an arriving customer belongs to segment g with probability α^g. Since the total arrival rate is 1, we have $\sum_g \alpha^g = 1$. We let $P_j^g(S)$ denote the probability that a segment g arrival purchases product j when S is the offer set. Since a segment g customer is indifferent to products outside its consideration set, we have $P_j^g(S) = P_j^g(S \cap \mathcal{J}^g) = P_j^g(S^g)$, where $S^g = S \cap \mathcal{J}^g$.

6.2 Tractable Lagrangian relaxation

We modify the Lagrangian relaxation approach for the case with multiple customer segments in the following manner: We introduce segment-based decision variables $h_{i,S^g,t}^g$, which denote the frequency with which set $S^g \subset \mathcal{J}^g$ is offered to segment g on resource i at time period t. As in *LRo*, we relax the condition that the frequencies be identical across the resources by associating Lagrange multipliers $\lambda_{i,S^g,t}^g$ with them and solve the optimality equation

$$\vartheta_{i,t}^\lambda(r_i) = \max_{h_i} \sum_{g \in \mathcal{G}_i} \alpha^g \left\{ \sum_{S^g \in C_i} h_{i,S^g,t}^g \left(\lambda_{i,S^g,t}^g + \sum_{j \in \mathcal{J}_i} P_j^g(S^g) \left[\vartheta_{i,t+1}^\lambda(r_i - 1) - \vartheta_{i,t+1}^\lambda(r_i) \right] \right) \right\}$$
$$+ \vartheta_{i,t+1}^\lambda(r_i)$$

s.t.
$$h_{i,S^g,t}^g \le r_i \, \forall g \in \mathcal{G}_i, S^g \subset \mathcal{J}^g, S^g \in C_i$$
$$\sum_{S^g \in C_i} h_{i,S^g,t}^g \le 1 \, \forall g \in \mathcal{G}_i$$
$$h_{i,S^g,t}^g \ge 0 \, \forall g \in \mathcal{G}_i, S^g \subset \mathcal{J}^g, S^g \in C_i$$

for resource i, with the boundary condition that $\vartheta_{i,\tau+1}^\lambda(\cdot) = 0$, where $\mathcal{G}_i = \{g | \exists j \in \mathcal{J}^g \text{ with } j \in \mathcal{J}_i\}$ can be interpreted as the set of customer segments that use resource i. We modify optimality equation (11) in a similar manner for the case with multiple segments:

$$\vartheta_{\phi,t}^{\lambda} = \max_{h_\phi} \sum_g \alpha^g \left\{ \sum_{S^g} h_{\phi,S^g,t}^g \left[\sum_{j \in S^g} f_j P_j^g(S^g) - \sum_{i \in \mathcal{I}_{S^g}} \lambda_{i,S^g,t}^g \right] \right\} + \vartheta_{\phi,t+1}^{\lambda}$$

s.t.

$$\sum_{S^g} h_{\phi,S^g,t}^g \leq 1 \,\forall g$$
$$h_{\phi,S^g,t}^g \geq 0 \,\forall g, S^g \subset \mathcal{J}^g$$

with the boundary condition that $\vartheta_{\phi,\tau+1}^{\lambda} = 0$.

We refer to the above described modification as the segment-based Lagrangian relaxation (*sLRo*). Many of the results from Sect. 4 carry over to *sLRo*. In particular, $V_t^{\lambda}(\boldsymbol{r}) = \sum_i \vartheta_{i,t}^{\lambda}(r_i) + \vartheta_{\phi,t}^{\lambda}$ gives us an upper bound on the value function. We find the tightest upper bound by solving

$$V^{sLRo} = \min_{\lambda} V_1^{\lambda}(\boldsymbol{r}^1). \tag{23}$$

$V_t^{\lambda}(\boldsymbol{r})$ is a convex function of λ and an expression for its subgradient can be derived in a manner analogous to Proposition 1; we omit the details. Therefore, we can solve minimization problem (23) using subgradient search. However, *sLRo* can be weaker than *LRo* and we can have $V^{LRo} < V^{sLRo}$. Still, an advantage of *sLRo* is its computational tractability. Note that solving *LRo* requires $\mathcal{O}(2^{|\mathcal{J}|})$ Lagrange multipliers which quickly becomes intractable. On the other hand, *sLRo* requires $\mathcal{O}(\sum_g 2^{|\mathcal{J}^g|})$ Lagrange multipliers, a much more manageable number provided the consideration set for each segment is small in size. Moreover, when the consideration sets of the segments overlap, it is possible to further tighten *sLRo* by adding the product-cut equalities (PC-equalities) described in Meissner et al. [17] and Talluri [21]; we omit the details.

7 Computational experiments

In this section, we compare the upper bound and revenue performance of *sLRo* with benchmark solution methods. We consider two choice models in our numerical experiments. The first one is the MNL choice model and the second is the exponomial choice model. We begin by describing the two choice models and then describe the different benchmark solution methods and the experimental setup.

7.1 MNL choice model with multiple customer segments

The total demand is comprised of multiple customer segments and within each segment choice is according to the MNL model. The MNL model associates a preference weight ω_j^g with product j that is in the consideration set of segment g. Similarly

it associates a preference weight ω_0^g with a segment g arrival not purchasing anything. The probability that a segment g arrival purchases product j when S is the offer set is (Ben-Akiva and Lerman [3])

$$P_j^g(S) = \frac{\omega_j^g}{\omega_0^g + \sum_{k \in S \cap \mathcal{J}^g} \omega_k^g},$$

while the probability it purchases none of the offered products is

$$P_0^g(S) = \frac{\omega_0^g}{\omega_0^g + \sum_{k \in S \cap \mathcal{J}^g} \omega_k^g}.$$

In our computational experiments, we have $\mathcal{J}^g \cap \mathcal{J}^{g'} = \emptyset$ so that the consideration sets of the segments are disjoint.

