Operations Research

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

Constant Regret Resolving Heuristics for Price-Based Revenue Management

Yining Wang, He Wang

To cite this article:
Yining Wang, He Wang (2022) Constant Regret Resolving Heuristics for Price-Based Revenue Management. Operations Research 70(6):3538-3557. https://doi.org/10.1287/opre.2021.2219

informs
https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 70, No. 6, November–December 2022, pp. 3538–3557
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Methods

Constant Regret Resolving Heuristics for Price-Based Revenue Management

Yining Wang,[a] He Wang[b]

[a] Warrington College of Business, University of Florida, Gainesville, Florida 32611; [b] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Contact: yining.wang@warrington.ufl.edu, https://orcid.org/0000-0001-9410-0392 (YW); he.wang@isye.gatech.edu, https://orcid.org/0000-0001-7444-2053 (HW)

Received: September 9, 2020
Revised: December 28, 2020; May 7, 2021; August 15, 2021
Accepted: October 13, 2021
Published Online in Articles in Advance: January 31, 2022

Area of Review: Stochastic Models

https://doi.org/10.1287/opre.2021.2219

Abstract. Price-based revenue management is an important problem in operations management with many practical applications. The problem considers a seller who sells one or multiple products over T consecutive periods and is subject to constraints on the initial inventory levels of resources. Whereas, in theory, the optimal pricing policy could be obtained via dynamic programming, computing the exact dynamic programming solution is often intractable. Approximate policies, such as the resolving heuristics, are often applied as computationally tractable alternatives. In this paper, we show the following two results for price-based network revenue management under a continuous price set. First, we prove that a natural resolving heuristic attains $O(1)$ regret compared with the value of the optimal policy. This improves the $O(\ln T)$ regret upper bound established in the prior work by Jasin in 2014. Second, we prove that there is an $\Omega(\ln T)$ gap between the value of the optimal policy and that of the fluid model. This complements our upper bound result by showing that the fluid is not an adequate information-relaxed benchmark when analyzing price-based revenue management algorithms.

Funding: This work was supported in part by the National Science Foundation [Grant CMMI-2145661].

Keywords: resolving • self-adjusting controls • price-based revenue management • dynamic pricing

1. Introduction

We study a classic price-based network revenue management problem (Gallego and Van Ryzin 1994, 1997), in which a seller sells one or multiple products over a finite horizon given an initial resource inventory. Suppose there are n products, m resources, and T consecutive selling periods. Throughout the paper, we index the periods backward: $t = T, T-1, \ldots, 1$. We denote the initial inventory level of the resources by $\mathbf{y}_T \in \mathbb{R}_+^m$. Let $\mathbf{A} = [a_{ij}]_{1\le i\le m, 1\le j\le n}$ be a constant matrix indicating that a product in column j requires a_{ij} units of resource in row i.

At the beginning of period t, let $\mathbf{y}_t$ be the current inventory level of resources. The seller posts a price vector $\mathbf{p}_t \in \mathcal{P} \subset \mathbb{R}_+^n$. Let $\mathbf{f}: \mathcal{P} \to \mathbb{R}_+^n$ be the mean demand function. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t=1}^T, \mathbb{P})$ be a filtered probability space. We assume that the realized demand $\mathbf{d}_t$ is given by $\mathbf{d}_t = \mathbf{f}(\mathbf{p}_t) + \boldsymbol{\xi}_t$, where the demand noise $\{\boldsymbol{\xi}_t\}_{t=1}^T$ is a martingale difference sequence adapted to the filtration $\{\mathcal{F}_t\}_{t=1}^T$. Let $\mathbf{s}_t$ be the demand satisfied in period t that maximizes revenue subject to inventory constraints, namely,

$$\mathbf{s}_t = \arg\max_{\mathbf{s}} \{\mathbf{p}_t^{\mathrm{T}}\mathbf{s} \mid \mathbf{0} \le \mathbf{s} \le \mathbf{d}_t, \mathbf{A}\mathbf{s} \le \mathbf{y}_t\}. \tag{1}$$

The realized revenue $\mathbf{r}_t$ in period t and remaining inventory level $\mathbf{y}_{t-1}$ in the next period are given by

$$r_t = \langle \mathbf{p}_t, \mathbf{s}_t \rangle, \quad \mathbf{y}_{t-1} = \mathbf{y}_t - \mathbf{A}\mathbf{s}_t. \tag{2}$$

When a resource is exhausted, the seller may not sell any product using this resource afterward.

The seller's objective is to design an admissible pricing policy π to maximize the expected revenue over the T periods. A pricing policy π can be represented by $\pi = (\pi_T, \ldots, \pi_1)$, where π_t is a mapping from the inventory level vector $\mathbf{y}_t$ to a price vector $\mathbf{p}_t \in \mathcal{P}$. A pricing policy π is *admissible* or *nonanticipating* if the posted price $\mathbf{p}_t$ only depends on the history up to the end of the previous period $t+1$, namely, $\mathbf{p}_t$ is measurable with respect to $\mathcal{F}_{t+1}$. Given the initial (normalized) inventory level $\mathbf{x}_T := \mathbf{y}_T/T$, the expected revenue of an admissible policy π is denoted by

$$R^\pi(T, \mathbf{x}_T) := \mathbb{E}\left[\sum_{t=1}^T r_t \,\middle|\, \mathbf{p}_t = \pi_t(\mathbf{y}_t),\ \forall t = T, \ldots, 1\right]. \tag{3}$$

1.1. Existing Results on the Fluid Model and the Resolving Heuristic

In principle, an optimal policy π^* maximizing $R^\pi(T, \mathbf{x}_T)$ defined in Equation (3) can be obtained via dynamic programming (DP). However, it is well known that the exact DP algorithm is often computationally intractable and suffers from the curse of dimensionality as the state space grows exponentially in size with the number of resources.

The seminal work of Gallego and Van Ryzin (1994) proposes a fluid approximation model of the optimal dynamic pricing problem. To define the fluid model, suppose there exists an inverse function $\mathbf{f}^{-1}$ of the demand rate function of $\mathbf{f}$. Let $r(\boldsymbol{\lambda}) := \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{f}^{-1}(\boldsymbol{\lambda})$ be the mean revenue in one period given the price vector $\mathbf{f}^{-1}(\boldsymbol{\lambda})$. We assume $r(\boldsymbol{\lambda})$ is strictly concave and smooth on its domain $\mathcal{D} \subset \mathbb{R}^n_+$ (see Sections 3 and 6 for the precise statements of these assumptions). The fluid model for the dynamic pricing problem is

$$\begin{aligned} \max_{\boldsymbol{\lambda}\in\mathcal{D}} \ & r(\boldsymbol{\lambda}) \\ \text{s.t.} \ & \mathbf{0} \le \mathbf{A}\boldsymbol{\lambda} \le \mathbf{x}_T. \end{aligned} \tag{4}$$

Let r^{F} and $\boldsymbol{\lambda}^{\mathrm{F}}$ denote the optimal value and the optimal solution of the fluid model, respectively. The optimization problem (4) chooses a demand rate $\boldsymbol{\lambda}^{\mathrm{F}}$ (or equivalently, setting the price to $\mathbf{f}^{-1}(\boldsymbol{\lambda}^{\mathrm{F}})$) such that the revenue function is maximized subject to the initial inventory constraint. The fluid approximation model ignores the randomness caused by demand noises and replaces the stochastic inventory constraint with a deterministic constraint.

The following results are established by Gallego and Van Ryzin (1994, 1997).

Theorem 1 (Gallego and Van Ryzin 1994, 1997). *For any admissible policy π and the initial inventory level $\mathbf{y}_T = \mathbf{x}_T T$, the expected revenue is upper bounded by $R^{\pi}(T, \mathbf{x}_T) \le Tr^{\mathrm{F}}$. Furthermore, for a static pricing policy π^s such that $\mathbf{p}_t = \mathbf{f}^{-1}(\boldsymbol{\lambda}^{\mathrm{F}})\ (\forall t = T, \dots, 1)$, the expected revenue is lower bounded by $R^{\pi^s}(T, \mathbf{x}_T) \ge Tr^{\mathrm{F}} - O(\sqrt{T})$.*

However, a main drawback of the static pricing policy is that it is not adaptive to random demand realization. Researchers propose various approaches to modify the static pricing policy, aiming to improve the $O(\sqrt{T})$ gap in Theorem 1 (see Section 2 for a survey). One intuitive approach is to *resolve* the fluid model in every period using the current inventory level. At period t, we solve the updated problem

$$\begin{aligned} \max_{\boldsymbol{\lambda}\in\mathcal{D}} \ & r(\boldsymbol{\lambda}) \\ \text{s.t.} \ & \mathbf{0} \le \mathbf{A}\boldsymbol{\lambda} \le \mathbf{x}_t, \end{aligned}$$

where $\mathbf{x}_t := \mathbf{y}_t/t$ is the normalized inventory level realized at the beginning of period t, namely, the amount of remaining inventory divided by the remaining periods. Let the solution to the problem be $\boldsymbol{\lambda}_t$. We then set the price in period t as $\mathbf{p}_t = \mathbf{f}^{-1}(\boldsymbol{\lambda}_t)$. We refer to this policy as the *resolving heuristic policy*. Using the resolving heuristic, Jasin (2014) reduces the revenue gap from $O(\sqrt{T})$ to $O(\ln T)$ as shown by the following result.

