Management Science

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

A Re-Solving Heuristic with Uniformly Bounded Loss for Network Revenue Management

Pornpawee Bumpensanti, He Wang

informs
http://pubsonline.informs.org/journal/mnsc
MANAGEMENT SCIENCE
Vol. 66, No. 7, July 2020, pp. 2993–3009
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

A Re-Solving Heuristic with Uniformly Bounded Loss for Network Revenue Management

Pornpawee Bumpensanti,[a] **He Wang**[a]

[a] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Contact: pornpawee@gatech.edu, http://orcid.org/0000-0002-4279-8909 (PB); he.wang@isye.gatech.edu, http://orcid.org/0000-0001-7444-2053 (HW)

Received: May 23, 2018
Revised: November 27, 2018; March 19, 2019
Accepted: March 24, 2019
Published Online in Articles in Advance: March 16, 2020

https://doi.org/10.1287/mnsc.2019.3365

Abstract. We consider a canonical quantity-based network revenue management problem where a firm accepts or rejects incoming customer requests irrevocably in order to maximize expected revenue given limited resources. Because of the curse of dimensionality, the exact solution to this problem by dynamic programming is intractable when the number of resources is large. We study a family of re-solving heuristics that periodically re-optimize an approximation to the original problem known as the deterministic linear program (DLP), where random customer arrivals are replaced by their expectations. We find that, in general, frequently re-solving the DLP produces the same order of revenue loss as one would get without re-solving, which scales as the square root of the time horizon length and resource capacities. By re-solving the DLP at a few selected points in time and applying thresholds to the customer acceptance probabilities, we design a new re-solving heuristic with revenue loss that is uniformly bounded by a constant that is independent of the time horizon and resource capacities.

History: Accepted by Kalyan Talluri, revenue management and market analytics.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2019.3365.

Keywords: revenue management • resource allocation • dynamic programming • linear programming

1. Introduction

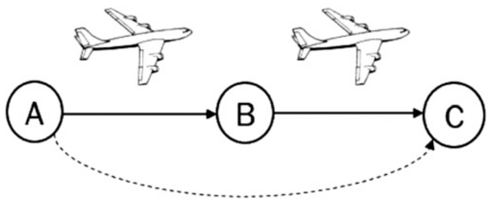
The *network revenue management* (NRM) problem (Williamson 1992, Gallego and van Ryzin 1997) is a classical model that has been extensively studied in the revenue management literature for over two decades. The problem is concerned with maximizing revenue given limited resource and time, and it has a wide range of applications in the airline, retail, advertising, and hospitality industries (see examples in Talluri and van Ryzin 2006). However, the exact solution to the NRM problem is difficult to compute when the number of resources is large. Heuristics proposed in the previous literature typically have optimality gaps (i.e., expected revenue losses compared with the optimal solution) that increase with the time horizon and the resource capacities. In this paper, we propose a new heuristic for the NRM problem for which the revenue loss is independent of the time horizon and the resource capacities.

The NRM problem is stated as follows: there is a set of resources with finite capacities that are available for a finite time horizon. Heterogeneous customers arrive sequentially over time. Customers are divided into different classes based on their consumption of resources and the prices that they pay. Each class of customer may request multiple types of resources and multiple units of each resource. On a customer's arrival, a decision maker must irrevocably accept or reject the customer. If the customer is accepted and there is enough remaining capacity, she consumes the resources requested and pays a fixed price associated with her class. Otherwise, if the customer is rejected, no revenue is collected, and no resources are used. Unused resources at the end of the finite horizon are perishable and have no salvage value. The decision maker's objective is to maximize the expected revenue earned during the finite horizon.

We note that the formulation stated above is more specifically known as the "quantity-based" NRM problem. In another formulation referred to as the "price-based" NRM problem, the decision maker chooses posted prices rather than accept/reject decisions. The two formulations are different, but they are equivalent in some special cases (Maglaras and Meissner 2006). We focus on the quantity-based formulation in this paper.

A classical application of the NRM problem is in airline seat revenue management (Williamson 1992, Gallego and van Ryzin 1997). Here, the resources correspond to flight legs, and the capacity corresponds to the number of seats on each flight. The resources are perishable on the date of flight departure. Arriving customers are divided into separate classes defined by combinations of itinerary and fare. A simple flight network of two flight legs and three

Figure 1. A Flight Network of Two Flight Legs ($A \rightarrow B, B \rightarrow C$) and Three Itineraries



itineraries is shown in Figure 1. The objective of the airline is to maximize the expected revenue earned from allocating available seats to different classes of customers. Notice that the problem cannot be decomposed for each individual flight leg, because some itineraries use multiple resources simultaneously (e.g., in Figure 1, customers traveling from A to C would request itinerary $A \rightarrow B \rightarrow C$). In practice, the huge size of airline networks makes solving this problem challenging.

1.1. Deterministic Linear Program Approximation and Re-Solving Heuristics

In theory, the NRM problem can be solved by dynamic programming; however, because the state space grows exponentially with the number of resources, the dynamic programming formulation is often intractable. Therefore, we focus on heuristics with provable performance guarantees in this paper. We define *revenue loss* as the gap between the expected revenue of a heuristic policy and that of the optimal policy. As is common in the revenue management literature, the effectiveness of heuristic polices is evaluated in an asymptotic regime where resource capacities and customer arrivals are both scaled proportionally by a factor of k ($k = 1, 2, \ldots$). Intuitively, this asymptotic regime increases market size while keeping resource scarcity (i.e., the ratio of capacity to demand) at a constant level. We assume this standard asymptotic setting throughout the paper.

One popular heuristic for the NRM problem that is extensively studied in the academic literature and widely used in practice is based on the deterministic linear programming (DLP) approximation, where the customer demand distributions are replaced by their expectations. The solution of the DLP can then be used to construct heuristic policies. Under the asymptotic scaling defined above, Gallego and van Ryzin (1994, 1997) have shown that the revenue loss of DLP-based static control policies is $\Theta(\sqrt{k})$ when the system size is scaled by k. The book by Talluri and van Ryzin (2006) provides a comprehensive overview of different types of DLP-based control policies, such as booking limit control, bid-price control, and so forth, and their variations.

An apparent weakness of the DLP approximation is that it ignores randomness in the arrival process and fails to incorporate information acquired through time. To include updated information, a simple approach is to re-optimize the DLP from time to time while replacing the initial capacity in the DLP with the remaining capacity at each re-solving point. The new solution to the updated DLP is then used to adjust control policies. The re-solving approach is intuitive and widely used in practice. We refer to this family of solution techniques as *re-solving heuristics*. One might expect that re-solving the DLP would yield better performance, because it includes updated information. Surprisingly, Cooper (2002) provides a counterexample where the performance of booking limit control deteriorates by re-solving the DLP. Furthermore, Chen and Homem-de Mello (2010) give an example where re-solving the DLP worsens the performance for bid-price control. Jasin and Kumar (2013) analyze the performance of re-solving both booking limit and bid-price controls. They showed that, when the initial capacity and customer arrival rates are both scaled by k, the revenue loss of re-solving heuristics is $\Omega(\sqrt{k})$, even by optimizing over the re-solving schedule or increasing re-solving frequency.

Despite those negative results, we note that there are several ways to construct control policies from the DLP; therefore, it is possible that some control policies are suitable for applying the re-solving technique, whereas others are not. Some recent literature draws attention to a specific type of control policy called *probabilistic allocation*, which seems suitable for applying the re-solving technique. Probabilistic allocation control is a randomized algorithm that accepts each arriving customer with some probability. Using the probabilistic allocation control, Reiman and Wang (2008) propose a heuristic policy that re-solves the DLP exactly once during the horizon. In their proposed policy, the re-solving time is random and determined endogenously by the heuristic policy. In the asymptotic setting, Reiman and Wang (2008) show that the revenue loss of their policy is $o(\sqrt{k})$. This is an improvement over the $\Theta(\sqrt{k})$ revenue loss of DLP-based static policies.

Jasin and Kumar (2012) consider the NRM problem with customer choice, which generalizes the quantity-based NRM problem. They analyze an algorithm that is based on probabilistic allocation control and re-solves the DLP after each unit of time. They show that the algorithm has a revenue loss of $O(1)$ when the system size is scaled by $k \rightarrow \infty$. A similar $O(1)$ revenue loss is obtained by Wu et al. (2015) for the case of one resource. However, both the results of Jasin and Kumar (2012) and the results of Wu et al. (2015) require the optimal solution to DLP (before any updating) to be nondegenerate; this assumption will be formally stated in Section 3, which seems to be central to the hardness of the NRM problem. Moreover, Wu et al. (2015)

show that, when the optimal solution is nondegenerate but nearly degenerate, the constant factor in $O(1)$ can become arbitrarily large. In this paper, we aim to establish a uniform $O(1)$ loss for the general NRM problem without assuming nondegeneracy.

1.2. Main Contributions

We propose a new re-solving heuristic that has a uniformly bounded revenue loss when the system size is scaled by $k \to \infty$. (Recall that the rate of revenue loss is defined for a sequence of problems indexed by $k = 1, 2, \ldots$, where the capacities and arrival rates are multiplied by k, whereas other parameters are treated as constants.) The bound is uniform in the sense that it does not depend on the ratio between capacities and time. Therefore, this result does not require the nondegeneracy assumption. Our $O(1)$ bound improves the $o(\sqrt{k})$ bound in Reiman and Wang (2008) and also improves the $O(1)$ bound in Jasin and Kumar (2012), where the constant factor requires nondegeneracy assumption and depends implicitly on problem instances. (However, as we noted before, Jasin and Kumar (2012) considered the NRM problem with customer choice, which generalizes the quantity-based NRM problem.) We call our new algorithm infrequent re-solving with thresholding (IRT). The intuition behind the IRT algorithm is that it is not necessary to update the DLP at early stage of the horizon, because the solution to the DLP barely changes after updating. It is sufficient to re-solve the DLP at a few carefully selected time points near the *end* of the horizon. In total, the IRT algorithm has $O(\log\log k)$ re-solving times for a system with scaling size k. Furthermore, a "thresholding" technique is applied in the case when the DLP solution after re-solving is nearly degenerate. The re-solving schedule and the thresholds of the IRT algorithm are designed in such a way that the accumulated random deviations before the re-solving point can be corrected after re-solving with high probability.

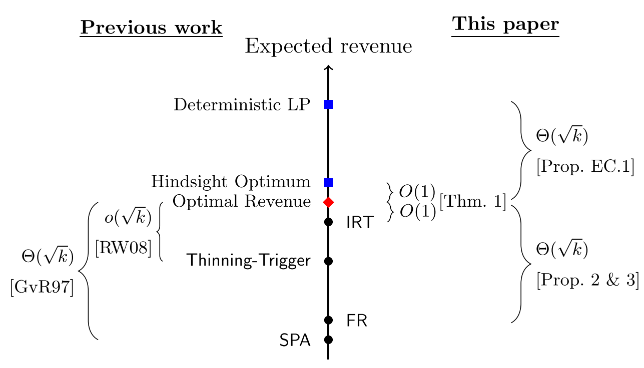
Then, we give a tight performance bound of the re-solving heuristic proposed by Jasin and Kumar (2012) but without assuming that the optimal solution to the DLP is nondegenerate. The heuristic in Jasin and Kumar (2012), which we call frequent re-solving (FR), re-solves the DLP after each unit of time. One would expect that, by re-solving the DLP frequently and thus, constantly updating capacity information, the decision maker can improve the expected revenue. Indeed, Jasin and Kumar (2012) have shown that, under the nondegeneracy assumption, the revenue loss of this policy is $O(1)$ when the system size is scaled by $k \to \infty$. However, we find that the revenue loss of this policy is $\Theta(\sqrt{k})$ in general, which has the same order of revenue loss as DLP-based static heuristics without any re-solving (Gallego and van Ryzin 1997, Talluri and van Ryzin 1998, Cooper 2002). In particular, Proposition 2 shows that there exists a problem instance where the revenue loss of this policy is at least $\Omega(\sqrt{k})$. To analyze this instance, we used the Berry–Esseen bound and Freedman's inequality to show that the probability of revenue loss being larger than $\Omega(\sqrt{k})$ is bounded away from zero. This result suggests that the nondegeneracy assumption made by Jasin and Kumar (2012) is necessary to obtain $O(1)$ revenue loss and explains why the $O(1)$ factor in Wu et al. (2015) must be arbitrarily large when the DLP optimal solution is converging to a degenerate point. Then, Proposition 3 shows that the revenue loss of this policy is bounded above by $O(\sqrt{k})$ in the general case, which also improves the $o(k)$ bound in Maglaras and Meissner (2006). The proof is based on a key inequality that bounds the average remaining capacity as a function of the remaining time.