7.2 Exponomial choice model with multiple customer segments

We again have that the total demand is comprised of multiple customer segments. Each segment has a preference weight (ideal utility) associated with purchasing a product that is in its consideration set and a preference weight associated with not purchasing anything. However, each segment's choice is governed by the exponomial choice model (Alptekinoğlu and Semple [2]). That is, if the preference weights are ordered so that $\omega_0^g \leq \omega_1^g \leq \ldots \omega_{n_l}^g$, where n_g is the number of products in the consideration set of segment g that are part of the offer set, then the probability that a segment g arrival purchases product j when S is the offer set is (Alptekinoğlu and Semple [2])

$$P_j^g(S) = \frac{\exp\left(-\sum_{k=j}^{n_g}[\omega_k^g - \omega_j^g]\right)}{n_g - j + 1} - \sum_{q=1}^{j-1} \frac{\exp\left(-\sum_{k=q}^{n_g}[\omega_k^g - \omega_q^g]\right)}{(n_g - q)(n_g - q + 1)}.$$

In our computational experiments, we have $\mathcal{J}^g \cap \mathcal{J}^{g'} = \emptyset$ so that the consideration sets of the segments are disjoint.

7.3 Benchmark methods

Choice Deterministic Linear Program (CDLP) This is the solution method described in Sect. 3.1. If customer choice is according to the MNL model and the consideration sets are disjoint, then *CDLP* has an equivalent, compact sales based formulation that is described in Gallego et al. [8]. Therefor, we solve *CDLP* as a compact linear program. On the other hand, if customer choice is according to the exponomial model, to our knowledge there is no compact formulation of *CDLP*. Since *CDLP* has a large number of variables, we use column generation to solve *CDLP*. To our knowledge, there is no efficient method to solve the column generation sub-problem

Fig. 1 Structure of the airline network with a single hub and six spokes
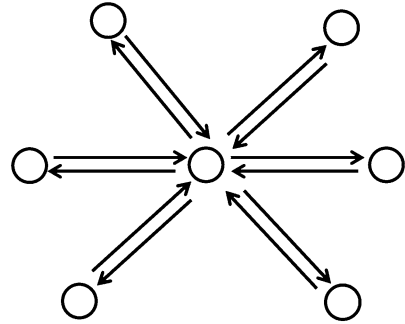
under the exponomial choice model and so we enumerate over the offer sets to solve the column generation sub-problem. The disjoint nature of the consideration sets makes the *CDLP* column generation sub-problem separable by the customer segments and this makes the enumeration procedure manageable. We stop the column generation procedure when we are within 1% of optimality. We obtain a bound on the optimality gap by using the optimal dual variables from the restricted master problem (see for example, Proposition 3 in Adelman [1]).

Affine Approximation (AF) This is the solution method described in Sect. 3.2. We use the reduced formulation described in Vossen and Zhang [25] to solve *AF*. While the number of variables in the reduced formulation is manageable, it still has a large number of constraints. We solve *AF* by generating constraints on the fly and stop when we are within 1% of optimality (Adelman [1]). For the MNL choice model, the separation problem of *AF* can be solved as a mixed-integer linear program (Zhang and Adelman [26]). However, we are not aware of a similar formulation for the exponomial choice model. Solving the separation problem by brute force involves enumerating over all possible offer sets *and* resource levels and becomes intractable.

Segment-based Lagrangian Relaxation (sLRo) This is the solution method described in Sect. 6. In our computational experiments, we use subgradient search to solve problem (23). We use a step size of $250/\sqrt{k}$ at iteration k of the subgradient algorithm and run the algorithm for 200 iterations. Although, our step size selection does not guarantee convergence (Bertsekas [4], Chapter 7), it provided good solutions and stable performance in our test problems.

7.4 Hub-and-spoke network

We have a network with a single hub serving N spokes. There is one flight from each spoke to the hub and one flight from the hub to each spoke, so that there are $2N$ flights in total. Figure 1 shows the structure of the network with $N = 6$. Note that the flight legs correspond to the resources in our DC-NRM formulation.

The total number of fare-products is $2N(N + 1)$. There are $4N$ fare-products connecting hub-to-spoke and spoke-to-hub origin-destination pairs. Of these, half are high fare-products whose revenues are drawn from the Poisson distribution with a mean of 30, while the remaining are low fare-products whose revenues are drawn from the Poisson distribution with a mean of 10. There are $2N(N - 1)$ fare products

connecting spoke-to-spoke origin-destination pairs. Half of them are high fare-products whose revenues are drawn from the Poisson distribution with a mean of 300, while the remaining are low fare-products whose revenues are drawn from the Poisson distribution with a mean of 100.

Each origin-destination pair is associated with a customer segment and a segment is only interested in the fare-products connecting its origin-destination pair. For the MNL choice model, the preference weights of the high fare-products are drawn from the Poisson distribution with a mean of 200, while that of the low fare-products are drawn from the Poisson distribution with a mean of 80. The no-purchase preference weights are drawn from the Poisson distribution with a mean of 10. We remark that we set the problem parameters in the same manner as Meissner and Strauss [16]. For the exponomial choice model, the preference weight (ideal utility) of each product is drawn uniformly from the interval $[-4, 18]$, while the preference weight for the no-purchase option is set to 1. We follow Alptekinoğlu and Semple [2] in setting the parameters of the exponomial choice model.

We measure the tightness of the leg capacities using the nominal load factor. Letting

$$S^{g*} \in \text{argmax}_{S^g \subset \mathcal{J}^g} \sum_{j \in S^g} P_j(S^g) f_j$$

be the offer set that maximizes expected revenues from segment g when there is ample capacity on all the flight legs, the nominal load factor is

$$\zeta = \frac{\sum_t \sum_i \left[\sum_g \alpha^g [\sum_{j \in \mathcal{J}_i} P_j^g(S^{g*})] \right]}{\sum_i r_i^1}.$$

We have $\tau = 400$ time periods in all of our test problems. We label our test problems by the pair (N, ζ) where $N \in \{10, 12, 14\}$ and $\zeta \in \{1.0, 1.2\}$, which gives us a total of six test problems for each choice model.