Theorem 2 (Jasin 2014). *Let $\pi^r = (\pi^r_1, \dots, \pi^r_T)$ be the resolving policy defined as $\mathbf{p}_t = \mathbf{f}^{-1}(\boldsymbol{\lambda}_t)$. Assume that the optimal dual variables of the fluid model (4) are strictly positive. The expected revenue is bounded by $R^{\pi^r}(T, \mathbf{x}_T) \ge Tr^{\mathrm{F}} - O(\ln T)$, where $\mathbf{x}_T = \mathbf{y}_T/T$.*

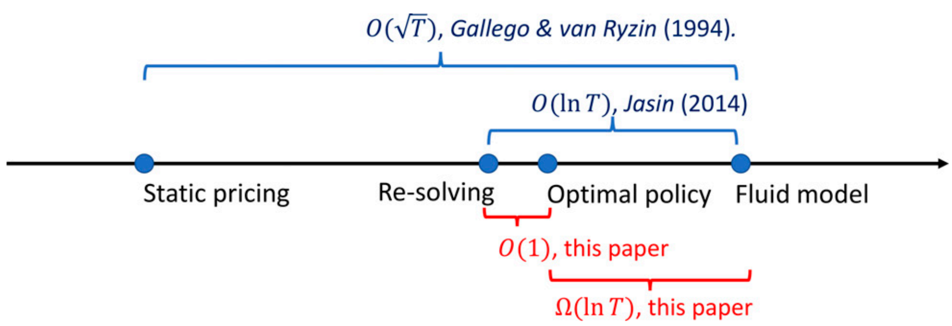
Note that the assumption by Jasin (2014) on dual variables implies that all resources have constrained initial inventory. However, we do not make such an assumption in this paper.

1.2. Our Results: Constant Regret and Logarithmic Gaps

In this paper, we establish two main results: constant regret for the resolving heuristic and a logarithmic regret lower bound on the gap between the value of the optimal pricing policy and the fluid model. Figure 1 summarizes the results established in this paper and compares them with existing results in the prior literature. We establish an upper bound result in Theorem 3, a lower bound result in Theorem 4, and finally an upper bound result in Theorem 4, with details to follow.

Our first main result, as stated in Theorems 3 and 5, asserts that, for any initial inventory level (except for certain boundary cases), the cumulative regret of the resolving heuristic π^r compared against the expected reward of the optimal dynamic pricing policy π^* is

Figure 1. (Color online) An Illustration of Existing Results (above) Compared with Our Results (Below)



upper bounded by a constant that is independent of T. Apart from the obvious improvement from $O(\ln T)$ to $O(1)$ in regret bound, our proof technique differs from the existing work as the latter typically compares the expected reward of the resolving policy to a certain information-relaxed benchmark, such as the fluid model or the hindsight optimum (HO) benchmark (see Section 2). Instead, we compare the value of π^r *directly* with the value of the optimal DP policy π^* by carefully analyzing the stochastic inventory levels under the two policies.

Our second main result, stated in Theorem 4, shows that there is an $\Omega(\ln T)$ *lower bound* on the gap between the expected revenue of the resolving heuristics π^r and the optimal objective of the fluid model Tr^{F}. Coupled with the $O(1)$ regret upper bound established in Theorem 3, this shows that there is an $\Omega(\ln T)$ lower bound on the gap between the value of the optimal policy π^* and the fluid model as well. This demonstrates a fundamental limitation of using the fluid model as a benchmark to analyze the regret for the price-based revenue management problem.

2. Related Work

The idea of using simple, easy-to-compute pricing policies to approximate optimal dynamic pricing policies originates from the seminal work of Gallego and Van Ryzin (1994, 1997), who propose a static price policy using fluid models and establish $O(\sqrt{T})$ regret. Later, Maglaras and Meissner (2006) show that the resolving heuristic has $o(T)$ regret in the single resource case. Atar and Reiman (2012) propose a pricing policy mimicking the behavior of a Brownian bridge (their policy is closely related to the resolving heuristic) and show it has $o(\sqrt{T})$ regret in the network revenue management setting. The most relevant prior research to our paper is the work by Jasin (2014), who studies a price-based network revenue management problem and shows that resolving heuristics attain $O(\ln T)$ asymptotic regret upper bound under mild conditions. Jasin (2014) also shows that infrequent resolving has similar theoretical performance guarantees and is much more computationally efficient. In this paper, we improve the regret of the resolving heuristic to an $O(1)$ constant bound independent of T. Our analysis is different from the one in Jasin (2014) in the sense that we directly compare the expected revenue of resolving with the value of the optimal DP policy instead of the fluid model. Additionally, we complement the result in Jasin (2014) by establishing an $\Omega(\ln T)$ lower bound between the expected revenue of the optimal DP policy and the fluid model.

Although we focus on price-based revenue management in this paper, we note that resolving heuristics have also been studied in *quantity-based* revenue management (Williamson 1992, Cooper 2002, Reiman and Wang 2008, Secomandi 2008, Jasin and Kumar 2013, Bumpensanti and Wang 2020). The decisions involved in a quantity-based revenue management problem are the opening and closing of available products, so the feasible decisions form a discrete set. Vera et al. (2021) propose a constant regret algorithm that works for both quantity- and price-based revenue management problems. However, the dynamic pricing model considered in Vera et al. (2021) assumes a single product and a discrete price set, whereas our paper considers multiple products and *continuous* price sets.

There are two major differences in the analysis of resolving heuristics as one moves from a discrete decision set to a continuous decision set. First of all, when the decision set is discrete (which is the case for quantity-based revenue management or price-based revenue management with a discrete price set), resolving the fluid model leads to fractional solutions that require rounding; it is shown that the design of the rounding procedures (e.g., by randomization, thresholding, etc.) plays a critical role in the performance of resolving heuristics (Arlotto and Gurvich 2019, Bumpensanti and Wang 2020, Vera and Banerjee 2021). In contrast, when the decision set is continuous, resolving algorithms must precisely track the updated inventory level in the fluid model without rounding the fluid model solution. As such, analysis for the continuous price-based revenue management problem is much different from the previous results for either quantity- or discrete price–based revenue management. Second, existing proofs of constant regret in quantity-based revenue management often use an information-relaxation bound called the *hindsight-optimum* benchmark. However, the hindsight-optimum benchmark does not admit a constant regret in the continuous price-based revenue management problem (see Appendix B), and our analysis does not rely on such a benchmark.

In this paper, we restrict our attention to a basic setting with stationary demand. Several other papers study revenue management problems with nonstationary, time-correlated, or arbitrary demand sequences (e.g., Chen and Farias 2013, Ma and Simchi-Levi 2020, Ma et al. 2020). Revenue management problems with nonstationary demand are harder, and therefore, these papers consider different performance metrics, such as approximation or competitive ratios rather than regret. Chen and Farias (2013) study a dynamic pricing problem under a general nonstationary demand setting. They show that resolving heuristics can achieve constant competitive ratios, whereas static pricing policies have asymptotically diminishing competitive ratios. Because of different model assumptions, the results in Chen and Farias (2013) are not directly comparable to ours.

Another stream of related literature studies dynamic pricing with *demand learning*, in which the underlying demand function is unknown and needs to be learned on the fly from sales data. Many papers consider demand learning settings using the price-based finite-inventory revenue management model from Gallego and Van Ryzin (1994) as the ground truth model (e.g., Aviv and Pazgal 2005; Besbes and Zeevi 2009, 2012; Lei et al. 2014; Wang et al. 2014; den Boer and Zwart 2015; Ferreira et al. 2018). Resolving heuristics are also applied to demand learning algorithms (Jasin 2015, Ferreira et al. 2018). In contrast, our paper assumes that the seller has full information about the demand curve and demand distributions, so we are able to show $O(1)$ regret, which is tighter than typical regret bounds in the learning literature. Also, the lower bound result in this paper (Theorem 4) is proved using different techniques from lower bound proofs in the learning setting (e.g., Broder and Rusmevichientong 2012, Wang et al. 2021) as the latter relies on information-theoretical lower bounds.

3. Main Results

In this section, we focus on the single product setting. The extension to a network revenue management model with multiple products and multiple resources is deferred to Section 6. Although some of the results in this section are special cases of those in the network revenue management setting, presenting the single product setting helps illustrate the key insight of our analysis without the technical complication of multiple resources.