In Figure 2, we summarize the performance of existing re-solving heuristics for the NRM problem. In this figure, the vertical axis represents the expected revenue, which increases from the bottom to the top. We highlight the gap between different heuristics and upper bounds compared with the optimal revenue, which in principle, can be obtained from dynamic programming but is hard to compute directly. The main result of the paper (Theorem 1) simultaneously establishes an $O(1)$ upper bound of the *hindsight optimum* (HO) and an $O(1)$ revenue loss of the IRT algorithm.

1.3. Other Related Work

The re-solving heuristics defined in the NRM context are generally known as *certainty equivalent control* in dynamic programming. In certainty equivalent control, each random disturbance is fixed at a nominal

Figure 2. (Color online) Summary of the Results in the Previous Literature (Left Side) and Our Main Results (Right Side)



Notes. The diamond node represents the expected revenue of the optimal policy (hard to compute), the square nodes represent upper bounds to the optimal revenue, and the circle nodes refer to revenues earned under different heuristics. The factor k is the scale of both time horizon and capacities.

value (e.g., its mean), and then, an optimal control sequence for the certainty equivalence approximation is found. Only the first control in the sequence is applied, the rest of them are discarded, and the same process is repeated in the next stage. An introduction to certainty equivalent control can be found in section 6.1 in Bertsekas (2005). Secomandi (2008) discussed whether certainty equivalent control guarantees performance improvement in the network revenue management setting.

The quantity-based NRM model can be generalized in several ways. One extension assumes that the decision maker offers a set of products to each arriving customer and that customers choose some products from the offered set based on some discrete choice model (Talluri and van Ryzin 2004, Liu and van Ryzin 2008). Another stream of literature assumes that either the customers' arrival process or the distribution of their reservation price is unknown and requires the decision maker to learn the distribution exclusively from past observations (Besbes and Zeevi 2012, Jasin 2015, Ferreira et al. 2018). Maglaras and Meissner (2006) and Talluri and van Ryzin (2006) discussed the case where the decision maker posts price (price-based NRM) versus the case where the decision maker chooses accept/reject (quantity-based NRM).

The NRM problem considered here is related to the online knapsack/secretary problem studied by Kleywegt and Papastavrou (1998), Kleinberg (2005), Babaioff et al. (2007), Arlotto and Gurvich (2019), and Arlotto and Xie (2019). In particular, Arlotto and Gurvich (2019) consider a multiselection secretary problem, where the decision maker sequentially selects i.i.d. random variables in order to maximize the expected value of the sum given a fixed budget. As such, by viewing each random variable as a customer arrival, the multiselection secretary problem is a special case of the NRM problem in which there is only a single resource and each customer requests exactly one unit of the resource. Arlotto and Gurvich (2019) propose an online policy that has a uniformly bounded regret compared with the optimal offline policy. Their policy accepts or rejects an arriving customer by comparing the budget ratio (i.e., ratio of remaining budget to remaining arrivals) with some fixed thresholds. However, it is unclear whether their technique can be generalized to the general NRM setting with multiple resources, because the thresholds in their policy are specifically defined for a single resource.

Recently, Vera and Banerjee (2019) studied an online packing problem, which has the same mathematical formulation as the network revenue management problem. They proposed a re-solving heuristic that achieves $O(1)$ revenue loss without the nondegeneracy assumption and under mild assumptions on the customer arrival processes. Unlike the IRT algorithm, their proposed algorithm re-solves the DLP every time there is an arrival; the algorithm then accepts that arrival if the acceptance probability from the DLP is greater than 0.5 and rejects it otherwise. Their proof is based on a novel argument that compensates for the optimal offline algorithm and forces it to follow the decisions of their online algorithm. The design of their algorithm and their proof idea are significantly different from those in this paper.

1.4. Notation

For a positive integer n, let $[n]$ denote the set $\{1,\ldots,n\}$. Given two real numbers $a\in\mathbb{R}$ and $b\in\mathbb{R}$, let $a\wedge b :=\min\{a,b\}$, $a\vee b := \max\{a,b\}$, and $a^+ := a\vee 0$. For any real number x, let $\lfloor x\rfloor$ be the largest integer less than or equal to x, and let $\lceil x\rceil$ be the smallest integer greater than or equal to x. For a set S, let $|S|$ denote the cardinality of S. For two functions $f(T)$ and $g(T)>0$, we write $f(T)=O(g(T))$ if there exists a constant M_1 and a constant T_1 such that $f(T)\le M_1 g(T)$ for all $T\ge T_1$; we write $f(T)=\Omega(g(T))$ if there exists a constant M_2 and a constant T_2 such that $f(T)\ge M_2 g(T)$ for all $T\ge T_2$. If $f(T)=O(g(T))$ and $f(T)=\Omega(g(T))$ both hold, we denote it by $f(T)=\Theta(g(T))$.

2. Problem Formulation and Approximations

Suppose that there is a finite horizon with length T. There are n classes of customers indexed by $j\in[n]$. The arrival process of customers in class j, $\{\Lambda_j(t), 0\le t\le T\}$, follows a Poisson process of rate λ_j. We let $\Lambda_j(t_1,t_2)$ denote the number of the arrivals of class j customers during $(t_1,t_2]$ for $0\le t_1<t_2\le T$: that is, $\Lambda_j(t_1,t_2)=\Lambda_j(t_2)-\Lambda_j(t_1)$. Arrival processes of different classes are independent. On arrival, each customer must be either accepted or rejected. Let r_j denote the revenue received by accepting a class j customer and $\boldsymbol{r}=[r_1,\ldots,r_n]^\top$ be the vector of such revenues. There are m resources indexed by $l\in[m]$, where resource l has initial capacity C_l. The vector of the initial capacities is given by $\boldsymbol{C}=[C_1,\ldots,C_m]^\top$. If a customer is accepted, a_{lj} units of resource l are consumed to serve a class j customer; let $\boldsymbol{A}_j=[a_{1j},\ldots,a_{mj}]^\top$ be the column vector associated with class j customers. Let $A\in\mathbb{R}^{m\times n}$ be the *bill-of-materials* (BOM) matrix defined as $A=[\boldsymbol{A}_1;\ldots;\boldsymbol{A}_n]$. If a customer is rejected, no revenue is collected, and no resource is used. Unused resources at the end of the horizon are perishable and have no salvage value. The objective of the decision maker is to maximize the expected revenue earned during the entire horizon by deciding whether to accept each arriving customer.

For a control policy π, let $z_j^\pi(t_1,t_2)$ be the number of class j customers admitted during $(t_1,t_2]$ $(\forall j\in[n], 0\le t_1<t_2\le T)$ under that policy. We call a

policy *admissible* if it is nonanticipating and satisfies $\sum_{j=1}^n \boldsymbol{A}_j z_j^\pi(0,T) \le \boldsymbol{C}$ a.s., and $z_j^\pi(t_1,t_2) \le \Lambda_j(t_1,t_2)$ a.s., $\forall j \in [n], 0 \le t_1 < t_2 \le T$. Let Π be the set of all admissible policies. The expected revenue under policy $\pi \in \Pi$ is defined as $v^\pi = \mathrm{E}[\sum_{j=1}^n r_j z_j^\pi(0,T)]$. We use $v^* = \sup_{\pi\in\Pi} v^\pi$ to denote the expected revenue under the optimal policy. If v^π is the expected revenue of a feasible policy $\pi \in \Pi$, we call $v^* - v^\pi$ the revenue loss of policy π.

2.1. Asymptotic Framework

The standard asymptotic framework in revenue management measures performance of heuristics when the capacities and customer arrivals are scaled up proportionally. Under this asymptotic scaling, we consider revenue loss of a sequence of problems indexed by $k = 1, 2, \ldots$, where the capacities and arrival rates are multiplied by k, whereas all other problem parameters are treated as constants.

To avoid cumbersome notation where lots of variables and quantities are indexed by k, in the rest of the paper, we consider a different but equivalent asymptotic scaling, where the customer arrival rates λ_j ($j \in [n]$) are kept as constants, the time horizon is scaled up by $T = 1, 2, \ldots$, and the resource capacities are scaled up proportionally by $C_l = b_l T$ ($l \in [m]$). Because the arrivals follow Poisson processes, scaling up the arrival rates and scaling up the horizon length have the same effect. We will thus express the revenue loss of heuristics in the order of T. Note that the horizon length (T) plays the same role as the scaling factor (k) in the standard asymptotic regime. For example, if we say that the revenue loss of an algorithm is $O(\sqrt{T})$, it implies that revenue loss of that algorithm is $O(\sqrt{k})$ under the standard scaling regime.

2.2. Previous Work on Upper-Bound Approximations

2.2.1. DLP. The DLP formulation is obtained by replacing all random variables with their expectations. Because the expected number of arrivals of class j customers during the horizon is $\lambda_j T$ for $j \in [n]$, the DLP formulation is given by

$$v^{\mathrm{DLP}} = \max_{\boldsymbol{y}} \left\{ \sum_{j=1}^n r_j y_j \;\middle|\; \sum_{j=1}^n \boldsymbol{A}_j y_j \le \boldsymbol{C}, \text{ and } 0 \le y_j \le \lambda_j T,\ \forall j \in [n] \right\}. \quad (1)$$

In this formulation, decision variables y_j can be viewed as the expected number of class j customers to be accepted in $[0, T]$. The first constraint specifies that the expected usage of all m resources cannot exceed their initial capacities, $\boldsymbol{C} = [C_1, \ldots, C_m]^\top$, and the second constraint specifies that the number of accepted customers from class j cannot exceed the expected number of arrivals, $\lambda_j T$.

Suppose that y^* is an optimal solution to (1). The optimal value of DLP is given by $v^{\mathrm{DLP}} = \sum_{j=1}^n r_j y_j^*$. It can be shown that v^{DLP} is an upper bound of the expected revenue of the optimal policy, v^*, namely $v^* \le v^{\mathrm{DLP}}$ (Gallego and van Ryzin 1997). Intuitively, DLP is a relaxation of the original problem, because it only requires the capacity constraints to be satisfied in expectation; therefore, v^{DLP} is an upper bound of v^*.

Equivalently, we can reformulate the DLP in (1) by letting x_j be the average number of class j customers accepted per unit time (i.e., $x_j = y_j/T$). Then, we get

$$v^{\mathrm{DLP}} = \max_{\boldsymbol{x}} \left\{ T \sum_{j=1}^n r_j x_j \;\middle|\; \sum_{j=1}^n \boldsymbol{A}_j x_j \le \boldsymbol{b}, \text{ and } 0 \le x_j \le \lambda_j,\ \forall j \in [n] \right\}, \quad (2)$$

where $\boldsymbol{b} = [b_1, \ldots, b_m]^\top$ refers to the vector of available resources per unit time (i.e., $b_l = C_l/T, \forall l \in [m]$). Let x_j^* for $j \in [n]$ be an optimal solution to (2). The optimal value to the DLP is given by $v^{\mathrm{DLP}} = T\sum_{j=1}^n r_j x_j^*$.

2.2.2. Hindsight Optimum. The hindsight optimum is the optimal revenue obtained when the total number of arrivals is known in advance. Recall that the random variable $\Lambda_j(T)$ represents the total arrivals of class j customers in $[0, T]$. If the values of $\Lambda_j(T)$ are known, let z_j be the number of class j customers accepted in $[0, T]$; the optimal acceptance policy is given by

$$V^{\mathrm{HO}} = \max_{\boldsymbol{z}} \left\{ \sum_{j=1}^n r_j z_j \;\middle|\; \sum_{j=1}^n \boldsymbol{A}_j z_j \le \boldsymbol{C}, \text{ and } 0 \le z_j \le \Lambda_j(T),\ \forall j \in [n] \right\}. \quad (3)$$

Let V^{HO} be the optimal objective value and $\bar{z}_j, j \in [n]$ be the optimal solution; note that V^{HO} and $\bar{z}_j$ are random variables that depend on $\Lambda_j(T)$. The HO is defined as the expectation of the optimal objective value (i.e., $v^{\mathrm{HO}} = \mathrm{E}[V^{\mathrm{HO}}] = \mathrm{E}[\sum_{j=1}^n r_j \bar{z}_j]$).

The hindsight optimum is obviously an upper bound to the optimal revenue of the original problem, because the decision maker does not know the future arrivals at time $t = 0$. In fact, it can be shown that hindsight optimum is a *tighter* upper bound than the DLP, namely $v^* \le v^{\mathrm{HO}} \le v^{\mathrm{DLP}}$ (Talluri and van Ryzin 1998). This is easily verified, because the expectation of the hindsight optimal solution, $\mathrm{E}[\bar{z}_j]$, is a feasible solution to the DLP.

We use the following definition throughout the paper.