Table 1 Comparison of the upper bounds on the optimal expected total revenue and the CPU times for the test problems with the MNL choice model

Problem	Upper bound			% Gap with sLRo		CPU s		
(N, ζ)	CDLP	AF	sLRo	CDLP	AF	CDLP	AF	sLRo
(10, 1.0)	30,296	30,025	29,495	2.72	1.80	0.60	655.67	302.41
(10, 1.2)	29,345	29,069	28,265	3.82	2.84	0.58	954.98	259.06
(12, 1.0)	29,863	29,543	28,988	3.02	1.91	0.97	1004.97	381.74
(12, 1.2)	29,109	28,754	27,801	4.71	3.43	0.96	1707.46	320.48
(14, 1.0)	29,820	29,417	28,803	3.53	2.13	1.49	1800.82	467.31
(14, 1.2)	28,946	28,468	27,442	5.48	3.74	1.49	3081.48	389.43
			Avg.	3.88	2.64			

Table 2 Comparison of the expected revenues for the test problems with the MNL choice model

Problem	Expected revenue			% Gap with sLRo	
(N, ζ)	CDLP	AF	sLRo	CDLP	AF
(10, 1.0)	28,727	28,830	29,390	2.26 ✓	1.91 ✓
(10, 1.2)	27,434	27,495	27,974	1.93 ✓	1.71 ✓
(12, 1.0)	28,196	28,405	28,803	2.11 ✓	1.38 ✓
(12, 1.2)	26,796	27,337	27,432	2.32 ✓	0.35 ⊙
(14, 1.0)	28,161	28,527	28,551	1.37 ✓	0.09 ⊙
(14, 1.2)	26,441	26,605	26,881	1.64 ✓	1.03 ✓
			Avg.	1.94	1.08

We first describe the results for the test problems where choice is according to the MNL model. Table 1 gives the upper bounds obtained by the different solution methods along with their run times. The first column in the table describes the problem characteristics by using (N, ζ). The second to fourth columns, respectively, give the upper bounds obtained by *CDLP, AF* and *sLRo*. The next two columns, respectively, give the percentage gap between the upper bounds obtained by *CDLP* and *AF* with respect to *sLRo*. The last three columns give the CPU seconds required by *CDLP*, *AF* and *sLRo*, respectively. All of our computational experiments are carried out on a Core i7 desktop with 3.4-Ghz CPU and 16-GB RAM. We use CPLEX 12.2 to solve the *CDLP* and *AF* linear programs. We implement the subgradient search algorithm to solve *sLRo* in C++. We see that *sLRo* obtains significantly tighter upper bounds than *CDLP* and *AF*. The average gap between the upper bounds obtained by *CDLP* and *sLRo* is around 4%, while that between *AF* and *sLRo* is around 3%. In terms of computation time, *CDLP* has a compact linear programming representation for the MNL model and so it solves in a fraction of a second. The running times of *AF* and *sLRo* are in minutes. However *sLRo* solves much faster than *AF* and the differences in run times are more noticeable as the size of the network, as measured by the number of spokes, increases.

We next compare the revenue performance of the three solution methods. We evaluate revenues through simulation and use common random numbers in our simulations. All three solution methods obtain value function approximations of the

Table 3 Comparison of the upper bounds on the optimal expected total revenue and the CPU times for the test problems with the exponomial choice model

Problem	Upper bound		% Gap with sLRo	CPU s	
(N, ζ)	CDLP	sLRo	CDLP	CDLP	sLRo
(10, 1.0)	18,966	18,102	4.77	11.68	245.66
(10, 1.2)	18,241	17,283	5.54	15.78	210.11
(12, 1.0)	18,024	17,246	4.51	30.04	277.27
(12, 1.2)	17,395	16,526	5.26	38.23	245.68
(14, 1.0)	20,613	20,001	3.06	35.30	330.26
(14, 1.2)	19,930	19,001	4.89	57.93	292.71
		Avg.	4.67		

Table 4 Comparison of the expected revenues for the test problems with the exponential choice model

Problem	Expected revenue		% Gap with sLRo
(N, ζ)	CDLP	sLRo	CDLP
(10, 1.0)	18,932	19,756	4.17 ✓
(10, 1.2)	17,530	18,636	5.93 ✓
(12, 1.0)	16,031	17,086	6.18 ✓
(12, 1.2)	15,079	16,332	7.67 ✓
(14, 1.0)	17,238	18,096	4.74 ✓
(14, 1.2)	15,847	17,178	7.75 ✓
		Avg.	6.07

form $\widetilde{V}_t(\boldsymbol{r})$ that we use to make the control decisions. In the case of *CDLP*, we use $\widetilde{V}_t(\boldsymbol{r}) = \sum_i \hat{\mu}_i r_i$, where $\hat{\mu} = \{\hat{\mu}_i | \forall i\}$ are the optimal values of the dual variables associated with constraints (3). For *AF*, we use $\widetilde{V}_t(\boldsymbol{r}) = \hat{\theta}_t + \sum_i \hat{V}_{i,t} r_i$, where $\hat{\theta} = \{\hat{\theta}_t | \forall t\}$, $\hat{V} = \{\hat{V}_{i,t} | \forall i, t\}$ is an optimal solution to *AF*. For *sLRo*, we use $\widetilde{V}_t(\boldsymbol{r}) = \sum_i \vartheta_{i,t}^{\hat{\lambda}}(r_i) + \vartheta_{\phi,t}^{\hat{\lambda}}$, where $\hat{\lambda}$ is an optimal solution to problem (23). For each solution method, we use the corresponding value function approximation $\widetilde{V}_t(\boldsymbol{r})$, and if $\boldsymbol{r}^t$ is the vector of remaining resource capacities at time period t, we use the policy of offering the set that attains the maximum in the optimization problem

$$\max_{S \subset \mathcal{Q}(\boldsymbol{r}^t)} \left\{ \sum_{j \in S} P_j(S) \left[f_j + \widetilde{V}_{t+1}\left(\boldsymbol{r}^t - \sum_{i \in \mathcal{I}_j} \boldsymbol{e}^i\right) \right] + P_0(S) \widetilde{V}_{t+1}\left(\boldsymbol{r}^t\right) \right\}.$$