In the single product setting, given the initial normalized inventory level x_T, the fluid model (4) becomes $\max_{\lambda\in\mathcal{D}}\{r(\lambda)\,|\,0\le\lambda\le x_T\}$. The resolving heuristic sets the price as $p_t=f^{-1}(\lambda_t)$ in each period ($\forall t=T,\dots,1$), where λ_t is the solution to $\max_{\lambda\in\mathcal{D}}\{r(\lambda)\,|\,0\le\lambda\le x_t\}$. Let the realized demand be $d_t=\lambda_t+\xi_t$, where ξ_t is the demand noise. We make the following assumptions on the single product demand model:

A1. (Monotonicity). The demand rate function $f:[\underline{p},\overline{p}]\to\mathbb{R}_+$ is strictly decreasing with $f(\underline{p})=\overline{d}$, $f(\overline{p})=\underline{d}$. (We allow either $\overline{p}<+\infty$ or $\overline{p}=+\infty$.) This implies the existence of f's inverse function f^{-1} on $[\underline{d},\overline{d}]$.

A2. (Strict Concavity). The expected revenue $r(\lambda)=\lambda f^{-1}(\lambda)$ as a function of the demand rate λ is strictly concave. There exists a positive constant $m>0$ such that $r''(\lambda)\le -m$ for all $\lambda\in[\underline{d},\overline{d}]$. The maximizer of $r(\lambda)$ is in the interior of the domain, that is, $\arg\max_\lambda r(\lambda)\in(\underline{d},\overline{d})$.

A3. (Smoothness). The third derivative of $r(\lambda)$ exists and satisfies $|r'''(\lambda)|\le M$ for all $\lambda\in[\underline{d},\overline{d}]$. Furthermore, there exists a constant $C>0$ such that $|r(\lambda)-r(\lambda')|\le C\,|\lambda-\lambda'|$ for all $\lambda,\lambda'\in[\underline{d},\overline{d}]$.

A4. (Martingale Difference Sequence). Conditional on the price p_t ($\forall t=T,\dots,1$), the demand noise ξ_t satisfies $\mathbb{E}[\xi_t\,|\,\mathcal{F}_{t+1}]=\mathbb{E}[\xi_t\,|\,p_t]=0$ *a.s.*, where $\mathcal{F}_{t+1}$ contains the history up to the previous period $t+1$. The conditional distribution of ξ_t given p_t is denoted by $\xi_t\sim Q(p_t)$. In addition, $|\xi_t|\le B_\xi$ *a.s.* for some constants $B_\xi<\infty$.

A5. (Bounded Wasserstein Distance). There exists a constant $L>0$ such that, for any $p,p'\in[\underline{p},\overline{p}]$, it holds that $\mathcal{W}_2(Q(p),Q(p'))\le L|f(p)-f(p')|$, where $\mathcal{W}_2(Q,Q'):=\inf_\Xi\sqrt{\mathbb{E}_{\xi,\xi'\sim\Xi}[|\xi-\xi'|^2]}$ is the L_2-Wasserstein distance between Q,Q', with Ξ being an arbitrary joint distribution with marginal distributions being Q and Q', respectively.

Assumptions A1–A3 are standard assumptions in price-based revenue management (Gallego and Van Ryzin 1994). In particular, the strict concavity of $r(\lambda)$ stems from the economic principle of diminishing marginal returns. Assumption A4 allows demand noise to have a rather general dependence on price. Assumption A5 states that, if two prices are close to each other, then the demand noises given these prices should also have similar distributions. This assumption is naturally satisfied when demand is modeled by a parametric family of distributions (e.g., Bernoulli, binomial, truncated normal) whose parameters depend continuously on price.

3.1. Constant Regret of the Resolving Heuristic

Let $r^u:=\max_{\lambda\ge0}r(\lambda)$ and $\lambda^u:=\arg\max_{\lambda\ge0}r(\lambda)$ be the unconstrained optimal revenue rate and its maximizer, respectively. Because $r(\lambda)$ is strictly concave, it is easily verified that the unique optimal solution to the fluid model $\max_{\lambda\in\mathcal{D}}\{r(\lambda)\,|\,0\le\lambda\le x_T\}$ is equal to $\lambda^{\mathrm{F}}=\min\{x_T,\lambda^u\}$, and hence, the fluid optimal price is $f^{-1}(\min\{x_T,\lambda^u\})$.

When the normalized initial inventory level x_T exceeds the (unconstrained) optimal demand rate λ^{pu}, it is easy to prove that both the static policy $\pi^s:p_t\equiv f^{-1}(\lambda^u)$ and the resolving heuristic π^r have constant regret (see proof in Appendix A).

Proposition 1. *Suppose $x_T>\lambda^u$. Let $\pi^s:p_t\equiv f^{-1}(\lambda^u)$ be the static pricing policy. Then, $R^{\pi^s}(T,x_T)\ge Tr^u-O(1)\ge R^{\pi^*}(T,x_T)-O(1)$. In addition, the resolving heuristic satisfies $R^{\pi^r}(T,x_T)\ge Tr^u-O(1)\ge R^{\pi^*}(T,x_T)-O(1)$.*

However, analysis for the limited inventory case ($x_T<\lambda^u$) is much more complicated. The static price policy $\pi^s\equiv f^{-1}(x_T)$ typically suffers $\Omega(\sqrt{T})$ regret in this case. Jasin (2014) shows that the regret of the resolving heuristic π^r when measured against the fluid benchmark $Tr(x_T)$ is at most $O(\ln T)$. Our next theorem improves the regret of π^r to a constant.

Theorem 3. *Suppose $x_T\in(\underline{d},\lambda^u)$. Let π^r be the resolving heuristic and π^* be the optimal policy. For $T\ge1$, it holds that $R^{\pi^r}(T,x_T)\ge R^{\pi^*}(T,x_T)-O(1)$.*

Theorem 3 is the main result of this section, and its proof is given in Section 4. Unlike the previous results by Gallego and Van Ryzin (1994) and Jasin (2014), Theorem 3 compares the expected revenue of π^r directly with the optimal DP pricing policy π^* rather than comparing it with the fluid approximation value $Tr(x_T)$. This change of benchmark allows us to derive a tighter regret bound. As we establish in the next section, it is impossible to obtain $O(1)$ regret using the fluid model as the benchmark.

In light of Proposition 1 and Theorem 3, we know that the resolving heuristic π^r has constant regret in either the sufficient inventory case ($x_T > \lambda^u$) or the limited inventory case ($x_T < \lambda^u$). So the only remaining scenario is the boundary case ($x_T = \lambda^u$). We investigate this scenario using numerical experiments in Section 5. Surprisingly, the numerical experiment indicates that π^r does *not* have constant regret in the boundary case. This observation is analogous to the situation for the quantity-based revenue management problem, in which the resolving heuristic has constant regret when the fluid model (a linear program in that setting) has nondegenerate solutions but does not have constant regret in certain boundary cases when the fluid model has degenerate solutions (Jasin and Kumar 2012, Bumpensanti and Wang 2020).

3.2. Logarithmic Gap of the Fluid Model Benchmark

In this section, we show that, in the limited inventory case ($x_T < \lambda^u$), the regret of the resolving policy π^r measured against the fluid approximation value $Tr(x_T)$ must be at least $\Omega(\ln T)$.

Theorem 4. *Suppose $x_T \in (\underline{d}, \lambda^u)$ and let π^r be the resolving policy defined in Theorem 2. Suppose there exists $\sigma > 0$ such that $\mathbb{E}_{\xi\sim Q(p)}[\xi^2] \geq \sigma^2$ for any $p \in [\underline{p}, \overline{p}]$. For $T \geq 2$, it holds that $R^{\pi^r}(T, x_T) \leq Tr(x_T) - \Omega(\ln T)$.*

Theorem 4 is proved in Section 4. It implies that the logarithmic regret by Jasin (2014) (see the statement of Theorem 2), which uses the fluid model benchmark, is tight and cannot be improved. Because $R^{\pi^*}(T, x_T) - R^{\pi^r}(T, x_T) = O(1)$ by Theorem 3, Theorem 4 also shows that there is a logarithmic gap $\Omega(\ln T)$ between the value of the *optimal DP policy* and the fluid model. This is why our constant regret analysis does not use the fluid model as the benchmark.

Prior work on the quantity-based revenue management problem (Reiman and Wang 2008, Bumpensanti and Wang 2020, Vera et al. 2021) also considers different regret benchmarks that are tighter than the fluid model, including various versions of the hindsight optimum benchmark. The hindsight optimum model assumes a clairvoyant who knows the *aggregate* realized demands for the entire horizon at the start. In Appendix B of this paper, we show that one version of the hindsight optimum benchmark proposed by Vera et al. (2021) has $O(1)$ regret when measured against the fluid approximation benchmark. By Theorem 4, this hindsight optimum benchmark is also $\Omega(\ln T)$ away from the value of the optimal DP policy.