Definition 1. Let v^{π} be the expected revenue associated with an admissible control policy π. We refer to $v^{\text{HO}} - v^{\pi}$ as the *regret* of that policy. (Note that, because $v^* \leq v^{\text{HO}}$, the revenue loss of the control policy, $v^* - v^{\pi}$, is upper bounded by its regret.)

2.3. Static Probabilistic Allocation Heuristic

There are various ways to construct heuristic policies using the optimal solution of DLP. An overview can be found in chapter 2 of Talluri and van Ryzin (2006). One intuitive approach is to interpret the solution to DLP as acceptance probabilities. Suppose that x^* is an optimal solution to DLP in (2). For each arriving customer, if the customer belongs to class j, she or he would be accepted independently with probability x_j^*/λ_j throughout the time horizon. Because customers from each class are accepted with probabilities that are static, we call this heuristic static probabilistic allocation (SPA). The SPA policy is formally stated in Algorithm 1.

Algorithm 1 (Static Probabilistic Allocation Heuristic)
initialize $\boldsymbol{x}^* \leftarrow \arg\max_x\{\sum_{j=1}^n r_j x_j \mid \sum_{j=1}^n \boldsymbol{A}_j x_j \leq \boldsymbol{C}/T$, and $0 \leq x_j \leq \lambda_j, \forall j \in [n]\}$; $\boldsymbol{C}' \leftarrow \boldsymbol{C}$
for all customers arriving in $[0, T]$, **do**
  if the customer belongs to class j and $\boldsymbol{A}_j \leq \boldsymbol{C}'$ $(\forall j \in [n])$, **then**
    accept the customer with probability x_j^*/λ_j
    if the customer is accepted, update capacity $\boldsymbol{C}' \leftarrow \boldsymbol{C}' - \boldsymbol{A}_j$
  else
    reject the customer
  end if
end for

The expected revenue of the SPA policy, denoted by v^{SPA}, can be computed as follows. Because the total number of arrivals from class j follows a Poisson distribution with mean $\lambda_j T$, the number of customers that the algorithm *attempts* to accept from class j follows a Poisson distribution with mean $(\lambda_j T) \cdot x_j^*/\lambda_j = x_j^* T = y_j^*$. Because of limited capacity, we must reject any customer from class j if the remaining capacity C' does not satisfy $A_j \leq C'$. It is straightforward to show that the expected number of customers who are turned away because of capacity limits is $O(\sqrt{T})$ (e.g., Gallego and van Ryzin 1997, Reiman and Wang 2008). Thus, we have

$$v^{\text{SPA}} = \sum_{j=1}^{n} r_j y_j^* - O\left(\sqrt{T}\right) = v^{\text{DLP}} - O\left(\sqrt{T}\right).$$

Recall from Section 2.2.1 that v^{DLP} is an upper bound of the expected revenue under the optimal policy, namely $v^* \leq v^{\text{DLP}}$. Thus, the revenue of SPA is bounded by $v^{\text{SPA}} \geq v^* - O(\sqrt{T})$.

3. Frequent Re-Solving and Degeneracy

An obvious drawback of the SPA policy constructed from the DLP is that it does not take into account the randomness of demand or the updated information after $t = 0$. This motivates us to consider re-solving heuristics, which periodically re-optimize the DLP using the updated capacity information to adjust customer admission controls.

In particular, the following re-solving heuristic, which we referred to as FR, has been studied by Jasin and Kumar (2012) and Wu et al. (2015). The FR policy divides the horizon into T periods and re-solves the linear program (LP) at the beginning of each period. At time $t = 0, 1, \ldots, T-1$, let $C_l(t)$ denote the remaining capacity of resource $l \in [m]$. We let $b_l(t) := \frac{C_l(t)}{T-t}$ be the average available capacity of resource l in period t. Let $\boldsymbol{C}(t)$ and $\boldsymbol{b}(t)$ denote the vectors of the remaining capacities and the average remaining capacities per unit time at time t, respectively, for all of the resources. We outline the FR policy in Algorithm 2.

Algorithm 2 (Frequent Re-Solving Heuristic)
initialize: set $\boldsymbol{C}(0) = \boldsymbol{C}$ and $\boldsymbol{b}(0) = \boldsymbol{C}/T$
for $t = 0, 1, \ldots, T-1$, **do**
  set $\boldsymbol{x}(t) \leftarrow \arg\max_{\boldsymbol{x}}\{\sum_{j=1}^n r_j x_j \mid \sum_{j=1}^n \boldsymbol{A}_j x_j \leq \boldsymbol{b}(t)$, and $0 \leq x_j \leq \lambda_j, \forall j \in [n]\}$
  set $\boldsymbol{C}' \leftarrow \boldsymbol{C}(t)$
  for all customers arriving in $[t, t+1)$, **do**
    if the customer belongs to class j and $\boldsymbol{A}_j \leq \boldsymbol{C}'$ $(\forall j \in [n])$, **then**
      accept the customer with probability $x_j(t)/\lambda_j$
      if the customer is accepted, update $\boldsymbol{C}' \leftarrow \boldsymbol{C}' - \boldsymbol{A}_j$
    else
      reject the customer
    end if
  end for
  set $\boldsymbol{C}(t+1) \leftarrow \boldsymbol{C}'$ and $\boldsymbol{b}(t+1) \leftarrow \frac{\boldsymbol{C}(t+1)}{T-t-1}$
end for

Jasin and Kumar (2012) show that, when the optimal solution to DLP (2) is *nondegenerate*, FR has a revenue loss of $O(1)$, namely, the revenue loss is bounded when the problem size k grows. The optimal solution $\boldsymbol{x}^*$ is nondegenerate if

$$\left|\left\{j \in [n] : x_j^* = 0 \text{ or } x_j^* = \lambda_j\right\}\right| + \left|\left\{l \in [m] : \sum_{j=1}^{n} a_{lj} x_j^* = b_j\right\}\right| = n. \tag{4}$$

The $O(1)$ loss is a significant improvement from the $O(\sqrt{T})$ revenue loss of SPA.

However, the assumption of nondegenerate DLP solution is critical to achieve the $O(1)$ loss. The proofs by Jasin and Kumar (2012) and Wu et al. (2015) are built on a key observation that the ratio of remaining capacities to remaining time, $\boldsymbol{b}(t)$, is a martingale (see also Arlotto and Gurvich 2019 for a discussion on this martingale property). If the optimal solution $\boldsymbol{x}^*$ is safely far from any degenerate solutions, with high probability, the adjusted solution $\boldsymbol{x}(t)$ in Algorithm 2 shares the same basis with $\boldsymbol{x}^*$; therefore, the revenue loss of FR can be bounded. It is unclear from the analyses of Jasin and Kumar (2012) and Wu et al. (2015) whether the nondegeneracy assumption is just an artifact of their analysis techniques or something intrinsic to the performance of FR. This motivates us to examine closely the role of the nondegeneracy assumption.

3.1. A Degenerate Example

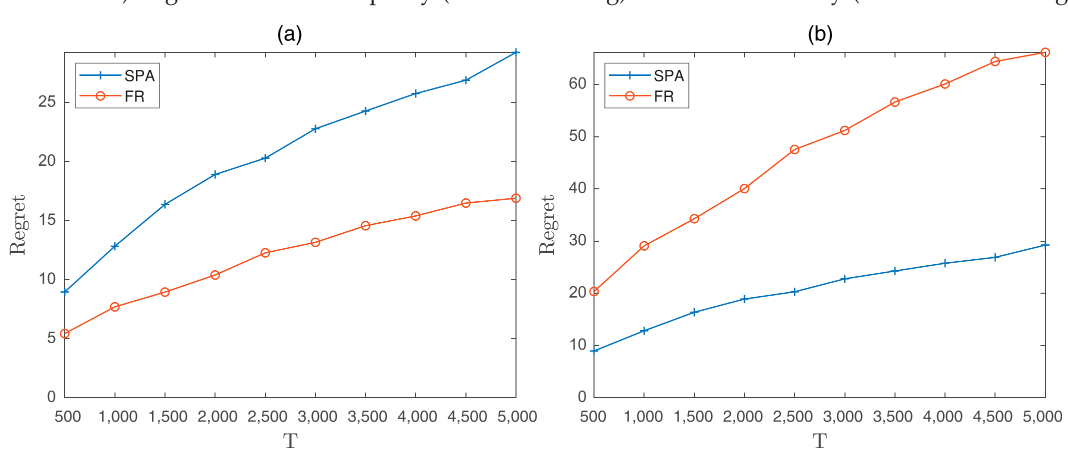
We will illustrate the issue of degenerate DLP solutions using the following numerical example while deferring the theoretical analysis of the FR policy to Section 5.

Suppose that there are two classes of customers and one resource. Customers from each class arrive according to a Poisson process with rate 1. Customers from both classes, if accepted, consume one unit of resource but pay different prices, r_1 and r_2. We first compare the expected revenue losses of the FR policy and the SPA policy, which does not re-solve after $t = 0$, to examine the effect of frequent re-solving. We simulate the FR policy and the SPA policy when the average capacity per unit time $b = 1$ (therefore, the total capacity is T) for two price scenarios: (a) $r_1 = 2$ and $r_2 = 1$ and (b) $r_1 = 5$ and $r_2 = 1$ and for varying horizon length $T = 500, \ldots, 5{,}000$. In both scenarios, the optimal solution to the DLP (2) is $x_1^* = 1, x_2^* = 0$. From Equation (4), we have $|\{j \in [n] : x_j^* = 0 \text{ or } x_j^* = \lambda_j\}| + |\{l \in [m] : \sum_{j=1}^n a_{lj} x_j^* = b_j\}| = 3 > n = 2$, and thus, the DLP solution in this example is degenerate.

Recall that the expected revenue loss of the FR policy is defined as $v^* - v^{\mathsf{FR}}$. Because calculating v^* requires solving dynamic programs, we use the regret $v^{\mathrm{HO}} - v^{\mathsf{FR}}$ (see Definition 1 in Section 2.2.2) as a proxy of the expected revenue loss. In Section 4, we will show that $v^{\mathrm{HO}} - v^* = O(1)$, and therefore, this substitution does not affect the rate of revenue loss. Figure 3 plots the regrets under the FR policy and the SPA policy over 1,000 sample paths.

We make the following observations from Figure 3. First, although the revenue loss of FR in scenario (a) (Figure 3(a)) is lower than that obtained from applying the SPA policy, the relationship is reversed in scenario (b) (Figure 3(b)). In other words, re-solving the DLP does not always lead to better performance. The intuition behind this result is that, when the ratio r_1/r_2 is large, such as in scenario (b), rejecting a class 2 customer to save the capacity for a potential future class 1 customer is more profitable. The SPA policy accepts every customer from class 1 and rejects all customers from class 2, because the solution to the DLP (without re-solving) is $x_1^* = 1, x_2^* = 0$. This static policy is indeed optimal when $r_1/r_2 \to \infty$. In contrast, the FR policy constantly adjusts accepting probabilities and starts to accept class 2 customers when the actual arrival of class 1 customers falls below its average. Second, we observe from Figure 3 that the revenue losses of both SPA and FR seem to have the same growth rate as horizon length T increases. (It is well known that the revenue loss of SPA is of order $\Theta(\sqrt{T})$ (see Section 2.3 and Proposition EC.2 in Online Appendix EC.1).) This result is in contrast with the work of Jasin and Kumar (2012), which shows that, when the solution to the DLP is nondegenerate, the expected revenue loss of FR is $O(1)$. However, we note that the nondegeneracy assumption made by Jasin and Kumar (2012) does not hold in this example, because the DLP has a unique solution that is degenerate.

Figure 3. (Color online) Regret Under the FR policy (with Re-Solving) and the SPA Policy (Without Re-Solving)



Notes. (a) Scenario (a) when $r_1 = 2$ and $r_2 = 1$. (b) Scenario (b) when $r_1 = 5$ and $r_2 = 1$.

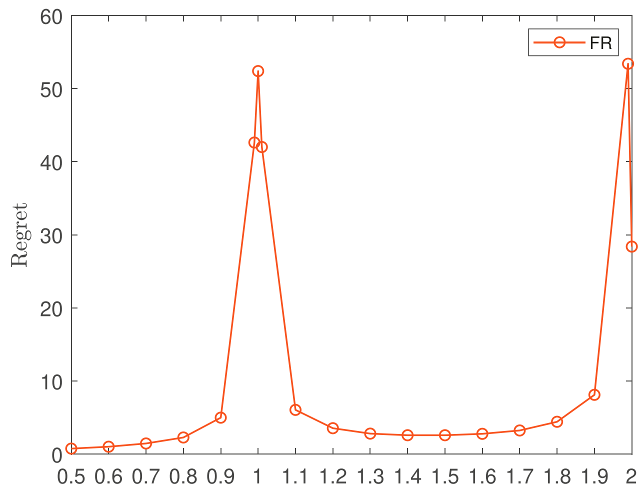
Next, we simulate the FR policy when $r_1 = 2, r_2 = 1$, and $T = 50{,}000$ for varying average capacity per unit time $b = 0.5, \ldots, 2$. Note that, when $b = 1$ and $b = 2$, the optimal solutions to the DLP (2) are $x_1^* = 1, x_2^* = 0$ and $x_1^* = 1, x_2^* = 1$, respectively, which are degenerate according to Equation (4). When $b \neq 1$ and $b \neq 2$, the solution to the DLP is nondegenerate. Therefore, by changing the value of b, we can evaluate the performance of FR with either degenerate or nondegenerate DLP solutions. Figure 4 shows the regret under the FR policy over 1,000 sample paths.