We refresh the respective value function approximations periodically by reoptimizing the solution methods at five equally spaced intervals over the booking horizon. Table 2 reports the expected revenues obtained by the three solution methods. The columns have a similar interpretation as in Table 1, except that they give revenues. In the last two columns, we use ✓ to indicate that *sLRo* generates higher revenues than the corresponding benchmark method at the 95% level and ⊙ if the revenue differences are not statistically significant at the 95% level. *sLRo* on average generates revenues that are about 2% higher than *CDLP* and about 1% higher than *AF*.

Now we describe the results for the test problems where choice is according to the exponential model. Table 3 gives the upper bounds obtained by *CDLP* and *sLRo* along with their run times. As mentioned, solving the *AF* separation problem becomes intractable for the exponential choice model and so we do not have *AF* as a benchmark solution method in the table. We see that *sLRo* continues to obtain significantly tighter upper bounds than *CDLP* and the *sLRo* bound is on average about 5% tighter than the *CDLP* bound. *sLRo* is agnostic to the form of the choice model and its solution time does not change very much compared to the MNL case. However, for the exponential choice model, *CDLP* has to be solved using column generation and consequently the solution times are larger compared to the MNL case. Still, *CDLP* solves in under a minute even

for the exponomial choice model. Table 4 gives the expected revenues obtained by *CDLP* and *sLRo*. The results are similar to the MNL case and we see that *sLRo* obtains significantly higher revenues than *CDLP*, with an average revenue improvement of around 6%.

8 Conclusions

In this paper we develop a new Lagrangian relaxation approach for DC-NRM with strong theoretical properties and numerical performance. We show that the Lagrangian relaxation equals the piecewise-linear approximation with the number of constraints scaling linearly with the resource capacities, compared to the exponential number when solving the piecewise-linear approximation directly. We build on these ideas and proposed a segment-based relaxation that is tractable when the consideration sets of the different customer segments are small. Our numerical experiments show that the proposed approach can provide significant benefits, both in terms of tighter upper bounds and higher expected revenues. Finally, we note that our results apply at the highest level of generality, for any discrete-choice model of customer demand behavior. An interesting future research direction is to see if more tractable methods can be obtained by specializing to specific choice models such as MNL or nested logit.

Appendix

Proof of Lemma 1

The first inequality follows from Meissner and Strauss [16]. So, we only show the second inequality. Let

$$\bar{v}_{i,t}(r_i) = \vartheta_{i,t}^{\lambda}(r_i) + \vartheta_{\phi,t}^{\lambda}/|\mathcal{I}| \ \forall t, i, r_i \in \mathcal{R}_i \tag{24}$$

where $|\mathcal{I}|$ is the total number of resources. Fix t, $\boldsymbol{r}$ and $\hat{S} \subset \mathcal{Q}(\boldsymbol{r})$. Since $\hat{S} \subset \mathcal{Q}(\boldsymbol{r})$, $\mathbb{1}_{[j \in \mathcal{J}_i]} \leq r_i$ for all $j \in \hat{S}$, which implies that $\mathbb{1}_{[i \in \mathcal{I}_{\hat{S}}]} = \max_{j \in \hat{S}} \mathbb{1}_{[j \in \mathcal{J}_i]} \leq r_i$. It follows that setting $\hat{h}_{i,\hat{S},t} = 1$ for all $i \in \mathcal{I}_{\hat{S}}$ and $\hat{h}_{i,S,t} = 0$ for all $S \neq \hat{S}$ and $i \in \mathcal{I}_S$ yields a feasible solution to optimality equation (10) and we have

$$\vartheta_{i,t}^{\lambda}(r_i) \geq \mathbb{1}_{[i \in \mathcal{I}_{\hat{S}}]} \left\{ \lambda_{i,\hat{S},t} + \sum_{j \in \mathcal{J}_i} P_j(\hat{S}) \left[\vartheta_{i,t+1}^{\lambda}(r_i - 1) - \vartheta_{i,t+1}^{\lambda}(r_i) \right] \right\} + \vartheta_{i,t+1}^{\lambda}(r_i). \tag{25}$$

We also have that $\hat{h}_{\phi,\hat{S},t} = 1$ and $\hat{h}_{\phi,S,t} = 0$ for $S \neq \hat{S}$ is feasible to optimality equation (11) and we have

$$\vartheta_{\phi,t}^{\lambda} \geq \sum_{j \in \hat{S}} P_j(\hat{S}) f_j - \sum_{i \in \mathcal{I}_{\hat{S}}} \lambda_{i,\hat{S},t} + \vartheta_{\phi,t+1}^{\lambda}. \tag{26}$$