4. Proofs

Before presenting our proof, we first define some notations. Let $\phi_t^*(x) = R^{\pi^*}(t, x)$ and $\phi_t^r(x) = R^{\pi^r}(t, x)$ be the expected cumulative revenue of the optimal DP pricing policy π^* and the resolving policy π^r, respectively, starting from period t given xt units of remaining inventory. Let x_t^* and x_t^r be the normalized inventory levels under policy π^* and π^r at the beginning of period t. These notations are summarized in Table 1 with some additional notations defined later in the proof.

The rest of this section is organized as follows. In the first section, we establish some properties of the optimal policy π^* and the resolving policy π^r. More specifically, we establish upper and lower bounds of the expected rewards $\phi_t^*(\cdot)$, $\phi_t^r(\cdot)$ using the key quantities of $\{\overline{\Delta}_{\to t}\}$ (harmonic series of optimal demand corrections), $\{\overline{\xi}^*_{\to t}, \overline{\xi}^r_{\to t}\}$ (harmonic series of stochastic noise variables), and $T^\sharp$ (a stopping time until which the demand noise process is well behaved). We then proceed with the proofs of Theorems 3 and 4 by carefully analyzing the differences in the Taylor series expansions of $\phi_t^*(\cdot)$, $\phi_t^r(\cdot)$.

Table 1. Notations Used in the Proof

Notation	Definition	Meaning
λ^u	$\lambda^u = \arg\max_{\lambda\in[\underline{d},\overline{d}]} r(\lambda)$	The optimal demand rate without inventory constraints
$\phi_t^*(x)$	$\phi_t^*(x) = R^{\pi^*}(t, x)$	Reward of DP π^* given t periods and xt inventory
$\phi_t^r(x)$	$\phi_t^r(x) = R^{\pi^r}(t, x)$	Reward of resolving π^r given t periods and xt inventory
$\mathcal{F}_t$	σ-algebra of $\{\xi_\tau\}_{\tau\geq t}$	All the events known up to the end of period t
x_t^*, x_t^r	Remaining inventory divide by t	Normalized inventory levels under policy π^* and π^r
Δ_t	See Equation (5)	The optimal DP demand correction in period t
$\overline{\Delta}_{\to t}$	$\frac{\Delta_T}{T-1} + \frac{\Delta_{T-1}}{T-2} + \cdots + \frac{\Delta_{t+1}}{t}$	Harmonic series of demand corrections up to t
ξ_t^*, ξ_t^r	$\xi_t^* \sim Q(f^{-1}(\lambda_t^*))$, $\xi_t^r \sim Q(f^{-1}(\lambda_t))$	The stochastic demand noises at time t under π^*, π^r
$\overline{\xi}^*_{\to t}, \overline{\xi}^r_{\to t}$	$\frac{\xi_T}{T-1} + \frac{\xi_{T-1}}{T-2} + \cdots + \frac{\xi_{t+1}}{t}$	Weighted demand noises up to t under π^* and π^{pr}
$\overline{\xi}^\delta_{\to t}$	$\overline{\xi}^r_{\to t} - \overline{\xi}^*_{\to t}$	Difference in harmonic demand noise series
Ξ_t	A joint distribution over (ξ_t^*, ξ_t^r)	The joint distribution that minimizes $\sqrt{\mathbb{E}_{\Xi_t}[\lvert\xi_t^* - \xi_t^r\rvert^2]}$
ℓ_t	$(\lambda_t + \xi_t - y_t)^+$	Lost sales at time t
$T^\sharp$	See Equation (8)	A stopping time such that $\{x_\tau^r\}_{\tau > T^\sharp}$ is well-behaved

4.1. Properties of the Optimal Policy and the Resolving Heuristic

For any $\tau \geq 1$, let the random variable x^*_τ be the normalized inventory level (i.e., remaining inventory divided by remaining periods) at period τ under policy π^*. Let $y_\tau = x^*_\tau \tau$ and $y_{\tau-1} = x^*_{\tau-1}(\tau-1)$ be the actual inventory levels at periods τ and $\tau-1$, respectively. If the policy selects the demand rate $\lambda \in [\underline{d}, \overline{d}]$, let the realized noise be ξ_τ and lost sales be $\ell_\tau = (\lambda + \xi_\tau - y_\tau)^+$; then, we have $y_{\tau-1} = y_\tau - (\lambda + \xi_\tau) + \ell_\tau$, which implies $x^*_{\tau-1} = y_{\tau-1}/(\tau-1) = x^*_\tau - (\lambda + \xi_\tau - \ell_\tau - x^*_\tau)/(\tau-1)$. Therefore, the value of the optimal policy π^* is given by the following Bellman equation:

$$\phi^*_\tau(x^*_\tau) = \max_{\lambda \in [\underline{d}, \overline{d}]} \left\{ r(\lambda) + \mathbb{E}_{\xi_\tau \sim Q(f^{-1}(\lambda))} \times \left[\phi^*_{\tau-1}\left(x^*_\tau - \frac{\lambda + \xi_\tau - \ell_\tau - x^*_\tau}{\tau - 1} \right) - f^{-1}(\lambda)\ell_\tau \right] \right\},$$

where the first term is the expected revenue, the second term represents the revenue-to-go function from period $\tau-1$ onward, and the last term subtracts revenue resulting from lost sales. Let λ^*_τ denote the maximizer of the right-hand side and define $\Delta_\tau := \lambda^*_\tau - x^*_\tau$. Let ξ^*_τ be the realized demand noise, which by Assumption A4 is drawn from the distribution $Q(f^{-1}(\lambda^*_\tau))$, and let ℓ^*_τ be realized lost sales. The normalized inventory level at the start of the next period is $x^*_{\tau-1} = x^*_\tau - (\Delta_\tau + \xi^*_\tau - \ell^*_\tau)/(\tau-1)$. The Bellman equation can be rewritten as

$$\phi^*_\tau(x^*_\tau) = r(x^*_\tau + \Delta_\tau) + \mathbb{E}\left[\phi^*_{\tau-1}\left(x^*_\tau - \frac{\Delta_\tau + \xi^*_\tau - \ell^*_\tau}{\tau-1} \right) - f^{-1}(\lambda^*_\tau)\ell^*_\tau \right] \quad \forall\, 2 \leq \tau \leq T,\, x^*_\tau \geq 0. \tag{5}$$

This equation implies $\underline{d} \leq x^*_\tau + \Delta_\tau \leq \overline{d}$. However, x^*_τ can be outside the domain $[\underline{d}, \overline{d}]$.

For any $t < T$, let

$$\overline{\Delta}_{\to t} := \frac{\Delta_T}{T-1} + \frac{\Delta_{T-1}}{T-2} + \ldots + \frac{\Delta_{t+1}}{t},$$
$$\overline{\xi}^*_{\to t} := \frac{\xi^*_T}{T-1} + \frac{\xi^*_{T-1}}{T-2} + \ldots + \frac{\xi^*_{t+1}}{t}, \text{ and}$$
$$\overline{\ell}^*_{\to t} := \frac{\ell^*_T}{T-1} + \frac{\ell^*_{T-1}}{T-2} + \ldots + \frac{\ell^*_{t+1}}{t}$$

be the harmonic series of corresponding variables up to time t under the optimal policy. We also define $\overline{\Delta}_{\to T} = \overline{\xi}^*_{\to T} = \overline{\ell}^*_{\to T} = 0$. It then holds that

$$x^*_\tau = x_T - \overline{\Delta}_{\to\tau} - \overline{\xi}^*_{\to\tau} + \overline{\ell}^*_{\to\tau}, \qquad \forall\, 1 \leq \tau \leq T. \tag{6}$$

Next, we consider the resolving heuristic π^r. For any $\tau \geq 1$, let the random variable x^r_τ be the normalized inventory level under the resolving heuristic. If $x^r_\tau \in [\underline{d}, \lambda^u]$, the resolving heuristic selects the price $f^{-1}(x^r_\tau)$. Let ξ^r_τ be the realized demand noise under this price, which is drawn from the distribution $Q(f^{-1}(x^r_\tau))$, and the realized lost sales be $\ell^r_\tau = (x^r_\tau + \xi^r_\tau - y_\tau)^+$. The normalized inventory level in the next period is equal to $x^r_{\tau-1} = (x_\tau \tau - x_\tau - \xi^r_\tau + \ell^r_\tau)/(\tau-1) = x^r_\tau - (\xi^r_\tau - \ell^r_\tau)/(\tau-1)$. Thus, the value of the resolving policy π^r can be written as

$$\phi^r_t(x^r_\tau) = r(x^r_\tau) + \mathbb{E}\left[\phi^r_{t-1}\left(x^r_\tau - \frac{\xi^r_\tau - \ell^r_\tau}{\tau-1} \right) - f^{-1}(x^r_\tau)\ell^r_\tau \right] \quad \forall\, 2 \leq \tau \leq T,\, \underline{d} \leq x^r_\tau \leq \lambda^u. \tag{7}$$

Note that Equation (7) does *not* hold for $x^r_\tau > \lambda^u$, in which case the resolving policy π^r chooses the unconstrained optimal demand rate λ^{pu} instead of x^r_τ. Comparing Equation (7) with Equation (5), we remark that the resolving heuristics π^r can be viewed as a special case of the dynamic programming policy with the decision rule restricted to $\Delta_\tau \equiv 0$ when the condition $x^r_\tau \leq \lambda^u$ is satisfied.