The simulation result from Figure 4 shows that the expected revenue loss under the FR policy is sensitive to the value of capacity rate b. When b is far away from the degenerate points (i.e., $b = 1$ and $b = 2$), FR performs well and has small revenue loss. However, the revenue loss increases significantly when the optimal DLP solution is close to degenerate (e.g., $b = 0.95$).

We notice that the observation from Figure 4 is consistent with the analysis by Jasin and Kumar (2012). Even though Jasin and Kumar (2012) prove that the revenue loss of FR is bounded by a constant whenever the DLP solution is nondegenerate, their analysis does not imply that the constant is uniform over all b values. Rather, the constant bound from their analysis critically depends on the distance between b and its nearest degenerate point. When the optimal DLP solution is close to degenerate, the bound in Jasin and Kumar (2012) can be arbitrarily large. Figure 4 shows that this phenomenon is not merely a consequence of the analysis technique from Jasin and Kumar (2012) but reflects the actual performance of the FR policy.

Let us now turn our attention to the dynamics of the DLP solutions after updating under the FR policy. Figure 5 shows the trajectory of the DLP solutions of

Figure 4. (Color online) Regret Under the FR Policy for $r_1 = 2, r_2 = 1$, and $T = 50{,}000$



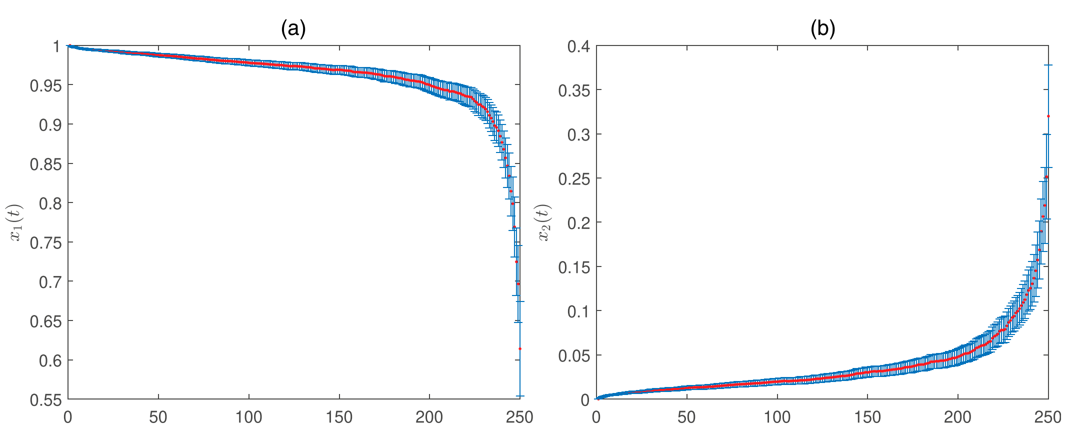
Note. The average capacity per unit time.

FR when $r_1 = 2, r_2 = 1, b = 1$, and $T = 250$. In the figure, we simulated the policy 1,000 times and plotted the median, the 5th percentile, and the 95th percentile of the DLP solutions over the 1,000 sample paths. It can be observed that the DLP solution barely changes near the beginning of the horizon but changes significantly near the end of the horizon. These plots shed some light on the importance of each re-solving time: re-solving near the beginning of the horizon is not as important as re-solving near the end of the horizon.

4. A Re-Solving Heuristic with Uniformly Bounded Loss

In this section, we propose a new re-solving algorithm. The main result of this section shows that

Figure 5. (Color online) The 5th and 95th Percentiles of the DLP Solutions (i.e., Acceptance Probabilities) of the Two Customer Classes Under FR in Each Period for $T = 250$



Notes. (a) Class 1 customer. (b) Class 2 customer.

this algorithm has uniformly bounded revenue loss given any horizon length T and starting capacity C without requiring the nondegeneracy assumption.

4.1. Definition of the IRT Algorithm

We propose an algorithm called IRT. The IRT policy has two distinct features compared with the FR policy: (1) the DLP is not re-solved in every period, and (2) customer acceptance probabilities are adjusted by some thresholds.

Unlike the FR policy, the IRT policy re-solves the DLP for only $O(\log\log T)$ times during a horizon of length T. The re-solving schedule is defined as follows. Given horizon length T, we set $K = \lceil \frac{\log\log T}{\log(6/5)} \rceil$. Let $\{t^*_u, \forall u \in [K]\}$ denote a sequence of re-solving times, where $\tau_u = T^{(5/6)^u}$ and $t^*_u = T - \tau_u$ for all $u \in [K]$. In addition, let $t^*_{K+1} = T$. Thus, the re-solving times divide the entire horizon into $K+1$ epochs: $[0, t^*_1)$, $[t^*_1, t^*_2), \cdots, [t^*_K, t^*_{K+1}]$. Figure 6 illustrates the re-solving schedule of the IRT policy.

At the beginning of each epoch u ($0 \le u \le K$), the algorithm solves an LP approximation to the dynamic programming problem—this LP is identical to the LP used in the FR algorithm, which uses information about remaining capacities and the mean of remaining customer arrivals. The optimal solution of the LP is then used to construct a probabilistic allocation control policy. The IRT policy applies thresholds to the allocation probabilities. In particular, in epoch $u \in \{0\} \cup [K-1]$ (except for the last epoch), the allocation probability for each class is rounded down to 0 if it is less than $\tau_u^{-1/4}$ or rounded up to 1 if it is larger than $1 - \tau_u^{-1/4}$. The complete definition of IRT is given in Algorithm 3.

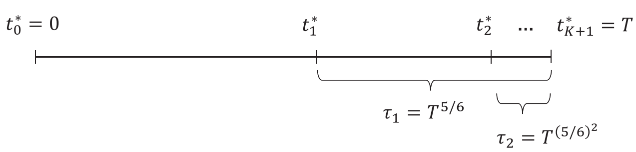
Algorithm 3 (Infrequent Re-Solving with Thresholding)

initialize: set $\tau_u = T^{(5/6)^u}$ and $t^*_u = T - \tau_u$ for all $u \in \{0\} \cup [K]$, where $K = \lceil \frac{\log\log T}{\log(6/5)} \rceil$
for $u = 0, 1, \ldots, K$, **do**
  set $\boldsymbol{x}^u \leftarrow \arg\max_x \{\sum_{j=1}^n r_j x_j \mid \sum_{j=1}^n \boldsymbol{A}_j x_j \le \boldsymbol{C}(t^*_k)/\tau_k$, and $0 \le x_j \le \lambda_j, \forall j \in [n]\}$
  if $u < K$, **then**
    for $j \in [n]$, **do**
      if $x^u_j < \lambda_j \tau_u^{-1/4}$, **then**
        set $p^u_j \leftarrow 0$
      else if $x^u_j > \lambda_j(1 - \tau_u^{-1/4})$, **then**
        set $p^u_j \leftarrow 1$
      else
        set $p^u_j \leftarrow x^u_j/\lambda_j$
      end if
    end for
  else
    set $p^u_j \leftarrow x^u_j/\lambda_j$ for all $j \in [n]$
  end if
  set $\boldsymbol{C}' \leftarrow \boldsymbol{C}(t^*_u)$
  for $t \in [t^*_u, t^*_{u+1})$, **do**
    observe requests from all arrival of customers
    if an arriving customer belongs to class j and $\boldsymbol{A}_j \le \boldsymbol{C}'$ $(\forall j \in [n])$, **then**
      accept the customer with probability p^u_j
      if accepted, update $\boldsymbol{C}' \leftarrow \boldsymbol{C}' - \boldsymbol{A}_j$
    else
      reject the customer
    end if
  end for
  set $\boldsymbol{C}(t^*_{u+1}) \leftarrow \boldsymbol{C}'$
end for

Before we present the formal analysis of the IRT algorithm, it might be helpful to discuss the intuition behind the design of this algorithm. We start with the choice of the first re-solving time, t^*_1. The analysis by Reiman and Wang (2008) shows that, by setting $t^*_1 \approx T - O(\sqrt{T})$, one re-solving of DLP is sufficient to reduce the regret to $O(T^{1/4})$. However, we note that, if t^*_1 is defined as in Reiman and Wang (2008), additional re-optimizations after t^*_1 cannot improve the regret rate. In the IRT algorithm, we choose the first re-solving time to be $t^*_1 = T - T^{5/6}$, which is earlier than the re-solving time in Reiman and Wang (2008). If no additional re-solving is used, this policy leads to a regret rate of $O(T^{5/12})$ (Proposition 1). Even though the $O(T^{5/12})$ rate is worse than the $O(T^{1/4})$ rate in Reiman and Wang (2008), because we choose an earlier re-solving time, more time is left for making additional adjustments. After we establish the $O(T^{5/12})$ regret rate with the first re-solving, we then use induction to prove that subsequent re-optimizations of the DLP can further reduce the regret, eventually reducing it to a constant. By definition, τ_u, the length of epoch u, satisfies the recursive relationship $\tau_{u+1} = \tau_u^{5/6}$, $\forall u \in [K]$. This enables us to apply the induction hypothesis to epochs $u \ge 1$.

The $\tau_u^{-1/4}$ thresholds in the algorithm are critical to bounding the regret. As we have seen from the numerical example in Section 3.1, large losses can occur when the DLP solution is nearly degenerate. If we use a nearly degenerate solution to construct probabilistic allocation controls, some customer classes would have acceptance probabilities that are either very close to zero or very close to one. As a result, the mean number of accepted or rejected customers is dominated by its standard deviation, making the control

Figure 6. The Re-Solving Times of the IRT Policy Are Constructed Recursively



policy ineffective. More specifically, if the acceptance probability of class j customers is $\epsilon \to 0$, the coefficient of variation of the number of customers accepted in one unit time is $1/\sqrt{\lambda_j \epsilon} \to +\infty$. Therefore, if the acceptance probability of a customer class is almost zero, we might as well reject all customers from that class in the current epoch as long as there is sufficient time left to accept customers in the next epoch. Similarly, if the acceptance probability of a customer class is almost one, we might as well accept all customers from that class in the current epoch. This is the intuition behind adding thresholds to the acceptance probabilities in the IRT policy.

4.2. Analysis of the IRT Policy

We now formally analyze the revenue loss (regret) of the IRT policy. The main result of this section is the following.

Theorem 1. *The regret of* IRT *policy defined in Algorithm 3 is bounded by $v^{\mathrm{HO}} - v^{\mathsf{IRT}} = O(1)$. The constant factor depends on the customer arrival rate λ_j ($\forall j \in [n]$), the revenues per customer r_j ($\forall j \in [n]$), and the BOM matrix A; however, this constant is independent of the time horizon T and the capacity vector $\boldsymbol{C}$.*

Theorem 1 states that the regret of IRT policy is $O(1)$. Moreover, this constant is independent of time horizon and capacities, and therefore, the performance of IRT is uniformly bounded when the capacity ratio $\boldsymbol{C}/T$ varies. Because degenerate DLP solution occurs only for some specific capacity ratios, the result in Theorem 1 does not require the nondegeneracy assumption in Jasin and Kumar (2012).

Because the hindsight optimum v^{HO} is an upper bound of the expected revenue of the optimal policy v^*, we immediately get a bound on its revenue loss: $v^* - v^{\mathsf{IRT}} \le v^{\mathrm{HO}} - v^{\mathsf{IRT}} = O(1)$. Moreover, Theorem 1 implied that the hindsight optimum is a tight upper bound, satisfying $v^{\mathrm{HO}} - v^* \le v^{\mathrm{HO}} - v^{\mathsf{IRT}} = O(1)$.