Plugging (25) and (26) into (24)

$$
\begin{aligned}
\sum_i \bar{v}_{i,t}(r_i) &= \sum_i \vartheta_{i,t}^{\lambda}(r_i) + \vartheta_{\phi,t}^{\lambda} \\
&\geq \sum_{j \in \hat{S}} P_j(\hat{S}) f_j + \sum_{i \in \mathcal{I}_{\hat{S}}} \sum_{j \in \mathcal{J}_i} P_j(\hat{S}) \Big[\vartheta_{i,t+1}^{\lambda}(r_i - 1) - \vartheta_{i,t+1}^{\lambda}(r_i) \Big] + \sum_i \vartheta_{i,t+1}^{\lambda}(r_i) + \vartheta_{\phi,t+1}^{\lambda} \\
&= \sum_{j \in \hat{S}} P_j(\hat{S}) \Big[f_j + \sum_{i \in \mathcal{I}_j} \bar{v}_{i,t+1}(r_i - 1) - \bar{v}_{i,t+1}(r_i) \Big] + \sum_i \bar{v}_{i,t+1}(r_i).
\end{aligned}
$$

Therefore, $\bar{v} = \{\bar{v}_{it}(r_i) | \forall i, t, r_i\}$ satisfies constraints (4) and is feasible for *PL*. Its objective function value is $\sum_i \bar{v}_{i,1}(r_i^1) = \sum_i \vartheta_{i,1}^{\lambda}(r_i^1) + \vartheta_{\phi,1}^{\lambda} = V_1^{\lambda}(\boldsymbol{r}^1)$. Therefore, $V^{PL} \leq V_1^{\lambda}(\boldsymbol{r}^1)$. □

Proof of Proposition 1

We show the results by induction over the time periods. We begin with the first part of the proposition. It is easy to show the result for the last time period. So we assume that the result holds for time period $t + 1$ and show that it holds for time period t. We have

$$
\begin{aligned}
\vartheta_{i,t}^{\hat{\lambda}}(r_i) &\geq \sum_{S \in \mathcal{C}_i} h_{i,S,t}^{\lambda}(r_i) \hat{\lambda}_{i,S,t} + \vartheta_{i,t+1}^{\hat{\lambda}}(r_i - 1) \Big[\sum_{S \in \mathcal{C}_i} h_{i,S,t}^{\lambda}(r_i) \sum_{j \in \mathcal{J}_i} P_j(S) \Big] \\
&\quad + \vartheta_{i,t+1}^{\hat{\lambda}}(r_i) \Big[1 - \sum_{S \in \mathcal{C}_i} h_{i,S,t}^{\lambda}(r_i) \sum_{j \in \mathcal{J}_i} P_j(S) \Big] \\
&\geq \vartheta_{i,t}^{\lambda}(r_i) + \sum_{S \in \mathcal{C}_i} h_{i,S,t}^{\lambda}(r_i) [\hat{\lambda}_{i,S,t} - \lambda_{i,S,t}] \\
&\quad + \Big[\sum_{k=t+1}^{\tau} \sum_{S \in \mathcal{C}_i} \Big\{ \sum_{x=0}^{r_i} \Pr\{X_{i,k}^{\lambda} = x \mid X_{i,t+1}^{\lambda} = r_i - 1\} h_{i,S,k}^{\lambda}(x) \Big\} [\hat{\lambda}_{i,S,k} - \lambda_{i,S,k}] \Big] \Pr\{X_{i,t+1}^{\lambda} \\
&= r_i - 1 \mid X_{i,t}^{\lambda} = r_i\} \\
&\quad + \Big[\sum_{k=t+1}^{\tau} \sum_{S \in \mathcal{C}_i} \Big\{ \sum_{x=0}^{r_i} \Pr\{X_{i,k}^{\lambda} = x \mid X_{i,t+1}^{\lambda} = r_i\} h_{i,S,k}^{\lambda}(x) \Big\} [\hat{\lambda}_{i,S,k} - \lambda_{i,S,k}] \Big] \Pr\{X_{i,t+1}^{\lambda} \\
&= r_i \mid X_{i,t}^{\lambda} = r_i\} \\
&= \vartheta_{i,t}^{\lambda}(r_i) + \sum_{k=t}^{\tau} \sum_{S \in \mathcal{C}_i} \Big\{ \sum_{x=0}^{r_i} \Pr\{X_{i,k}^{\lambda} = x \mid X_{i,t}^{\lambda} = r_i\} h_{i,S,k}^{\lambda}(x) \Big\} [\hat{\lambda}_{i,S,k} - \lambda_{i,S,k}],
\end{aligned}
$$

where the first inequality holds since $\{h^\lambda_{i,S,t}(r_i) | \forall S \in C_i\}$ is a feasible solution to problem (10) when the Lagrange multipliers are $\hat{\lambda}$. The second inequality holds by the induction assumption, and uses the facts that $\sum_{S \in C_i} [h^\lambda_{i,S,t}(r_i) \sum_{j \in J_i} P_j(S)] = \Pr\{X^\lambda_{i,t+1} = r_i - 1 \mid X^\lambda_{i,t} = r_i\}$ and $1 - \sum_{S \in C_i} [h^\lambda_{i,S,t}(r_i) \sum_{j \in J_i} P_j(S)] = \Pr\{X^\lambda_{i,t+1} = r_i \mid X^\lambda_{i,t} = r_i\}$. On the other hand, the last equality follows by noting that

$$
\Pr\left\{ X^\lambda_{i,k} = x \mid X^\lambda_{i,t} = r_i \right\}
$$
$$
= \Pr\left\{ X^\lambda_{i,k} = x \mid X^\lambda_{i,t} = r_i, X^\lambda_{i,t+1} = r_i \right\} \Pr\left\{ X^\lambda_{i,t+1} = r_i \mid X^\lambda_{i,t} = r_i \right\}
$$
$$
+ \Pr\left\{ X^\lambda_{i,k} = x \mid X^\lambda_{i,t} = r_i, X^\lambda_{i,t+1} = r_i - 1 \right\} \Pr\left\{ X^\lambda_{i,t+1} = r_i - 1 \mid X^\lambda_{i,t} = r_i \right\}
$$
$$
= \Pr\left\{ X^\lambda_{i,k} = x \mid X^\lambda_{i,t+1} = r_i \right\} \Pr\left\{ X^\lambda_{i,t+1} = r_i \mid X^\lambda_{i,t} = r_i \right\}
$$
$$
+ \Pr\left\{ X^\lambda_{i,k} = x \mid X^\lambda_{i,t+1} = r_i - 1 \right\} \Pr\left\{ X^\lambda_{i,t+1} = r_i - 1 \mid X^\lambda_{i,t} = r_i \right\}.
$$

This concludes the proof of the first part of the proposition.