This observation motivates the definition of a stopping time $T^\sharp$, which ensures that Equation (7) holds for all $\tau \geq T^\sharp$. Specifically, we define a time index $T^\sharp$ as

$$T^\sharp := \max\left\{ \tau \geq 1 : |\overline{\xi}^r_{\to\tau-1}| > \min\left(x_T - \underline{d}, \lambda^u - x_T, \frac{r'(x_T)}{2|r''(x_T)|} \right) \right\} \vee \left\lceil \frac{2\,|r''(x_T)|\,B_\xi}{r'(x_T)} + 1 \right\rceil, \tag{8}$$

where $\overline{\xi}^r_{\to\tau} = \frac{\xi^r_T}{T-1} + \ldots + \frac{\xi^r_{\tau+1}}{\tau}$ is the harmonic series of demand noises under the resolving policy. (If $T < \lceil 2\,|r''(x_T)|\,B_\xi / r'(x_T) \rceil$, let $T^\sharp := T$.) Note that $\overline{\xi}^r_{\to\tau-1}$ is a function of $\{\xi^r_t\}_{t=\tau}^T$ and is measurable with respect to $\mathcal{F}_\tau$, so $T^\sharp$ is indeed a stopping time. Intuitively, $T^\sharp$ is the first time that the inventory level in the next period $x^r_{\tau-1}$ falls outside of the interval $[\underline{d}, \lambda^u]$. In other words, the inventory level of the resolving heuristic x^r_τ satisfies Equation (7) up to time $T^\sharp$. The additional term involving $r'(x_T)/\,|2r''(x_T)|$ in Equation (8) is needed for a technical reason in the proof. Because $|\overline{\xi}_{\to\tau}| \leq \min\{x_T - \underline{d}, \lambda^u - x_T\}$ holds for all $\tau \geq T^\sharp$, we have

$$x^r_\tau = x_T - \overline{\xi}^r_{\to\tau}, \;\; \forall \tau \geq T^\sharp, \text{ and } x^r_{T^\sharp-1} = (x_T - \overline{\xi}^r_{\to T^\sharp-1})^+. \tag{9}$$

Note that the definition of $T^\sharp$ in Equation (8) implies that $x^r_\tau = x_T - \overline{\xi}^r_{\to\tau} \geq 0$ for all $\tau \geq T^\sharp$, so the lost sales $\ell^r_\tau = 0$ for all $\tau \geq T^\sharp + 1$; however, lost sales may occur at period $T^\sharp$.

The following lemma shows that $\mathbb{E}[T^\sharp]$ is bounded by a constant that is independent of the horizon length T. This critical fact is used later to prove constant regret as we reduce the expected regret to a function of $\mathbb{E}[T^\sharp]$.

Lemma 1. *For $x_T \in (\underline{d}, \lambda^u)$ and any $T \geq 1$, it holds that*

$$\mathbb{E}[T^\sharp] \leq 2 + \frac{2\,|r''(x_T)|\,B_\xi}{r'(x_T)} + \frac{4B_\xi^4}{\min\{x_T - \underline{d}, \lambda^u - x_T, -r'(x_T)/2r''(x_T)\}^4}. \quad (10)$$

We also use the following lemma in the proof to bound the expected total revenues of the optimal policy and resolving heuristic as the sum of single period revenues. The proofs of Lemmas 1 and 2 can be found in Appendix A.

Lemma 2. *Let $x_T \in (\underline{d}, \lambda^u)$. Let $T^\sharp$ be the stopping time defined in Equation (8). Then, it holds that*

$$\phi_T^*(x_T) \leq \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T} r(x_\tau^* + \Delta_\tau) + (T^\sharp - 1) r(\min\{x_{T^\sharp-1}^*, \lambda^u\})\right],$$

$$\mathbb{E}\left[\sum_{\tau=T^\sharp}^{T} r(x_\tau^r)\right] - r^u \leq \phi_T^r(x_T) \leq \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T} r(x_\tau^r) + (T^\sharp - 1) r(\min\{x_{T^\sharp-1}^r, \lambda^u\})\right].$$

4.2. Proof of Theorem 3

In this section, we prove Theorem 3. The proof is divided into three main steps. In the first step (from Equations (11)–(15)), we use Taylor expansion and the strict concavity and smoothness properties of the revenue function r to bound the expected revenue difference between the optimal DP policy and the resolving policy. In the second step (from Equations (16)–(23)), the regret expression is split into five terms denoted by $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}$, and each term is simplified separately. In the third step (from Equations (24)–(26)), the five terms are combined again, and the regret is eventually reduced to a function of $\mathbb{E}[T^\sharp]$, which is $O(1)$ by Lemma 1.

In the first step of our proof to Theorem 3, we use Taylor expansion and a "completing-the-square" trick to upper bound the difference between $\phi_T^*(x_T)$ and $\phi_T^r(x_T)$, namely, the difference between the expected revenues of the optimal DP solution and the resolving heuristic. By Lemma 2, for any $x_T \in (\underline{d}, \lambda^u)$, it holds that

$$\phi_T^*(x_T) \leq \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T} r(x_\tau^* + \Delta_\tau) + (T^\sharp - 1) r(\min\{x_{T^\sharp-1}^*, \lambda^u\})\right], \quad (11)$$

$$\phi_T^r(x_T) \geq \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T} r(x_\tau^r)\right] - r^u. \quad (12)$$

Recall that $T^\sharp$ is the stopping time defined in Equation (8). For any $\tau \geq T^\sharp$, we have $x_\tau^* = x_T - \overline{\Delta}_{\to\tau} - \overline{\xi}_{\to\tau}^* + \overline{\ell}_{\to\tau}^*$ by Equation (6), and $x_\tau^r = x_T - \overline{\xi}_{\to\tau}^r$ by Equation (9). For notational simplicity, define $\overline{\xi}_{\to\tau}^\delta := \overline{\xi}_{\to\tau}^r - \overline{\xi}_{\to\tau}^*$. To bound the difference between $r(x_\tau^* + \Delta_\tau) - r(x_\tau^r)$, we have

$$\begin{aligned}
&r(x_\tau^* + \Delta_\tau) - r(x_\tau^r) \\
&\quad = r(x_T - \overline{\Delta}_{\to\tau} - \overline{\xi}_{\to\tau}^* + \overline{\ell}_{\to\tau}^* + \Delta_\tau) - r(x_T - \overline{\xi}_{\to\tau}^r) \\
&\quad \leq r'(\underline{d})\overline{\ell}_{\to\tau}^* + r(x_T - \overline{\Delta}_{\to\tau} - \overline{\xi}_{\to\tau}^* + \Delta_\tau) - r(x_T - \overline{\xi}_{\to\tau}^r) \\
&\quad \leq r'(\underline{d})\overline{\ell}_{\to\tau}^* + r'(x_T - \overline{\xi}_{\to\tau}^r)[\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}_{\to\tau}^\delta] \\
&\qquad - \frac{m}{2}|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}_{\to\tau}^\delta|^2 \\
&\quad \leq r'(\underline{d})\overline{\ell}_{\to\tau}^* + r'(x_T)[\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}_{\to\tau}^\delta] \\
&\qquad - r''(x_T)\overline{\xi}_{\to\tau}^r[\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}_{\to\tau}^\delta] \\
&\qquad + \frac{M}{2}|\overline{\xi}_{\to\tau}^r|^2|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}_{\to\tau}^\delta| - \frac{m}{2}|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}_{\to\tau}^\delta|^2,
\end{aligned} \quad (13)$$

where the first inequality holds because $\overline{\ell}_{\to\tau}^* \geq 0$ and $r'(\underline{d}) \geq r'(x)$ for any $x \in [\underline{d}, \overline{d}]$ by concavity, the second inequality uses the strict concavity Assumption A2, and the third inequality uses the smoothness Assumption A3.