The complete Proof of Theorem 1 can be found in Section A.1. We outline the main idea of the proof here. In the proof, we define a sequence of auxiliary re-solving policies with increasing re-solving frequency. Recall that $K = \lceil \frac{\log\log T}{\log(6/5)} \rceil$ is the number of re-optimizations made by the IRT algorithm. For any $u \in [K]$, we define a policy that follows the IRT heuristic exactly in $[0, t_u^*)$ but then, applies static allocation control in $[t_u^*, T]$. We refer to such a policy as IRT^u. Notice that, when $u = K$, IRT^u coincides with IRT. Similarly, we define HO^u as a policy that is exactly the same as IRT in $[0, t_u^*)$ but applies the hindsight optimal policy in $[t_u^*, T]$. Proof of Theorem 1 depends on the following proposition, which is proved in Section A.2.

Proposition 1. *Given horizon length T, suppose that the first re-solving time is $t_1^* = T - T^{5/6}$. Then,*

1. *the regret of* HO^1 *is* $O(Te^{-\kappa T^{1/6}})$, *and*
2. *the regret of* IRT^1 *is* $O(Te^{-\kappa T^{1/6}}) + O(T^{5/12})$.

Here, we define $\kappa = \frac{\lambda_{min}}{24(\alpha|J_\lambda|+1)^2}$, where $J_\lambda = \{j : x_j^* = \lambda_j\}$ (recall that $\boldsymbol{x}^*$ is the solution to DLP), $\lambda_{\min} = \min_{j\in[n]} \lambda_j$, and α is a positive constant that depends on the BOM matrix A.

Notice that IRT^1 is a nonanticipating and admissible policy, and its regret of $O(T^{5/12})$ is an improvement over the $O(\sqrt{T})$ bound of SPA. The policy HO^1 is not nonanticipating, because it requires access to future arrival information; thus, it is not practical, and its sole purpose is to bound the performance of IRT^1 in the proof.

We then use Proposition 1 to prove Theorem 1 by induction. We illustrate the induction step using IRT^2, a policy that re-solves at $t_1^* = T - T^{5/6}$ and again, at $t_2^* = T - T^{(5/6)^2}$. The regret of IRT^2 can be written as

$$\mathrm{E}\Big[V^{\mathrm{HO}} - V^{\mathsf{IRT}^2}\Big] = \underbrace{\mathrm{E}\Big[V^{\mathrm{HO}} - V^{\mathsf{HO}^1}\Big]}_{(*)} + \underbrace{\mathrm{E}\Big[V^{\mathsf{HO}^1} - V^{\mathsf{HO}^2}\Big]}_{(**)} + \underbrace{\mathrm{E}\Big[V^{\mathsf{HO}^2} - V^{\mathsf{IRT}^2}\Big]}_{(***)}.$$

The term $(*)$ is bounded by $O(Te^{-\kappa T^{1/6}})$ according to Proposition 1. For the term $(**)$, the policies HO^1 and HO^2 are identical up to time t_1^*. Therefore, applying Proposition 1(1) to the subproblem in $(t_1^*, T]$, we get $\mathrm{E}[V^{\mathsf{HO}^1} - V^{\mathsf{HO}^2}] = O(T^{5/6}e^{-\kappa(T^{5/6})^{1/6}}) = O(T^{5/6}e^{-\kappa T^{5/36}})$. For the last term $(***)$, using the well-known result that static probabilistic allocation has a squared root regret, we have $\mathrm{E}[V^{\mathsf{HO}^2} - V^{\mathsf{IRT}^2}] = \mathrm{E}[V^{\mathrm{HO}}(t_2^*, T) - V^{\mathsf{SPA}}(t_2^*, T)] = O(\sqrt{T - t_2^*}) = O(T^{(5/6)^2/2})$. Combining these three terms, we get

$$\begin{aligned} v^{\mathrm{HO}} - v^{\mathsf{IRT}^2} &= O\Big(Te^{-\kappa T^{1/6}}\Big) + O\Big(T^{5/6}e^{-\kappa T^{5/36}}\Big) + O\big(T^{25/72}\big) \\ &= O\big(T^{25/72}\big). \end{aligned}$$

By induction, we show that, if the decision maker re-solves for $K \ge 1$ times, where the uth ($u = 1, \cdots, K$) re-solving time is $t_u^* = T - T^{(5/6)^u}$, the regret is given by

$$\begin{aligned} v^{\mathrm{HO}} - v^{\mathsf{IRT}^K} &= \sum_{u=0}^{K-1} O\Big(T^{(5/6)^u} \exp\Big(-\kappa T^{(5/6)^u/6}\Big)\Big) \\ &\quad + O\Big(T^{(5/6)^K/2}\Big). \end{aligned}$$

When $K = \lceil \frac{\log\log T}{\log(6/5)} \rceil$, the right-hand side of the above equation is bounded by a constant. In additional, the policy IRT^K is the same as IRT, and therefore, we prove that the regret of IRT is $v^{\mathrm{HO}} - v^{\mathsf{IRT}} = O(1)$.

4.3. Revisiting the Degenerate Example in Section 3.1

In Section 3.1, we considered a numerical example with two classes and one resource. We simulated the

FR policy when $r_1 = 2$, $r_2 = 1$, and $T = 50{,}000$ for varying average capacity $b = 0.5, \ldots, 2$, and we showed that FR has poor performance when the DLP solution is either degenerate (i.e., $b = 1$ or $b = 2$) or nearly degenerate. We now test the IRT policy using the same example and compare it with the FR policy. Figure 7 plots the average regret under FR and IRT over 1,000 sample paths.

It can be observed from Figure 7 that the regret under the proposed IRT policy is not sensitive to the average capacity per unit time. This result verifies Theorem 1 in that the regret of IRT is uniformly bounded with respect to the ratio between capacity and time. In contrast, the regret under the FR policy has two spikes that are associated with the two degenerate points ($b = 1$ and $b = 2$).

5. Analysis of the Frequent Re-Solving Policy

5.1. Lower Bound of the Revenue Loss of FR

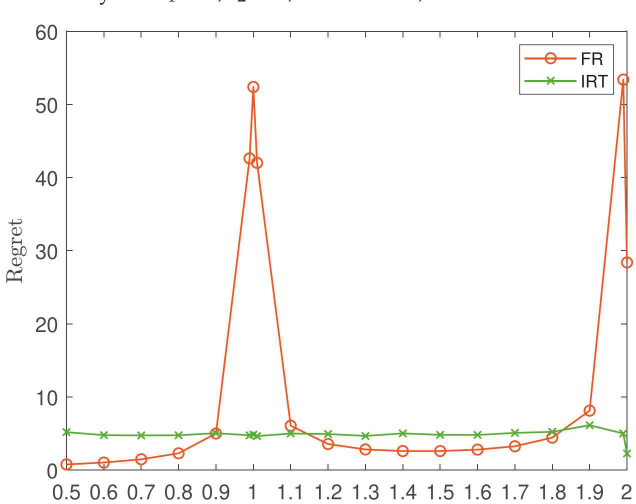
The simulation in Section 3.1 inspires us to analyze the performance of FR without the nondegeneracy assumption in order to gain a better understanding of the effect of frequent re-solving. We first show that the regret under the FR policy is bounded below by $\Omega(\sqrt{T})$.

Proposition 2. *There exists a problem instance for which the regret of the* FR *policy defined in Algorithm 2 is bounded below by*

$$v^{\mathrm{HO}} - v^{\mathrm{FR}} = \Omega\left(\sqrt{T}\right).$$

Proposition 2 implies that the expected revenue loss under FR policy is bounded below by $\Omega(\sqrt{T})$ as well, because the revenue gap between the hindsight

Figure 7. (Color online) Regret Under the FR policy and the IRT Policy for $r_1 = 2, r_2 = 1$, and $T = 50{,}000$



Note. The average capacity per unit time.

optimum (v^{HO}) and the optimal revenue (v^*) is $O(1)$ (Theorem 1). That is, we have

$$\begin{aligned} v^* - v^{\mathrm{FR}} &= -\left(v^{\mathrm{HO}} - v^*\right) + v^{\mathrm{HO}} - v^{\mathrm{FR}} \\ &= -O(1) + \Omega\left(\sqrt{T}\right) = \Omega\left(\sqrt{T}\right). \end{aligned}$$

To prove Proposition 2, we consider a problem instance with two classes of customers and one resource. We assume that customers from each class arrive according to a Poisson process with rate 1; the arrivals from two classes are independent. The initial resource capacity is T. Customers from both classes, if accepted, consume one unit of the resource but pay different prices, $r_1 > r_2$. We consider the event when the number of class 1 customers who arrive during T period is more than T. If this event happens, the hindsight optimum will accept T of class 1 customers and none of class 2 customers. Conditional on that event, we use Freedman's inequality (Freedman 1975) to show that, with positive probability, the FR policy accepts $\Omega(\sqrt{T})$ of class 2 customers and thus, at most, $T - \Omega(\sqrt{T})$ of class 1 customers. Therefore, the revenue of FR is at least $\Omega(\sqrt{T})$ less than the hindsight optimum. The complete proof can be found in Online Appendix EC.2.1.

5.2. Upper Bound of the Revenue Loss of FR

In this section, we provide an upper bound of the expected revenue loss of the FR policy.

Proposition 3. *The gap between the expected revenue of the* FR *policy defined in Algorithm 2 and the optimal value of the DLP is bounded by*

$$v^{\mathrm{DLP}} - v^{\mathrm{FR}} = O\left(\sqrt{T}\right).$$

The constant prefactor depends on the customer arrival rate λ_j ($\forall j \in [n]$), the revenues per customer r_j ($\forall j \in [n]$), and the BOM matrix A; however, it does not depend on the starting capacity C_l ($\forall l \in [m]$).

Because v^{DLP} is an upper bound of the expected revenue under the optimal policy (see Section 2.2.1), Proposition 3 immediately implies that the expected revenue loss of the FR policy when compared with the optimal revenue is bounded by $O(\sqrt{T})$. That is, $v^* - v^{\mathrm{FR}} \le v^{\mathrm{DLP}} - v^{\mathrm{FR}} = O(\sqrt{T})$. Combining Propositions 2 and 3 gives $v^* - v^{\mathrm{FR}} = \Theta(\sqrt{T})$.

The proof of Proposition 3 can be found in Online Appendix EC.2.2. The proof is based on the following idea. Because the LP solved under the FR policy and the DLP (2) only differ in the right-hand side of the capacity constraints, $\boldsymbol{b}(t)$ and $\boldsymbol{b}$, the expected revenue loss of the FR policy when compared with the optimal value of the DLP can be expressed in terms of $\boldsymbol{b}(t)$ and $\boldsymbol{b}$. More specifically, we show that the expected revenue loss during $[t, t+1)$ can be expressed as $O(\mathrm{E}[(b_l - b_l(t))^+])$

for each resource $l \in [m]$. Then, using the relationship between the average remaining capacity, $b(t)$, and the number of accepted customers up to time t, we prove that $O(\mathrm{E}[(b_l - b_l(t))^+]) = O(\frac{1}{\sqrt{T-t}})$. This completes the proof, because $\sum_{t=0}^{T-1} O(\frac{1}{\sqrt{T-t}}) = O(\sqrt{T})$.

Although the $O(\sqrt{T})$ bound in Proposition 3 is looser than the $O(1)$ bound of FR in Jasin and Kumar (2012), it does not require the additional condition that the optimal solution to the DLP is nondegenerate. Given that the expected revenue loss of SPA is also $O(\sqrt{T})$ (see Online Appendix EC.1.2), we conclude that re-solving at least guarantees the same order of revenue loss compared with no re-solving.

6. Numerical Experiment

In this section, we evaluate the numerical performance of five different heuristics, which include the following.

1. SPA: static probabilistic allocation heuristic (Algorithm 1)
2. FR: frequent re-solving heuristic (Algorithm 2)
3. IRT: infrequent re-solving with thresholding (Algorithm 3)
4. Infrequent re-solving (IR): this algorithm uses the same re-solving schedule as IRT but without applying thresholding (i.e., the acceptance probability at iteration u is always set to $p_j^u \leftarrow x_j^u/\lambda_j$)
5. Frequent re-solving with thresholding (FRT): this algorithm is motivated by IRT; we apply the same $\tau^{-1/4}$ thresholds from IRT to the frequent re-solving algorithm (see the complete description in Algorithm 4)

Algorithm 4 (Frequent Re-Solving with Thresholding)

initialize: set $\boldsymbol{C}(0) = \boldsymbol{C}$ and $\boldsymbol{b}(0) = \boldsymbol{C}/T$
for $t = 0, 1, \ldots, T-1$, **do**
  set $\boldsymbol{x}(t) \leftarrow \arg\max_{\boldsymbol{x}} \{\sum_{j=1}^n r_j x_j \mid \sum_{j=1}^n \boldsymbol{A}_j x_j \le \boldsymbol{b}(t),$ and $0 \le x_j \le \lambda_j, \forall j \in [n]\}$
  set $\boldsymbol{C}' \leftarrow \boldsymbol{C}(t)$
  for all customers arriving in $[t, t+1)$, **do**
    if the customer belongs to class j and $\boldsymbol{A}_j \le \boldsymbol{C}'$ $(\forall j \in [n])$, **then**
      if $x_j(t) < \lambda_j(T-t)^{-1/4}$, **then**
        reject the customer
      else if $x_j(t) > \lambda_j(1-(T-t)^{-1/4})$, **then**
        accept the customer
      else
        accept the customer with probability $x_j(t)/\lambda_j$
      end if
      if the customer is accepted, update $\boldsymbol{C}' \leftarrow \boldsymbol{C}' - \boldsymbol{A}_j$
    else
      reject the customer
    end if
  end for
  set $\boldsymbol{C}(t+1) \leftarrow \boldsymbol{C}'$ and $\boldsymbol{b}(t+1) \leftarrow \frac{\boldsymbol{C}(t+1)}{T-t-1}$
end for

Recall that IRT has two distinct features compared with FR: It uses an infrequent re-solving schedule and adds thresholds for acceptance probabilities. The motivation to include IR and FRT in this test is to evaluate which of the two features plays a more important role.