The proof of the second part proceeds in a similar fashion. We assume that the result holds for time period $t + 1$ and show that it holds for time period t. We have

$$
\vartheta^{\hat{\lambda}}_{\phi,t} \geq \sum_S h^\lambda_{\phi,S,t} \Big[\sum_{j \in S} P_j(S) f_j - \sum_{i \in I_S} \hat{\lambda}_{i,S,t} \Big] + \vartheta^{\hat{\lambda}}_{\phi,t+1}
$$
$$
\geq \vartheta^\lambda_{\phi,t} + \sum_{k=t}^\tau \sum_i \sum_{S \in C_i} h^\lambda_{\phi,S,k} \big[\hat{\lambda}_{\phi,S,k} - \lambda_{\phi,S,k} \big],
$$

where the first inequality holds because $\{h^\lambda_{\phi,S,t} | \forall S\}$ is feasible for problem (14) when the Lagrange multipliers are $\hat{\lambda}$ and the second inequality follows from the induction assumption. □

Proof of Lemma 2

Proof The proof is similar to that of Lemma 1. Let r^*, S^* be an optimal solution to $\Phi_t(v)$. Since $S^* \subset Q(r^*) \subset Q_i(r_i^*)$, we have

$$
\max_{r_i \in R_i, S \subset Q_i(r_i)} \left\{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} - \sum_{j \in J_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\} \geq \mathbb{1}_{[S^* \in C_i]} \lambda_{i,S^*,t}
$$
$$
- \sum_{j \in J_i} P_j(S^*) \psi_{i,t+1}(r_i^*) + \Delta_{i,t}(r_i^*).
$$

Therefore, for any $\lambda \in \Lambda$, we have

$$\sum_i \max_{r_i \in \mathcal{R}_i, S \subset Q_i(r_i)} \left\{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\} + \max_{S \subset \mathcal{J}} \left\{ \lambda_{\phi,S,t} \right\}$$

$$\geq \sum_i \left[\mathbb{1}_{[S^* \in C_i]} \lambda_{i,S^*,t} - \sum_{j \in \mathcal{J}_i} P_j(S^*) \psi_{i,t+1}(r_i^*) + \Delta_{i,t}(r_i^*) \right] + \lambda_{\phi,S^*,t}$$

$$= \sum_{i \in \mathcal{I}_{S^*}} \lambda_{i,S^*,t} + \lambda_{\phi,S^*,t} - \sum_j P_j(S^*) \sum_{i \in \mathcal{I}_j} \psi_{i,t+1}(r_i^*) + \sum_i \Delta_{i,t}(r_i^*)$$

$$= R(S^*) - \sum_j P_j(S^*) \sum_{i \in \mathcal{I}_j} \psi_{i,t+1}(r_i^*) + \sum_i \Delta_{i,t}(r_i^*) = \Phi_t(v).$$

It follows that $\Phi_t(v) \leq \Pi_t(v)$. $\qquad\square$

Proof of Lemma 3

Proof Note that if $S \subset Q_i(0)$, then $j \notin \mathcal{J}_i$ for all $j \in S$ and we have $S \notin C_i$. Therefore, $\max_{S \subset Q_i(0)} \{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(0) + \Delta_{i,t}(0) \} = \Delta_{i,t}(0)$.

For $r_i \in \{1, \dots, r_i^1\}$, $Q_i(r_i) = \mathcal{J} = C_i \cup C_i^c$. We have

$$\max_{S \subset \mathcal{J}} \left\{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\}$$

$$= \max \left\{ \max_{S \in C_i} \left\{ \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\}, \max_{S \in C_i^c} \left\{ \Delta_{i,t}(r_i) \right\} \right\}$$

$$= \max \left\{ \max_{S \in C_i} \left\{ \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\}, \Delta_{i,t}(r_i) \right\}.$$

Putting everything together

$$\max_{r_i \in \mathcal{R}_i, S \subset Q_i(r_i)} \{ \mathbb{1}_{[S \in C_i]} \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \}$$

$$= \max \left\{ \Delta_{i,t}(0), \max_{r_i \in \{1, \dots, r_i^1\}} \left\{ \max_{S \in C_i} \left\{ \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\}, \Delta_{i,t}(r_i) \right\} \right\}$$

$$= \max \left\{ \Delta_{i,t}(0), \max_{r_i \in \{1, \dots, r_i^1\}, S \in C_i} \left\{ \lambda_{i,S,t} - \sum_{j \in \mathcal{J}_i} P_j(S) \psi_{i,t+1}(r_i) + \Delta_{i,t}(r_i) \right\} \right\}$$

where the last equality follows from $\Delta_{i,t}(\cdot)$ being a decreasing function. We note that $\Delta_{i,t}(r_i) \leq \Delta_{i,t}(r_i - 1)$ is equivalent to $v_{i,t}(r_i) - v_{i,t}(r_i - 1) \geq v_{i,t+1}(r_i) - v_{i,t+1}(r_i - 1)$, and the latter inequality follows Kunnumkal and Talluri [12] (Lemma 1). Therefore

Table 5 Choice probabilities

S	$P_1(S)$	$P_2(S)$
{1}	1/2	0
{2}	0	10/11
{1, 2}	1/12	10/12

$\Delta_{i,t}(0)$ dominates $\Delta_{i,t}(r_i)$. Note that we do not make an assumption that $v_{i,t}(0) = 0$ for any t, and consequently cannot immediately assume $\Delta_{i,t}(0) = 0 \geq \Delta_{i,t}(r_i)$. $\square$