Define $\xi_\tau^\delta := \xi_\tau^r - \xi_\tau^*$. Taking expectations on both sides of Equation (13) and summing over $\tau = T, T-1, \ldots, T^\sharp$, we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{\tau=T^\sharp}^{T} (r(x_\tau^* + \Delta_\tau) - r(x_\tau^r))\right] \\
&\leq \mathbb{E}\Bigg[\sum_{\tau=T^\sharp}^{T} \Big(r'(x_T)[\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}_{\to\tau}^\delta - \xi_\tau^\delta] \\
&\quad - r''(x_T)\overline{\xi}_{\to\tau}^r[\Delta_\tau - \overline{\Delta}_{\to\tau}] + r''(x_T)\overline{\xi}_{\to\tau}^r[\xi_\tau^\delta - \overline{\xi}_{\to\tau}^\delta] \\
&\quad + \frac{M}{2}\left|\overline{\xi}_{\to\tau}^r\right|^2\left|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}_{\to\tau}^\delta\right| - \frac{m}{2}\left|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}_{\to\tau}^\delta\right|^2 \\
&\quad + r'(\underline{d})\overline{\ell}_{\to\tau}^*\Big)\Bigg],
\end{aligned} \quad (14)$$

where we add a few terms using $\mathbb{E}[\sum_{\tau=T^\sharp}^{T} \xi_\tau^\delta] = 0$ and $\mathbb{E}[\sum_{\tau=T^\sharp}^{T} \overline{\xi}_{\to\tau}^r \xi_\tau^\delta] = 0$ by Doob's optional stopping theorem (recall that $T^\sharp$ is a stopping time, and $\overline{\xi}_{\to\tau}^r$ is measurable with respect to $\mathcal{F}_{\tau+1}$).

Define $\eta = -(x_{T^\sharp-1}^* - \lambda^u)^+$. By applying Equation (13) to period $\tau = T^\sharp - 1$, we have

$$\begin{aligned}
&r(\min\{x_{T^\sharp-1}^*, \lambda^u\}) - r(x_{T^\sharp-1}^r) = r(x_{T^\sharp-1}^* + \eta) - r(x_{T^\sharp-1}^r) \\
&\quad \leq r'(x_T)[\eta - \overline{\Delta}_{\to T^\sharp-1} + \overline{\xi}_{\to T^\sharp-1}^\delta] \\
&\qquad - r''(x_T)\overline{\xi}_{\to T^\sharp-1}^r[\eta - \overline{\Delta}_{\to T^\sharp-1} + \overline{\xi}_{\to T^\sharp-1}^\delta] \\
&\qquad + \frac{M}{2}|\overline{\xi}_{\to T^\sharp-1}^r|^2\,|\eta - \overline{\Delta}_{\to T^\sharp-1} + \overline{\xi}_{\to T^\sharp-1}^\delta| \\
&\qquad - \frac{m}{2}|\eta - \overline{\Delta}_{\to T^\sharp-1} + \overline{\xi}_{\to T^\sharp-1}^\delta|^2 + r'(\underline{d})\overline{\ell}_{\to T^\sharp-1}^* \\
&\quad = r'(x_T)[\eta - \overline{\Delta}_{\to T^\sharp-1} + \overline{\xi}_{\to T^\sharp-1}^\delta] \\
&\qquad - r''(x_T)\overline{\xi}_{\to T^\sharp-1}^r[\eta - \overline{\Delta}_{\to T^\sharp-1} + \overline{\xi}_{\to T^\sharp-1}^\delta] \\
&\qquad + \frac{M^2}{8m}|\overline{\xi}_{\to T^\sharp-1}^r|^4 - \frac{M^2}{8m}|\overline{\xi}_{\to T^\sharp-1}^r|^4 \\
&\qquad + \frac{M}{2}|\overline{\xi}_{\to T^\sharp-1}^r|^2\,|\eta - \overline{\Delta}_{\to T^\sharp-1} + \overline{\xi}_{\to T^\sharp-1}^\delta| \\
&\qquad - \frac{m}{2}|\eta - \overline{\Delta}_{\to T^\sharp-1} + \overline{\xi}_{\to T^\sharp-1}^\delta|^2 + r'(\underline{d})\overline{\ell}_{\to T^\sharp-1}^*
\end{aligned}$$

$$\begin{aligned}&\le r'(x_T)[\eta-\overline{\Delta}_{\to T^\sharp-1}+\overline{\xi}^{\delta}_{\to T^\sharp-1}]\\&\quad -r''(x_T)\overline{\xi}^{r}_{\to T^\sharp-1}[\eta-\overline{\Delta}_{\to T^\sharp-1}+\overline{\xi}^{\delta}_{\to T^\sharp-1}]\\&\quad +\frac{M^2}{8m}|\overline{\xi}^{r}_{\to T^\sharp-1}|^4+r'(\underline{d})\overline{\ell}^{*}_{\to T^\sharp-1},\end{aligned}\tag{15}$$

where the last inequality follows by completing the square as $-\frac{M^2}{8m}|\overline{\xi}^{r}_{\to T^\sharp-1}|^4+\frac{M}{2}|\overline{\xi}^{r}_{\to T^\sharp-1}|^2\,|\eta-\overline{\Delta}_{\to T^\sharp-1}+\overline{\xi}^{\delta}_{\to T^\sharp-1}|-\frac{m}{2}|\eta-\overline{\Delta}_{\to T^\sharp-1}+\overline{\xi}^{\delta}_{\to T^\sharp-1}|^2=-\frac{m}{2}(|\eta-\overline{\Delta}_{\to T^\sharp-1}+\overline{\xi}^{\delta}_{\to T^\sharp-1}|-\frac{M}{2m}|\overline{\xi}^{r}_{\to T^\sharp-1}|^2)^2\le 0$ almost surely.

Subtracting (12) from (11), we obtain

$$\begin{aligned}&\phi^*_T(x_T)-\phi^r_T(x_T)\\&\le\mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}(r(x^*_\tau+\Delta_\tau)-r(x^r_\tau))+(T^\sharp-1)r(\min\{x^*_{T^\sharp-1},\lambda^u\})\right]+r^u\\&\le\mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}(r(x^*_\tau+\Delta_\tau)-r(x^r_\tau))+(T^\sharp-1)(r(x^*_{T^\sharp-1}+\eta)-r(x^r_{T^\sharp-1}))\right]\\&\quad+\mathbb{E}\big[(T^\sharp-1)r^u\big]+r^u\\&\le\mathbb{E}[r'(x_T)\mathcal{A}-r''(x_T)\mathcal{B}+r''(x_T)\mathcal{C}+\mathcal{D}+r'(\underline{d})\mathcal{E}]+r^u\mathbb{E}[T^\sharp],\end{aligned}\tag{16}$$

where the second inequality holds because $x_{T^\sharp-1}+\eta=\min\{x^*_{T^\sharp-1},\lambda^u\}$ and $r^u\ge r(x^r_{T^\sharp-1})$, the last inequality uses Equations (14) and (15), and the terms $\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D},\mathcal{E}$ are defined as

$$\begin{aligned}\mathcal{A}&=\sum_{\tau=T^\sharp}^{T}[\Delta_\tau-\overline{\Delta}_{\to\tau}+\overline{\xi}^{\delta}_{\to\tau}-\xi^\delta_\tau]\\&\quad-(T^\sharp-1)(\overline{\Delta}_{\to T^\sharp-1}-\overline{\xi}^{\delta}_{\to T^\sharp-1})+(T^\sharp-1)\eta,\\ \mathcal{B}&=\sum_{\tau=T^\sharp}^{T}\overline{\xi}^{r}_{\to\tau}[\Delta_\tau-\overline{\Delta}_{\to\tau}]-(T^\sharp-1)\overline{\xi}^{r}_{\to T^\sharp-1}\overline{\Delta}_{\to T^\sharp-1}\\&\quad+(T^\sharp-1)\overline{\xi}^{r}_{\to T^\sharp-1}\eta,\\ \mathcal{C}&=\sum_{\tau=T^\sharp}^{T}\overline{\xi}^{r}_{\to\tau}[\xi^\delta_\tau-\overline{\xi}^{\delta}_{\to\tau}]-(T^\sharp-1)\overline{\xi}^{r}_{\to T^\sharp-1}\overline{\xi}^{\delta}_{\to T^\sharp-1},\\ \mathcal{D}&=\sum_{\tau=T^\sharp}^{T}\left(\frac{M}{2}|\overline{\xi}^{r}_{\to\tau}|^2\,|\Delta_\tau-\overline{\Delta}_{\to\tau}+\overline{\xi}^{\delta}_{\to\tau}|-\frac{m}{2}|\Delta_\tau-\overline{\Delta}_{\to\tau}\right.\\&\quad\left.+\overline{\xi}^{\delta}_{\to\tau}|^2\right)+(T^\sharp-1)\frac{M^2}{8m}|\overline{\xi}^{r}_{\to T^\sharp-1}|^4,\\ \mathcal{E}&=\sum_{\tau=T^\sharp}^{T}\overline{\ell}^{*}_{\to\tau}+(T^\sharp-1)\overline{\ell}^{*}_{\to T^\sharp-1}.\end{aligned}$$