6.1. Single Resource

We consider a revenue management problem with a single resource and two classes of customers. We assume that customers from each class arrive according to a Poisson process with rate 1. The arrivals of two classes are independent. Customers from both classes, if accepted, consume one unit of resource but pay different prices, r_1 and r_2. We consider two cases: (1) $r_1 = 2, r_2 = 1$ and (2) $r_1 = 5, r_2 = 1$. We also test three settings for the average capacity per unit time: $b = 12$, 1.1, and 1.5. When the average capacity per unit time is one, the solution to the DLP is degenerate. The scenario where the average capacity is 1.1 represents a setting where the DLP solution is "nearly degenerate," and the scenario of 1.5 represents a setting where the DLP solution is far away from any degenerate point. We simulate the heuristics for two price and three average capacity per unit time scenarios defined above and for varying horizon length $T = 500, 1{,}000, \ldots, 5{,}000$.

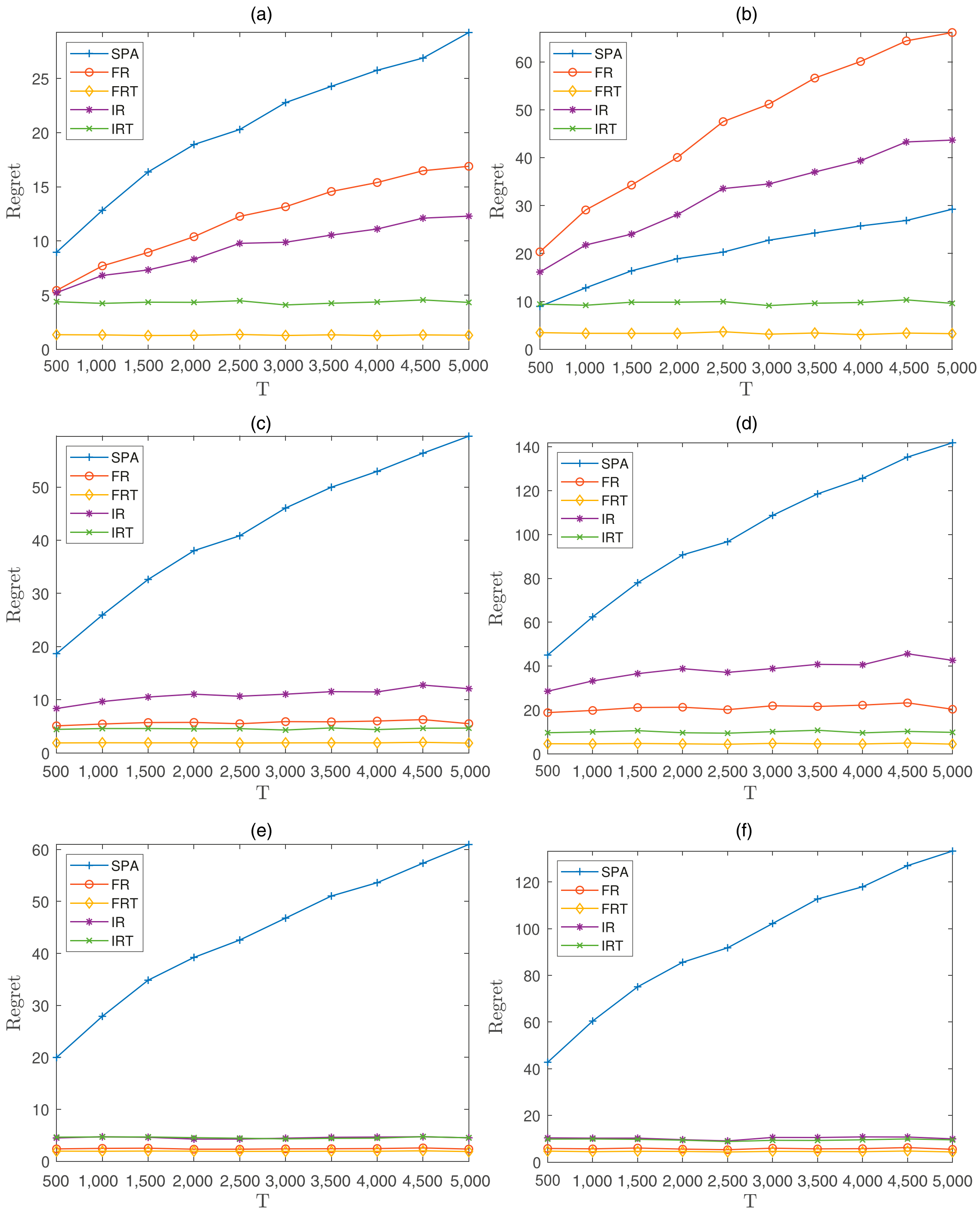
Figure 8 plots the regret under SPA, FR, FRT, IR, and IRT over 1,000 sample paths. Figure 8, (a), (c), and (e) shows the case when $r_1 = 2$ and $r_2 = 1$, whereas Figure 8, (b), (d), and (f) shows the case when $r_1 = 5$ and $r_2 = 1$. Figure 8, (a) and (b) illustrates the case when $b = 1$. Figure 8, (c) and (d) illustrates the case when $b = 1.1$. Figure 8, (e) and (f) illustrates the case when $b = 1.5$. We make the following observations.

1. When $r_1 = 2$ and $r_2 = 1$, the expected revenue loss under SPA is the largest for all average capacity per unit time and horizon length. This does not hold when $r_1 = 5$, $r_2 = 1$, and $b = 1$, where SPA is better than both FR and IR.
2. The expected revenue loss under IR is higher than the expected revenue loss under FR, except when the problem is degenerate ($b = 1$). We conclude that choosing an infrequent re-solving schedule alone is not enough to achieve $O(1)$ loss.
3. The expected revenue losses under IRT and FRT remain constant for all cases as the horizon length increases. Moreover, although we do not have theoretical guarantee for FRT, the expected revenue loss under FRT often seems smaller than the expected revenue loss under IRT. This implies that appropriate thresholding is the main factor that leads to uniformly bounded regret for re-solving heuristics.

6.2. Multiple Resources

Next, we consider a network revenue management problem with multiple resources. We consider the problem when there are five classes of customers

Figure 8. (Color online) Regret Under the SPA, the FR, the FRT, the IR, and the IRT Policies for $T = 500, 1{,}000, \ldots, 5{,}000$



Notes. (a) When $b = 1$, $r_1 = 2$, and $r_2 = 1$; (b) when $b = 1$, $r_1 = 5$, and $r_2 = 1$; (c) when $b = 1.1$, $r_1 = 2$, and $r_2 = 1$; (d) when $b = 1.1$, $r_1 = 5$, and $r_2 = 1$; (e) when $b = 1.5$, $r_1 = 2$, and $r_2 = 1$; (f) when $b = 1.5$, $r_1 = 5$, and $r_2 = 1$.

and four types of resources. We assume that customers from each class arrive according to a Poisson process with rate 1; the arrivals of different classes are independent. The vector of the average capacities per unit time is given by $\boldsymbol{b} = [1, 1, 1, 1]^\top$. The vector of the revenue earned by accepting customers is given by $\boldsymbol{r} = [10, 3, 6, 1, 2]^\top$. The bill-of-materials matrix is given by

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We simulate the heuristics for varying horizon length $T = 500, 1{,}000, \ldots, 5{,}000$. Notice that, in this example, the optimal solution to the DLP is degenerate.

Figure 9 plots the regrets under SPA, FR, FRT, IR, and IRT over 1,000 sample paths. The result shows that the revenue losses of SPA scale poorly with horizon length T. In comparison, the revenue losses of FR and IR increase more slowly when T increases, and IR seems to perform slightly better for large T. The revenue losses of FRT and IRT remain constant as T grows. Moreover, the expected revenue loss under the IRT is higher than the expected revenue loss under the FRT. Again, this result implies that, among the two factors infrequent re-solving and thresholding, the latter plays a more important role.

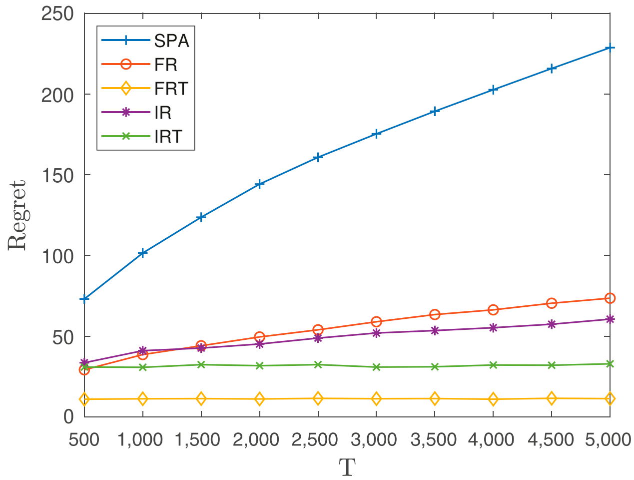
7. Conclusion and Discussion

We study re-solving heuristics for the NRM problem. A re-solving heuristic periodically re-optimizes a deterministic LP approximation of the original NRM problem. The main question considered in this paper is as follows: can we find a simple and computationally efficient re-solving heuristic with expected revenue loss compared with the optimal policy that is bounded by a constant even when both the time horizon and the resource capacities scale up?

We answer the above question in the affirmative by proposing a re-solving heuristic called IRT, in which revenue loss is bounded by a constant independent of time horizon and resource capacities. This finding improves a previous result by Jasin and Kumar (2012), showing that FR, an algorithm that re-solves the DLP after each unit of time, has $O(1)$ revenue loss but requires the optimal solution to the DLP to be nondegenerate for the quantity-based NRM problem. Moreover, we show that, when both time horizon and resource capacities scale up by $k = 1, 2, \ldots$, FR has a revenue loss of $\Theta(\sqrt{k})$. This is a negative result, because most DLP-based heuristics can achieve the same revenue loss rate without using re-solving at all. However, we note that the analysis by Jasin and Kumar (2012) considers customer choice behavior, whereas we focus on the classical quantity-based problem with independent demand model throughout this paper. Atar and Reiman (2012) and Jasin (2014) also consider re-solving heuristics for the price-based NRM problem. It is unclear if one can extend our analysis technique to either the NRM problem with customer choice or the price-based NRM problem, because our proof is based on the notion of hindsight optimum (Reiman and Wang 2008), which is not well defined for the NRM problem with customer choice or the price-based NRM problem. We will leave this question for future research.

Our simulation results show that, when the controls from FR are adjusted by some thresholds, the resulting algorithm FRT has very promising numerical performance and seems to have a bounded revenue loss as well. So far, we are not able to prove this result, mainly because the induction-based proof that we developed for IRT breaks down when the DLP is re-optimized every period. Recently, Vera and Banerjee (2019) proposed a different re-solving heuristic for the NRM problem, where the DLP is re-optimized every period and a fixed acceptance probability threshold of 0.5 is applied to every class for all periods. They show that their heuristic also achieves $O(1)$ regret. Although the fixed threshold used by Vera and Banerjee (2019) is different from the time-varying thresholds that we proposed in the FRT algorithm, we think that their analysis technique may be helpful to establish the revenue loss bound of FRT.

Figure 9. (Color online) Regret Under the SPA, the FR, the FRT, the IR, and the IRT Compared with the Hindsight Optimal for $T = 500, 1{,}000, \ldots, 5{,}000$



Appendix. Proofs for Section 4

In this section, we provide complete proofs for the results on the IRT policy in Section 4.