Lagrangian relaxation using product-specific multipliers

We first describe the Lagrangian relaxation approach using product-specific multipliers (*LRp*) and then give an example which illustrates it can be weaker than *PL* for DC-NRM. Under the independent-class assumption, Topaloglu [24] proposes relaxing the constraints that the same acceptance decisions be made for each product across the resources that it uses by associating Lagrange multipliers with them. The relaxation associates a multiplier $\lambda_{i,j,t}$ for each product j and $i \in \mathcal{I}_j$, and can be interpreted as the portion of the revenue of product j that is allocated to resource i. The Lagrangian relaxation approach of Topaloglu [24] can be easily extended to DC-NRM. We solve the optimality equation

$$v_{i,t}^\lambda(r_i) = \max_{S \subset \mathcal{Q}_i(r_i)} \sum_{j \in \mathcal{J}_i} P_j(S)\Big[\lambda_{i,j,t} + v_{i,t+1}^\lambda(r_i - 1) - v_{i,t+1}^\lambda(r_i)\Big] + v_{i,t+1}^\lambda(r_i) \qquad (27)$$

for resource i, where $\mathcal{Q}_i(r_i) = \Big\{j \mid \mathbb{1}_{[j \in \mathcal{J}_i]} \leq r_i\Big\}$. It is possible to show that $\sum_i v_{i,t}^\lambda(r_i)$ is an upper bound on $V_t(\mathbf{r})$ (Kunnumkal and Topaloglu [14]). So, we can find the tightest upper bound on the optimal expected revenue by solving

$$V^{LRp} = \min_{\left\{\lambda \mid \sum_{i \in \mathcal{I}_j} \lambda_{i,j,t} = f_j, \, \forall j, t\right\}} \sum_i v_{i,1}^\lambda(r_i^1),$$

where the constraint $\sum_{i \in \mathcal{I}_j} \lambda_{i,j,t} = f_j$ ensures that the Lagrange multipliers correspond to a valid fare allocation. In contrast to the independent demands setting (where $V^{PL} = V^{LRp}$), as the following example illustrates, we can have $V^{PL} < V^{LRp}$ for DC-NRM.

Example 1 Consider a network revenue management problem with two products, two resources and a single time period in the booking horizon. The first product uses only the first resource, while the second product uses only the second resource, and we have a single unit of capacity on each resource. Note that in the airline context, this example corresponds to a parallel flights network. The revenues associated with the products are $f_1 = 10$ and $f_2 = 1$. The choice probabilities are given in Table 5. In the Lagrangian relaxation, since we have Lagrange multipliers only for $j \in \mathcal{J}_i$, we

Table 6 Choice probabilities

S	$P_1(S)$	$P_2(S)$
$\{1\}$	50/99	0
$\{2\}$	0	51/101
$\{1,2\}$	1/2	1/2

have only two multipliers $\lambda_{1,1,1}$ and $\lambda_{2,2,1}$. Moreover, the constraint $\sum_{i\in\mathcal{I}_j}\lambda_{i,j,t}=f_j$ implies that $\lambda_{1,1,1}=f_1$ and $\lambda_{2,2,1}=f_2$. Noting that there is only a single time period in the booking horizon and that $\mathcal{Q}_i(1)=\mathcal{J}$ for $i=1,2$,

$$v_{1,1}^{\lambda}(1)=\max_{S\subset\mathcal{J}}\{P_1(S)f_1\}=5$$

and

$$v_{2,1}^{\lambda}(1)=\max_{S\subset\mathcal{J}}\{P_2(S)f_2\}=10/11$$

so that $V^{LRp}=65/11$.

On the other hand, the linear program associated with the piecewise-linear approximation is

$$V^{PL}=\min_{v}v_{1,1}(1)+v_{2,1}(1)$$

s.t

$$v_{1,1}(1)+v_{2,1}(1)\geq\max\{P_1(\{1,2\})f_1+P_2(\{1,2\})f_2,P_1(\{1\})f_1,P_2(\{2\})f_2,0\}$$
$$v_{1,1}(1)+v_{2,1}(0)\geq\max\{P_1(\{1\})f_1,0\}$$
$$v_{1,1}(0)+v_{2,1}(1)\geq\max\{P_2(\{2\})f_2,0\}$$
$$v_{1,1}(0)+v_{2,1}(0)\geq0.$$

Note that the first, second, third and fourth constraints correspond to the states vectors [1, 1], [1, 0], [0, 1] and [0, 0], respectively. It is easy to verify that an optimal solution to *PL* is $v_{1,1}(1)=5, v_{1,1}(0)=10/11, v_{2,1}(1)=0, v_{2,1}(0)=0$ and we have $V^{PL}=5<V^{LRp}$.

Optimal *LRo* multipliers

We give an example below which illustrates that $\sum_{i\in\mathcal{I}_j}\lambda_{i,j,t}=f_j$ for all j fails to hold for *LRo* multipliers.