In the second step of the proof to Theorem 3, we analyze the five terms $\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D},\mathcal{E}$ separately. We see that many of the terms in $\mathcal{A},\mathcal{B},\mathcal{C}$, and $\mathcal{D}$ are equal to zero either almost surely or in expectation and the term $\mathcal{E}$ is upper bounded by a constant; thus, the expression in the upper bound in Equation (16) can be greatly simplified. Recall the definition $\overline{\Delta}_{\to\tau}=\frac{\Delta_T}{T-1}+\ldots+\frac{\Delta_{\tau+1}}{\tau}$, $\overline{\xi}^{\delta}_{\to\tau}=\frac{\xi^\delta_T}{T-1}+\ldots+\frac{\xi^\delta_{\tau+1}}{\tau}$. For the term $\mathcal{A}$, with elementary algebra, it can be verified that

$$\sum_{\tau=T^\sharp}^{T}[\Delta_\tau-\overline{\Delta}_{\to\tau}+\overline{\xi}^{\delta}_{\to\tau}-\xi^\delta_\tau]-(T^\sharp-1)(\overline{\Delta}_{\to T^\sharp-1}-\overline{\xi}^{\delta}_{\to T^\sharp-1})=0,\tag{17}$$

which implies that $\mathbb{E}[\mathcal{A}]=\mathbb{E}[(T^\sharp-1)\eta]$.

Reorganizing all terms in $\mathcal{B}$ by ξ_t ($\forall t\ge T^\sharp$), we obtain

$$\begin{aligned}&\sum_{\tau=T^\sharp}^{T}\overline{\xi}^{r}_{\to\tau}[\Delta_\tau-\overline{\Delta}_{\to\tau}]-(T^\sharp-1)\overline{\xi}^{r}_{\to T^\sharp-1}\overline{\Delta}_{\to T^\sharp-1}\\&\quad=\sum_{t=T^\sharp}^{T}\frac{\xi^r_t}{t-1}\sum_{\tau=T^\sharp}^{t-1}(\Delta_\tau-\overline{\Delta}_{\to\tau})-(T^\sharp-1)\overline{\xi}^{r}_{\to T^\sharp-1}\overline{\Delta}_{\to T^\sharp-1}\\&\quad=\sum_{t=T^\sharp}^{T}\frac{\xi^r_t}{t-1}[-(t-1)\overline{\Delta}_{\to t-1}+(T^\sharp-1)\overline{\Delta}_{\to T^\sharp-1}]\\&\qquad-(T^\sharp-1)\overline{\xi}^{r}_{\to T^\sharp-1}\overline{\Delta}_{\to T^\sharp-1}\\&\quad=-\sum_{t=T^\sharp}^{T}\xi^r_t\overline{\Delta}_{\to t-1},\end{aligned}$$

where the last equality uses $\overline{\xi}^{\delta}_{\to T^\sharp-1}=\frac{\xi^\delta_T}{T-1}+\ldots+\frac{\xi^\delta_{T^\sharp}}{T^\sharp-1}$.

Note that the random variable $\overline{\Delta}_{\to\tau-1}=\frac{\Delta_T}{T-1}+\ldots+\frac{\Delta_\tau}{\tau-1}$ is measurable with respect to $\mathcal{F}_{\tau+1}$ because the DP policy is nonanticipating. By Doob's optional stopping theorem, we have $\mathbb{E}[-\sum_{\tau=T^\sharp}^{T}\xi^r_\tau\overline{\Delta}_{\to\tau-1}]=0$, and thus,

$$\mathbb{E}[\mathcal{B}]=\mathbb{E}\Big[(T^\sharp-1)\overline{\xi}^{r}_{\to T^\sharp-1}\eta\Big].\tag{18}$$

Next, we analyze the term $\mathcal{C}$. Note that $\mathcal{C}$ has a similar structure as $\mathcal{B}$; therefore,

$$\begin{aligned}\mathbb{E}[\mathcal{C}]&=\mathbb{E}\left[-\sum_{t=T^\sharp}^{T}\xi^r_t\overline{\xi}^{\delta}_{\to t-1}\right]=\mathbb{E}\left[-\sum_{t=T^\sharp}^{T}\xi^r_t\left(\frac{\xi^\delta_T}{T-1}+\ldots+\frac{\xi^\delta_{t+1}}{t}+\frac{\xi^\delta_t}{t-1}\right)\right].\\&=\mathbb{E}\left[-\sum_{t=T^\sharp}^{T}\frac{\xi^r_t\xi^\delta_t}{t-1}\right],\end{aligned}$$

where the last equality uses Doob's optional stopping theorem. Because $|\xi^r_t|\le B_\xi$ *a.s.*, we have

$$\begin{aligned}\left|\mathbb{E}\left[-\sum_{t=T^\sharp}^{T}\frac{\xi^r_t\xi^\delta_t}{t-1}\right]\right|&\le\mathbb{E}\left[\sum_{t=T^\sharp}^{T}\frac{|\xi^r_t||\xi^r_t-\xi^*_t|}{t-1}\right]\\&\le B_\xi\mathbb{E}\left[\sum_{t=T^\sharp}^{T}\frac{|\xi^r_t-\xi^*_t|}{t-1}\right]=B_\xi\mathbb{E}\left[\sum_{t=2}^{T}\mathbf{1}\{T^\sharp\le t\}\frac{|\xi^r_t-\xi^*_t|}{t-1}\right]\\&=B_\xi\mathbb{E}\left[\sum_{t=2}^{T}\mathbf{1}\{T^\sharp\le t\}\frac{\mathbb{E}[|\xi^r_t-\xi^*_t||\,\mathcal{F}_{t+1}]}{t-1}\right]\\&\le B_\xi\mathbb{E}\left[\sum_{t=2}^{T}\mathbf{1}\{T^\sharp\le t\}\frac{\sqrt{\mathbb{E}[|\xi^r_t-\xi^*_t|^2\,|\,\mathcal{F}_{t+1}]}}{t-1}\right],\end{aligned}\tag{19}$$

where the last equality holds because the event $\{T^\sharp\le t\}=\{T^\sharp\ge t+1\}^c\in\mathcal{F}_{t+1}$.

Because the regret is defined as the *difference* between the expected revenues under the optimal DP policy and the resolving heuristic, we can choose the joint distribution of (ξ_t^*, ξ_t^r) arbitrarily as long as their marginal distributions remain the same. We choose the joint distributions as follows. At each time period t with posted prices p_t^*, p_t^r and corresponding demand rates $x_t^* + \Delta_t$ and x_t^r, let $(\xi_t^*, \xi_t^r) \sim \Xi_t$ such that the marginal distributions are $Q_t(p_t^*), Q_t(p_t^r)$, and furthermore,

$$\sqrt{\mathbb{E}\Big[|\xi_t^* - \xi_t^r|^2 \mid \mathcal{F}_{t+1}\Big]} = \sqrt{\mathbb{E}_{(\xi_t^*,\xi_t^r)\sim\Xi_t}\Big[|\xi_t^* - \xi_t^r|^2\Big]} \le L\,|f(p_t^*) - f(p_t^r)| = L\,|x_t^* + \Delta_t - x_t^r| \quad a.s.$$

The existence of such a joint distribution Ξ_t is implied by $\mathcal{W}_2(Q(p_t^*), Q(p_t^r)) \le L\,|f(p_t^*) - f(p_t^r)|$ (see Assumption A5). As a result, Equation (19) can be simplified to

$$|\mathbb{E}[\mathcal{C}]| \le LB_\xi \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T} \frac{|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}^\delta_{\to\tau}|}{\tau - 1}\right]. \tag{20}$$

Combining Equation (16) with Equations (17), (18), and (20), we have

$$\begin{aligned}
&\mathbb{E}\Big[r'(x_T)\mathcal{A} - r''(x_T)\mathcal{B} + r''(x_T)\mathcal{C} + \mathcal{D}\Big]\\
&\le \mathbb{E}\Bigg[(T^\sharp - 1)\Big(r'(x_T)\eta - r''(x_T)\overline{\xi}^r_{\to T^\sharp-1}\eta + \frac{M^2}{8m}|\overline{\xi}^r_{\to T^\sharp-1}|^4\Big)\\
&\quad + \sum_{\tau=T^\sharp}^{T}\Big(\Big(|r''(x_T)|\frac{LB_\xi}{\tau-1} + \frac{M}{2}|\overline{\xi}^r_{\to\tau}|^2\Big)|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}^\delta_{\to\tau}|\\
&\quad - \frac{m}{2}|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\xi}^\delta_{\to\tau}|^2\Big)\Bigg]\\
&\le E\Bigg[(T^\sharp - 1)\Big(r'(x_T)\eta - r''(x_T)\overline{\xi}^r_{\to T^\sharp-1}\eta + \frac{M^2}{8m}|\overline{\xi}^r_{\to T^\sharp-1}|^4\Big)\\
&\quad + \sum_{\tau=T^\sharp}^{T}\frac{1}{2m}\Big(|r''(x_T)|\frac{LB_\xi}{\tau-1} + \frac{M}{2}|\overline{\xi}^r_{\to\tau}|^2\Big)^2\Bigg]\\
&\le E\Bigg[(T^\sharp - 1)\Big(r'(x_T)\eta - r''(x_T)\overline{\xi}^r_{\to T^\sharp-1}\eta + \frac{M^2}{8m}|\overline{\xi}^r_{\to T^\sharp-1}|^4\Big)\\
&\quad + \sum_{\tau=T^\sharp}^{T}\frac{1}{m}\Big(\Big(|r''(x_T)|\frac{LB_\xi}{\tau-1}\Big)^2 + \frac{M^2}{4}|\overline{\xi}^r_{\to\tau}|^4\Big)\Bigg].
\end{aligned} \tag{21}$$