A.1. Proof of Theorem 1

Proof of Theorem 1. Given remaining capacity $\boldsymbol{C}(t_1)$ at time $t_1 \in [0, T]$, let $\boldsymbol{x}(t_1)$ be an optimal solution to the following LP:

$$\max_{\boldsymbol{x}} \left\{ \sum_{j=1}^{n} r_j x_j \,\middle|\, \sum_{j=1}^{n} \boldsymbol{A}_j x_j \le \boldsymbol{C}(t_1)/(T - t_1), \text{ and } 0 \le x_j \le \lambda_j, \forall j \in [n] \right\}.$$

For any $t_2 \in (t_1, T]$, let $V^{\mathsf{SPA}}(t_1, t_2)$ denote the revenue earned in $[t_1, t_2)$ under a static probabilistic allocation policy, where class j customers are accepted with probability $x_j(t_1)/\lambda_j$. Let $V^{\mathsf{SPA}'}(t_1, t_2)$ be the revenue earned in $[t_1, t_2)$, where a class j customer is accepted with the following probability.

- 0 if $x_j(t_1) < \lambda_j(T - t_1)^{-1/4}$
- 1 if $x_j(t_1) > \lambda_j(1 - (T - t_1)^{-1/4})$
- $x_j(t_1)/\lambda_j$ otherwise

Let $V^{\mathrm{HO}}(t_1, T)$ denote the revenue earned from solving the hindsight optimum in $[t_1, T]$. That is, $V^{\mathrm{HO}}(t_1, T)$ is the optimal revenue given the remaining capacity at t_1 and a sample path of demand in $(t_1, T]$ given by

$$V^{\mathrm{HO}}(t_1, T) = \max_{y}\left\{\sum_{j=1}^{n} r_j y_j \;\middle|\; \sum_{j=1}^{n} \boldsymbol{A}_j y_j \le \boldsymbol{C}(t_1), \text{ and } 0 \le y_j \le \Lambda_j(T) - \Lambda_j(t_1), \forall j \in [n]\right\}.$$

Consider policy IRT^2, which re-solves at $t_1^* = T - T^{5/6}$ and re-solves again at $t_2^* = T - T^{(5/6)^2} = T - T^{25/36}$. Let v^{IRT^2} be the expected revenue of IRT^2. The regret of this policy can be decomposed as

$$\begin{aligned} v^{\mathrm{HO}} - v^{\mathsf{IRT}^2} &= \mathrm{E}\left[V^{\mathrm{HO}} - V^{\mathsf{IRT}^2}\right] \\ &= \mathrm{E}\left[V^{\mathrm{HO}} - V^{\mathsf{HO}^1}\right] + \mathrm{E}\left[V^{\mathsf{HO}^1} - V^{\mathsf{HO}^2}\right] \\ &\quad + \mathrm{E}\left[V^{\mathsf{HO}^2} - V^{\mathsf{IRT}^2}\right]. \end{aligned} \tag{A.1}$$

The first term of Equation (A.1) is bounded by $O(Te^{-\kappa T^{1/6}})$ as stated in Proposition 1. For the second term of Equation (A.1), the revenues of policies HO^1 and HO^2 are

$$\begin{aligned} V^{\mathsf{HO}^1} &= V^{\mathsf{SPA}'}\left(0, t_1^*\right) + V^{\mathrm{HO}}(t_1^*, T), \\ V^{\mathsf{HO}^2} &= V^{\mathsf{SPA}'}\left(0, t_1^*\right) + V^{\mathsf{SPA}'}\left(t_1^*, t_2^*\right) + V^{\mathrm{HO}}\left(t_2^*, T\right). \end{aligned}$$

Note that policies HO^1 and HO^2 are exactly the same during time $t \in [0, t_1^*)$, and therefore, we have

$$V^{\mathsf{HO}^1} - V^{\mathsf{HO}^2} = V^{\mathrm{HO}}\left(t_1^*, T\right) - \left(V^{\mathsf{SPA}'}\left(t_1^*, t_2^*\right) + V^{\mathrm{HO}}\left(t_2^*, T\right)\right).$$

Applying Proposition 1(1) to the remaining problem in $(t_1^*, T]$, which has a horizon length $T - t_1^* = \tau_1 = T^{5/6}$, we get

$$\begin{aligned} &\mathrm{E}\left[V^{\mathrm{HO}}\left(t_1^*, T\right) - \left(V^{\mathsf{SPA}'}\left(t_1^*, t_2^*\right) + V^{\mathrm{HO}}\left(t_2^*, T\right)\right)\right] \\ &\quad = O\left(T^{5/6} e^{-\kappa (T^{5/6})^{1/6}}\right) = O\left(T^{5/6} e^{-\kappa T^{5/36}}\right). \end{aligned} \tag{A.2}$$

Because the revenue of IRT^2 can be decomposed as

$$V^{\mathsf{IRT}^2} = V^{\mathsf{SPA}'}\left(0, t_1^*\right) + V^{\mathsf{SPA}'}\left(t_1^*, t_2^*\right) + V^{\mathsf{SPA}}\left(t_2^*, T\right),$$

the last term of (A.1) can be bounded by

$$\begin{aligned} \mathrm{E}\left[V^{\mathsf{HO}^2} - V^{\mathsf{IRT}^2}\right] &= \mathrm{E}\left[V^{\mathrm{HO}}\left(t_2^*, T\right) - V^{\mathsf{SPA}}\left(t_2^*, T\right)\right] \\ &= O\left(\sqrt{T - t_2^*}\right) \\ &= O\left(T^{(5/6)^2/2}\right). \end{aligned} \tag{A.3}$$

Equation (A.3) follows the well-known result that static probabilistic allocation has a regret of $O(\sqrt{k})$ for a problem with horizon length k (see Online Appendix EC.1.2).

Combining (A.2) and (A.3), Equation (A.1) is bounded by

$$v^{\mathrm{HO}} - v^{\mathsf{IRT}^2} = O\left(Te^{-\kappa T^{1/6}}\right) + O\left(T^{5/6} e^{-\kappa T^{5/36}}\right) + O\left(T^{25/72}\right). \tag{A.4}$$

Now, consider policy IRT^3, which follows IRT^2 during $t \in [0, t_3^*)$ but re-solves again at time $t_3^* = T - T^{(5/6)^3}$. By the same decomposition argument, the expected regret is given by

$$\begin{aligned} v^{\mathrm{HO}} - v^{\mathsf{IRT}^3} &= \mathrm{E}\left[V^{\mathrm{HO}} - V^{\mathsf{HO}^1}\right] + \mathrm{E}\left[V^{\mathsf{HO}^1} - V^{\mathsf{HO}^2}\right] \\ &\quad + \mathrm{E}\left[V^{\mathsf{HO}^2} - V^{\mathsf{HO}^3}\right] + \mathrm{E}\left[V^{\mathsf{HO}^3} - V^{\mathsf{IRT}^3}\right] \\ &= O\left(Te^{-\kappa T^{1/6}}\right) + O\left(T^{5/6} e^{-\kappa T^{5/36}}\right) \\ &\quad + O\left(T^{25/36} e^{-\kappa T^{25/216}}\right) + O\left(T^{(5/6)^3/2}\right). \end{aligned}$$

Let $K = \lceil \frac{\log\log T}{\log(6/5)} \rceil$. Note that the policy IRT^K coincides with IRT. By induction, if the decision maker re-solves K times, where the uth re-solving time is $t_u^* = T - T^{(5/6)^u}$, the regret is given by

$$\begin{aligned} v^{\mathrm{HO}} - v^{\mathsf{IRT}} &= v^{\mathrm{HO}} - v^{\mathsf{IRT}^K} \\ &= \sum_{u=0}^{K-1} O\left(T^{(5/6)^u} \exp\left(-\kappa T^{(5/6)^u/6}\right)\right) + O\left(T^{(5/6)^K/2}\right). \end{aligned} \tag{A.5}$$

For the first term of the right-hand side of Equation (A.5), using the fact that $T^{(5/6)^K} \le e$, we have

$$\begin{aligned} \sum_{u=0}^{K-1} T^{(5/6)^u} \exp\left(-\kappa T^{(5/6)^u/6}\right) &= \sum_{\ell=1}^{K} T^{(5/6)^{K-\ell}} \exp\left(-\kappa T^{(5/6)^{K-\ell}/6}\right) \\ &\le \sum_{\ell=1}^{K} e^{(6/5)^\ell} \exp\left(-\kappa e^{(6/5)^\ell/6}\right) \\ &\le \sum_{\ell=1}^{\infty} e^{(6/5)^\ell} \exp\left(-\kappa e^{(6/5)^\ell/6}\right) \\ &\le \int_0^\infty x \exp\left(-\kappa x^{1/6}\right) = O(1). \end{aligned}$$

Thus, the first term of the right-hand side of Equation (A.5) is $O(1)$. By the definition of constant K, we have $T^{(5/6)^K/2} \le e^{1/2}$. Thus, the second term in (A.5) is also $O(1)$. Therefore, we have $v^{\mathrm{HO}} - v^{\mathsf{IRT}} = O(1)$. In addition, this constant factor is independent of the time horizon T and the capacity vector $\boldsymbol{C}$. □

A.2. Proof of Proposition 1

Throughout this subsection, we focus on policies IRT^1 and HO^1 with only one re-solving at $t_1^* = T - T^{5/6}$. We write $t^* := t_1^*$ for simplicity. Let $\Gamma(T) = \alpha \sum_{j: x_j^* = \lambda_j} |\Lambda_j(T) - \lambda_j T|$, where α is a constant with a value that is determined by the BOM matrix $A = (a_{lj})_{l \in [m], j \in [n]}$. More specifically, α is the maximum absolute value of the elements in the inverses of all invertible submatrices of the BOM matrix A. In a special case when all entries of A are either 0 or 1, we have $\alpha \le \max\{1, m \wedge n - 1\}$. We let $\Delta_j(t)$ be the deviation of the number of arrivals of class j customers from its mean in $(t, T]$ (i.e., $\Delta_j(t) = \Lambda_j(T) - \Lambda_j(t) - \lambda_j(T - t)$). Define $\tilde{z}_j(t)$ as the number of class j customers accepted up to time t if the algorithm was allowed to go over the capacity limits. For all $j \in [n]$, we define the following events:

$$E_{1,j} = \left\{(T - t^*)x_j^* - \tilde{z}_j(t^*) + t^* x_j^* \geq \Gamma(T)\right\}, \tag{A.6}$$

$$E_{2,j} = \left\{(T - t^*)\left(\lambda_j - x_j^*\right) + \tilde{z}_j(t^*) - t^* x_j^* \geq \Gamma(T) + \left|\Delta_j(t^*)\right|\right\}. \tag{A.7}$$

The event E is defined as

$$E = \left(\bigcap_{j:x_j^* \geq \lambda_j T^{-1/4}} E_{1,j}\right) \cap \left(\bigcap_{j:x_j^* \leq \lambda_j(1-T^{-1/4})} E_{2,j}\right). \tag{A.8}$$

Now, we will prove Proposition 1.

Proof of Proposition 1. For all $j : x_j^* < \lambda_j T^{-1/4}$, $\tilde{z}_j(t^*) = 0 \leq Tx_j^*$. For all $j : x_j^* \geq \lambda_j T^{-1/4}$, event E in (A.8) implies $\tilde{z}_j(t^*) \leq (T - t^*)x_j^* + t^* x_j^* = Tx_j^*$. Therefore, suppose that event E holds, the capacity constraints for all resources are satisfied up to period t^*, and we have $z_j(t^*) = \tilde{z}_j(t^*)$.

If $x_j^* < \lambda_j T^{-1/4}$, we have $\bar{z}_j - z_j(t^*) = \bar{z}_j - 0 \geq 0$. Recall that $\bar{z}_j$ is the solution to the hindsight optimum (see Section 2.2.2). Otherwise, suppose that event E holds; by Lemma EC.6 in Online Appendix EC.3, we have

$$\begin{aligned}\bar{z}_j - z_j(t^*) &\geq Tx_j^* - \Gamma(T) - z_j(t^*) + t^* x_j^* - t^* x_j^* \\ &= (T - t^*)x_j^* - \Gamma(T) - \left(z_j(t^*) - t^* x_j^*\right) \geq 0,\end{aligned} \tag{A.9}$$

where (A.9) follows from condition (A.6) and the fact that $z_j(t^*) = \tilde{z}_j(t^*)$.