Example 2 Consider a network revenue management problem with two products, two resources and a single time period in the booking horizon. The first product

uses only the first resource, while the second product uses only the second resource, and we have a single unit of capacity on each resource. Note that in the airline context, this example corresponds to a parallel flights network. The revenues associated with the products are $f_1 = 99$ and $f_2 = 101$. The choice probabilities are given in Table 6. Letting $S_1 = \{1\}$, $S_2 = \{2\}$ and $S_3 = \{1, 2\}$, we have $\sum_{j \in S_1} P_j(S_1) f_j = 50$, $\sum_{j \in S_2} P_j(S_2) f_j = 51$ and $\sum_{j \in S_3} P_j(S_3) f_j = 100$, $\mathcal{I}_{S_1} = \{1\}$, $\mathcal{I}_{S_2} = \{2\}$ and $\mathcal{I}_{S_3} = \{1, 2\}$. If we impose the constraint $\sum_{i \in \mathcal{I}_S} \lambda_{i,S,t} = \sum_{j \in S} P_j(S) f_j$ on the Lagrange multipliers, then we have $\lambda_{1,S_1,1} = 50$, $\lambda_{2,S_2,1} = 51$ and $\lambda_{1,S_3,1} + \lambda_{2,S_3,1} = 100$. We have $\vartheta_{1,1}^{\lambda}(1) = \max\{50, \lambda_{1,S_3,1}\}$ and $\vartheta_{2,1}^{\lambda}(1) = \max\{51, \lambda_{2,S_3,1}\}$. It can be verified that

$$\min_{\lambda \, | \, \lambda_{1,S_1,1}=50, \lambda_{2,S_2,1}=51, \lambda_{1,S_3,1}+\lambda_{2,S_3,1}=100} \vartheta_{1,1}^{\lambda}(1) + \vartheta_{2,1}^{\lambda}(1) = 101 > 100 = V^{PL} = V^{LRo},$$

where the last equality follows from Proposition 2. Therefore, the optimal *LRo* multipliers in the above example satisfy $\sum_{i \in \mathcal{I}_S} \lambda_{i,S,t} < \sum_{j \in S} P_j(S) f_j$.

References

1. Adelman, D.: Dynamic bid-prices in revenue management. Oper. Res. **55**(4), 647–661 (2007)
2. Alptekinoğlu, Aydin, Semple, John H.: The exponential choice model: a new alternative for assortment and price optimization. Oper. Res. **64**(1), 79–93 (2016)
3. Ben-Akiva, M., Lerman, S.: Discrete-Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge (1985)
4. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont (1999)
5. Bront, J.J.M., Méndez-Díaz, I., Vulcano, G.: A column generation algorithm for choice-based network revenue management. Oper. Res. **57**(3), 769–784 (2009)
6. Chen, L., Homem de Mello, T.: Mathematical programming models for revenue management under customer choice. Eur. J. Oper. Res. **203**(2), 294–305 (2010)
7. Gallego, G., Iyengar, G., Phillips, R., Dubey, A.: Managing flexible products on a network. Tech. Rep. TR-2004-01, Dept of Industrial Engineering, Columbia University, NY (2004)
8. Gallego, G., Ratliff, R., Shebalov, S.: A general attraction model and sales-based linear program for network revenue management under customer choice. Oper. Res. **63**(1), 212–232 (2015)
9. Hauser, J.R., Wernerfelt, B.: An evaluation cost model of consideration sets. J. Consum. Res. **16**, 393–408 (1990)
10. Iyengar, S.S., Lepper, M.: When choice is demotivating: Can one desire too much of a good thing? J. Personal. Soc. Psychol. **76**, 995–1006 (2000)
11. Kunnumkal, S., Talluri, K.T.: A new compact linear programming formulation for choice network revenue management. Tech. Rep., Universitat Pompeu Fabra (2012)
12. Kunnumkal, S., Talluri, K.: Technical note—a note on relaxations of the choice network revenue management dynamic program. Oper. Res. **64**(1), 158–166 (2016a)
13. Kunnumkal, S., Talluri, K.T.: On a piecewise-linear approximation for network revenue management. Math. Oper. Res. **41**(1), 72–91 (2016b)
14. Kunnumkal, S., Topaloglu, H.: A new dynamic programming decomposition method for the network revenue management problem with customer choice behavior. Prod. Oper. Manag. **19**(5), 575–590 (2010)
15. Liu, Q., van Ryzin, G.J.: On the choice-based linear programming model for network revenue management. Manuf. Serv. Oper. Manag. **10**(2), 288–310 (2008)
16. Meissner, J., Strauss, A.K.: Network revenue management with inventory-sensitive bid prices and customer choice. Eur. J. Oper. Res. **216**(2), 459–468 (2012)
17. Meissner, J., Strauss, A.K., Talluri, K.T.: An enhanced concave programming method for choice network revenue management. Prod. Oper. Manag. **22**(1), 71–87 (2013)

18. Rusmevichientong, P., Shmoys, D., Tong, C., Topaloglu, H.: Assortment optimization under the multinomial logit model with random choice parameters. Prod. Oper. Manag. **23**(11), 2023–2039 (2014)

19. Schweitzer, P., Seidmann, A.: Generalized polynomial approximations in markovian decision processes. J. Math. Anal. Appl. **110**, 568–582 (1985)

20. Talluri, K.T.: Airline revenue management with passenger routing control: a new model with solution approaches. Int. J. Serv. Technol. Manag. **2**, 102–115 (2001)

21. Talluri, K.T.: New formulations for choice network revenue management. INFORMS J. Comput. **26**(2), 401–413 (2014)

22. Talluri, K.T., van Ryzin, G.J.: Revenue management under a general discrete choice model of consumer behavior. Manag. Sci. **50**(1), 15–33 (2004a)

23. Talluri, K.T., van Ryzin, G.J.: The Theory and Practice of Revenue Management. Kluwer, New York (2004b)

24. Topaloglu, H.: Using Lagrangian relaxation to compute capacity-dependent bid prices in network revenue management. Oper. Res. **57**, 637–649 (2009)

25. Vossen, T.W.M., Zhang, D.: Reductions of approximate linear program for network revenue management. Oper. Res. **63**(6), 1352–1371 (2015)

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

Affiliations

Sumit Kunnumkal[1] · Kalyan Talluri[2]

Sumit Kunnumkal
sumit_kunnumkal@isb.edu

1   Indian School of Business, Gachibowli 500032, India

2   Imperial College Business School, South Kensington Campus, London SW7 2AZ, UK