Similarly, if $x_j^* > \lambda_j(1 - T^{-1/4})$, we have $z_j(t^*) + \Lambda_j(T) - \Lambda_j(t^*) - \bar{z}_j \geq \Lambda_j(t^*) + \Lambda_j(T) - \Lambda_j(t^*) - \Lambda_j(T) = 0$. Otherwise, suppose that event E holds; by Lemma EC.6, we have

$$\begin{aligned}&z_j(t^*) + \Lambda_j(T) - \Lambda_j(t^*) - \bar{z}_j \\ &\quad = z_j(t^*) + (T - t^*)\lambda_j + \Delta_j(t^*) - \bar{z}_j \\ &\quad \geq z_j(t^*) + (T - t^*)\lambda_j + \Delta_j(t^*) - Tx_j^* - \Gamma(T) \\ &\quad = \left(z_j(t^*) - t^* x_j^*\right) + (T - t^*)\left(\lambda_j - x_j^*\right) + \Delta_j(t^*) - \Gamma(T) \\ &\quad \geq (T - t^*)\left(\lambda_j - x_j^*\right) + \left(z_j(t^*) - t^* x_j^*\right) - \left|\Delta_j(t^*)\right| - \Gamma(T) \\ &\quad \geq 0,\end{aligned} \tag{A.10}$$

where (A.10) follows from condition (A.7) and the fact that $z_j(t^*) = \tilde{z}_j(t^*)$.

Therefore, combining (A.9) and (A.10), we have $z_j(t^*) \leq \bar{z}_j \leq z_j(t^*) + \Lambda_j(T) - \Lambda_j(t^*)$. In other words, the decision maker would still be able to achieve the hindsight optimum if she uses probabilistic allocation up to t^* and then, gets perfect information from t^* onward, because she can accept $\bar{z}_j - z_j(t^*)$ of class j customers. If the decision maker re-solves once at t^*, then the regret can be written as

$$\begin{aligned}v^{\mathrm{HO}} - v^{\mathsf{IRT}^1} &= \mathrm{E}\left[V^{\mathrm{HO}} - V^{\mathsf{IRT}^1}\right] \\ &= \mathrm{E}\left[V^{\mathrm{HO}} - V^{\mathsf{HO}^1}\right] + \mathrm{E}\left[V^{\mathsf{HO}^1} - V^{\mathsf{IRT}^1}\right],\end{aligned} \tag{A.11}$$

where we can decompose the revenue earned under the policy HO^1 and IRT^1 as

$$\begin{aligned}V^{\mathsf{HO}^1} &= V^{\mathsf{SPA}'}(0, t^*) + V^{\mathrm{HO}}(t^*, T), \\ V^{\mathsf{IRT}^1} &= V^{\mathsf{SPA}'}(0, t^*) + V^{\mathsf{SPA}}(t^*, T).\end{aligned}$$

Consequently, we get

$$\begin{aligned}\mathrm{E}\left[V^{\mathsf{HO}^1} - V^{\mathsf{IRT}^1}\right] &= \mathrm{E}\left[V^{\mathrm{HO}}(t^*, T) - V^{\mathsf{SPA}}(t^*, T)\right] \\ &= O\left(\sqrt{T - t^*}\right) = O(T^{5/12}).\end{aligned} \tag{A.12}$$

Equation (A.12) follows the well-known result that static probabilistic allocation without re-solving has a regret rate of $O(\sqrt{k})$ for a problem with horizon length k, where the constant factor does not depend on the capacity vector C (e.g., Reiman and Wang 2008). For completeness, we give a proof of this result in Online Appendix EC.1.

Recall that, if the event E happens, the hindsight optimal is still attainable starting from t^*. In other words, conditioned on E, the regret of HO^1 is $V^{\mathrm{HO}} - V^{\mathsf{HO}^1} = 0$. Therefore, the first term of (A.11) is given by

$$\begin{aligned}\mathrm{E}\left[V^{\mathrm{HO}} - V^{\mathsf{HO}^1}\right] &= \mathrm{E}\left[V^{\mathrm{HO}} - V^{\mathsf{HO}^1}\middle| E\right]\mathbb{P}(E) \\ &\quad + \mathrm{E}\left[V^{\mathrm{HO}} - V^{\mathsf{HO}^1}\middle| E^c\right]\mathbb{P}(E^c) \\ &= \mathrm{E}\left[V^{\mathrm{HO}} - V^{\mathsf{HO}^1}\middle| E^c\right]\mathbb{P}(E^c) \\ &\leq \mathrm{E}\left[V^{\mathrm{HO}}\middle| E^c\right]\mathbb{P}(E^c).\end{aligned} \tag{A.13}$$

Note that the hindsight optimum is bounded almost surely by $V^{\mathrm{HO}} \leq \sum_{j=1}^n r_j \Lambda_j(T)$, where $\Lambda_j(T)$ is the total number of arrivals from class j. Moreover, $\Lambda_j(T)$ follows Poisson distribution with mean $\lambda_j T$. By the Poisson tail bound (see Lemma EC.5 in Online Appendix EC.3), we have

$$\begin{aligned}\mathrm{E}[(\Lambda_j(T) - 2\lambda_j T)^+] &= \int_0^\infty \mathbb{P}(\Lambda_j(T) - 2\lambda_j T \geq x)dx \\ &\leq \int_0^\infty 2\exp\left(-\frac{(x + \lambda_j T)^2}{2(x + 2\lambda_j T)}\right)dx \\ &= \int_0^{\lambda_j T} 2\exp\left(-\frac{(x + \lambda_j T)^2}{2(x + 2\lambda_j T)}\right)dx \\ &\quad + \int_{\lambda_j T}^\infty 2\exp\left(-\frac{(x + \lambda_j T)^2}{2(x + 2\lambda_j T)}\right)dx \\ &\leq \int_0^{\lambda_j T} 2\exp\left(-\frac{(x + \lambda_j T)^2}{2(\lambda_j T + 2\lambda_j T)}\right)dx \\ &\quad + \int_{\lambda_j T}^\infty 2\exp\left(-\frac{(x + \lambda_j T)^2}{2(x + 2x)}\right)dx \\ &\leq \int_0^{\lambda_j T} 2\exp\left(-\frac{(x + \lambda_j T)\lambda_j T}{6\lambda_j T}\right)dx \\ &\quad + \int_{\lambda_j T}^\infty 2\exp\left(-\frac{(x + \lambda_j T)x}{6x}\right)dx \\ &\leq \int_0^\infty 2\exp\left(-\frac{(x + \lambda_j T)}{6}\right)dx \\ &= 12\exp\left(-\frac{\lambda_j T}{6}\right).\end{aligned}$$

Combining the above inequality with Equation (A.13), we have

$$\begin{aligned}
&\mathrm{E}\Big[V^{\mathrm{HO}} - V^{\mathrm{HO}^1}\Big] \\
&\quad \le \mathrm{E}\Big[V^{\mathrm{HO}}\Big|E^c\Big]\mathbb{P}(E^c) \\
&\quad \le \mathrm{E}\left[\sum_{j=1}^n r_j\Lambda_j(T)\middle|E^c\right]\mathbb{P}(E^c) \\
&\quad \le \mathrm{E}\left[\sum_{j=1}^n r_j\big(2\lambda_jT + (\Lambda_j(T) - 2\lambda_jT)^+\big)\middle|E^c\right]\mathbb{P}(E^c) \\
&\quad \le \sum_{j=1}^n r_j\big(2\lambda_jT\mathbb{P}(E^c) + \mathrm{E}\big[(\Lambda_j(T) - 2\lambda_jT)^+|E^c\big]\mathbb{P}(E^c)\big) \\
&\quad \le \sum_{j=1}^n r_j\big(2\lambda_jT\mathbb{P}(E^c) + \mathrm{E}\big[(\Lambda_j(T) - 2\lambda_jT)^+\big]\big) \\
&\quad \le \sum_{j=1}^n r_j\big(2\lambda_jT\mathbb{P}(E^c) + 12e^{-\lambda_jT/6}\big).
\end{aligned} \tag{A.14}$$

Using the result from Lemma EC.7 in Online Appendix EC.3 as well as (A.13) and (A.14), we have

$$\begin{aligned}
\mathrm{E}\Big[V^{\mathrm{HO}} - V^{\mathrm{HO}^1}\Big] &\le \sum_{j=1}^n r_j\Big(2\lambda_jT\cdot O\Big(e^{-\kappa T^{1/6}}\Big) + 12e^{-\lambda_jT/6}\Big) \\
&= O\Big(Te^{-\kappa T^{1/6}}\Big).
\end{aligned} \tag{A.15}$$

We can conclude from (A.11), (A.12), and (A.15) that

$$v^{\mathrm{HO}} - v^{\mathrm{IRT}^1} = O\Big(Te^{-\kappa T^{1/6}}\Big) + O\big(T^{5/12}\big),$$

where the big O notation hides constants that are independent of the time horizon T and the capacity vector $\boldsymbol{C}$. □

References

Arlotto A, Gurvich I (2019) Uniformly bounded regret in the multisecretary problem. *Stochastic Systems* 9(3):231–260.

Arlotto A, Xie X (2019) Logarithmic regret in the dynamic and stochastic knapsack problem. Working paper, Duke University, Durham, NC.

Atar R, Reiman MI (2012) Asymptotically optimal dynamic pricing for network revenue management. *Stochastic Systems* 2(2): 232–276.

Babaioff M, Immorlica N, Kempe D, Kleinberg R (2007) A knapsack secretary problem with applications. Charikar M, Jansen K, Reingold O, Rolim JDP, eds. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, Lecture Notes in Computer Science, vol. 4627 (Springer-Verlag, Berlin), 16–28.

Bertsekas DP (2005) *Dynamic Programming and Optimal Control*, vol. 1, 3rd ed. (Athena Scientific, Belmont, MA).

Besbes O, Zeevi A (2012) Blind network revenue management. *Oper. Res.* 60(6):1537–1550.

Chen L, Homem-de Mello T (2010) Re-solving stochastic programming models for airline revenue management. *Ann. Oper. Res.* 177(1):91–114.

Cooper WL (2002) Asymptotic behavior of an allocation policy for revenue management. *Oper. Res.* 50(4):720–727.

Ferreira KJ, Simchi-Levi D, Wang H (2018) Online network revenue management using Thompson sampling. *Oper. Res.* 66(6): 1586–1602.

Freedman DA (1975) On tail probabilities for martingales. *Ann. Probab.* 3(1):100–118.

Gallego G, van Ryzin G (1994) Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* 40(8):999–1020.

Gallego G, van Ryzin G (1997) A multiproduct dynamic pricing problem and its applications to network yield management. *Oper. Res.* 45(1):24–41.

Jasin S (2014) Reoptimization and self-adjusting price control for network revenue management. *Oper. Res.* 62(5):1168–1178.

Jasin S (2015) Performance of an LP-based control for revenue management with unknown demand parameters. *Oper. Res.* 63(4): 909–915.

Jasin S, Kumar S (2012) A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Math. Oper. Res.* 37(2):313–345.

Jasin S, Kumar S (2013) Analysis of deterministic LP-based booking limit and bid price controls for revenue management. *Oper. Res.* 61(6):1312–1320.

Kleinberg R (2005) A multiple-choice secretary algorithm with applications to online auctions. *Proc. 16th Annual ACM-SIAM Sympos. Discrete Algorithms* (SIAM, Philadelphia), 630–631.

Kleywegt AJ, Papastavrou JD (1998) The dynamic and stochastic knapsack problem. *Oper. Res.* 46(1):17–35.

Liu Q, van Ryzin G (2008) On the choice-based linear programming model for network revenue management. *Manufacturing Service Oper. Management* 10(2):288–310.

Maglaras C, Meissner J (2006) Dynamic pricing strategies for multiproduct revenue management problems. *Manufacturing Service Oper. Management* 8(2):136–148.

Reiman MI, Wang Q (2008) An asymptotically optimal policy for a quantity-based network revenue management problem. *Math. Oper. Res.* 33(2):257–282.

Revuz D, Yor M (2013) *Continuous Martingales and Brownian Motion*, Grundlehren der mathematischen Wissenschaften, vol. 293 (Springer Science & Business Media, Berlin).

Secomandi N (2008) An analysis of the control-algorithm re-solving issue in inventory and revenue management. *Manufacturing Service Oper. Management* 10(3):468–483.

Talluri K, van Ryzin G (1998) An analysis of bid-price controls for network revenue management. *Management Sci.* 44(11):1577–1593.

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

Talluri KT, van Ryzin GJ (2006) *The Theory and Practice of Revenue Management*, International Series in Operations Research & Management Science, vol. 68 (Springer Science & Business Media, Berlin).

Vera A, Banerjee S (2019) The Bayesian prophet: A low-regret framework for online decision making. Working paper, Cornell University, Ithaca, NY.

Williamson EL (1992) Airline network seat inventory control: Methodologies and revenue impacts. PhD thesis, Massachusetts Institute of Technology, Cambridge.

Wu H, Srikant R, Liu X, Jiang C (2015) Algorithms with logarithmic or sublinear regret for constrained contextual bandits. Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, eds. *Advances in Neural Information Processing Systems* (Curran Associates, Red Hook, NY), 433–441.