In the second inequality, we use the fact that

$$-\frac{m}{2}u^2 + bu = -\frac{m}{2}(u - \frac{b}{m})^2 + \frac{b^2}{2m} \le \frac{b^2}{2m}$$

for any $b, u \in \mathbb{R}$ and $m > 0$. In the third inequality, we use the fact that $(a+b)^2 \le 2(a^2 + b^2)$ for any $a, b \in \mathbb{R}$.

Because $\overline{\xi}^r_{\to T^\sharp-1} = \overline{\xi}^r_{\to T^\sharp} + \xi^r_{T^\sharp}/(T^\sharp - 1) \ge \overline{\xi}^r_{\to T^\sharp} - B_\xi/(T^\sharp - 1) \ge r'(x_T)/r''(x_T)$ by the definition of $T^\sharp$ in Equation (8), it ensures that $r'(x_T) - r''(x_T)\overline{\xi}^r_{\to T^\sharp-1} \ge 0$. Recall that $\eta = -(x^*_{T^\sharp-1} - \lambda^u)^+ \le 0$, so $(r'(x_T) - r''(x_T)\overline{\xi}^r_{\to T^\sharp-1})\eta \le 0$, and hence, the first term in Equation (21) can be eliminated. We get

$$\begin{aligned}
&\mathbb{E}[r'(x_T)\mathcal{A} - r''(x_T)\mathcal{B} + r''(x_T)\mathcal{C} + \mathcal{D}]\\
&\le \mathbb{E}\Bigg[(T^\sharp - 1)\frac{M^2}{8m}|\overline{\xi}^r_{\to T^\sharp-1}|^4\\
&\quad + \sum_{\tau=T^\sharp}^{T}\frac{1}{m}\Big((|r''(x_T)|\frac{LB_\xi}{\tau-1}\Big)^2 + \frac{M^2}{4}|\overline{\xi}^r_{\to\tau}|^4\Big)\Bigg].
\end{aligned} \tag{22}$$

For the term $\mathcal{E}$, notice that, in any sample path, lost sales can occur in at most one period during the entire horizon. Suppose $\tau' \ge T^\sharp$ is the period in which the lost sales occur (otherwise, $\mathcal{E} = 0$ if lost sales do not occur before $T^\sharp$).
Because

$$\overline{\ell}^*_{\to\tau} = \frac{\ell^*_T}{T-1} + \frac{\ell^*_{T-1}}{T-2} + \ldots + \frac{\ell^*_{\tau+1}}{\tau} = \frac{\ell^*_{\tau'}}{\tau'-1}$$

for any $\tau < \tau'$, we have

$$\begin{aligned}
\mathcal{E} &= \sum_{\tau=T^\sharp}^{T}\overline{\ell}^*_{\to\tau} + (T^\sharp - 1)\overline{\ell}^*_{\to T^\sharp-1} = \sum_{\tau=T^\sharp}^{\tau'-1}\frac{\ell^*_{\tau'}}{\tau'-1} + (T^\sharp - 1)\frac{\ell^*_{\tau'}}{\tau'-1}\\
&= (\tau' - 1)\frac{\ell^*_{\tau'}}{\tau'-1} = \ell^*_{\tau'}.
\end{aligned}$$

The lost sales are upper bounded by the realized demand in period τ', so $\ell^*_{\tau'} \le \lambda^*_{\tau'} + \xi^*_{\tau'}$. Because $\lambda^*_{\tau'} \le \overline{d}$ by Assumption A1 and $\xi^*_{\tau'} \le B_\xi$ by Assumption A4, $\mathcal{E} \le \overline{d} + B_\xi$. By Equation (22), we have

$$\begin{aligned}
&\mathbb{E}[r'(x_T)\mathcal{A} - r''(x_T)\mathcal{B} + r''(x_T)\mathcal{C} + \mathcal{D} + r'(\underline{d})\mathcal{E}]\\
&\le \mathbb{E}\Bigg[(T^\sharp - 1)\frac{M^2}{8m}|\overline{\xi}^r_{\to T^\sharp-1}|^4\\
&\quad + \sum_{\tau=T^\sharp}^{T}\frac{1}{m}\Big((|r''(x_T)|\frac{LB_\xi}{\tau-1})^2 + \frac{M^2}{4}|\overline{\xi}^r_{\to\tau}|^4\Big)\Bigg] + r'(\underline{d})(\overline{d} + B_\xi).
\end{aligned} \tag{23}$$

In the final step of our proof to Theorem 3, we upper bound each term in Equation (23). First, it is easy to verify that

$$\begin{aligned}
&\sum_{\tau=T^\sharp}^{T}\left(\frac{|r''(x_T)|\,LB_\xi}{\tau-1}\right)^2\\
&\le (|r''(x_T)|\,LB_\xi)^2\sum_{j=1}^{T-1}\frac{1}{j^2} \le 2(|r''(x_T)|\,LB_\xi)^2 \quad a.s.
\end{aligned} \tag{24}$$

We next focus on the terms involving $|\overline{\xi}^r_{\to\tau}|^4$. Recall the definition $\overline{\xi}^r_{\to t} = \frac{\xi^r_T}{T-1} + \ldots + \frac{\xi^r_{t+1}}{t}$. Let $z_t := \overline{\xi}^r_{\to t-1}$. Then, $\{z_t\}$ is a martingale adapted to the filtration $\{\mathcal{F}_t\}_{t=1}^T$ by Assumption A4, which implies that $\{|z_t|^4\}$ is a submartingale. Let $S_t = \sum_{\tau=t}^T (t-1)(|z_\tau|^4 - |z_{\tau+1}|^4)$; then, $\{S_t\}$ is also a submartingale. Because $T^\sharp$ is a stopping time, we have

$$\begin{aligned}
&\mathbb{E}\left[(T^\sharp - 1)|\overline{\xi}^r_{\to T^\sharp-1}|^4 + \sum_{\tau=T^\sharp}^{T}|\overline{\xi}^r_{\to\tau}|^4\right] = \mathbb{E}[S_{T^\sharp}] \le \mathbb{E}[S_1]\\
&= \mathbb{E}\left[\sum_{\tau=2}^{T}|z_\tau|^4\right] = \mathbb{E}\left[\sum_{\tau=1}^{T-1}|\overline{\xi}^r_{\to\tau}|^4\right].
\end{aligned}$$

It is easy to verify that

$$\mathbb{E}[|\overline{\xi}^r_{\to t}|^4] = \sum_{j,k>t} \frac{\mathbb{E}[|\xi^r_j|^2 |\xi^r_k|^2]}{(j-1)^2(k-1)^2} \le B^4_\xi \left(\sum_{j>t} \frac{1}{(j-1)^2}\right)^2 \le \frac{4B^4_\xi}{t^2}.$$

Subsequently,

$$\mathbb{E}\left[\frac{M^2}{8m}(T^\sharp - 1)\,|\overline{\xi}^r_{\to T^\sharp-1}|^4 + \frac{M^2}{4m}\sum_{\tau=T^\sharp}^{T} |\overline{\xi}^r_{\to\tau}|^4\right] \le \frac{M^2}{4m}\sum_{\tau=1}^{\mathbb{T}-1}\frac{4B^4_\xi}{\tau^2} \le \frac{2M^2B^4_\xi}{m}. \tag{25}$$

Combining Equations (16) and (23)–(25), we have

$$\phi^*_T(x_T) - \phi^r_T(x_T) \le \frac{2M^2B^4_\xi}{m} + \frac{2(|r''(x_T)|\,LB_\xi)^2}{m} + r'(\underline{d})(\overline{d} + B_\xi) + r^u\mathbb{E}[T^\sharp] = O(1), \tag{26}$$

where the last equality holds by applying Lemma 1. This completes the proof of Theorem 3.

4.3. Proof of Theorem 4

Because this proof only concerns the resolving policy, for convenience we drop the superscript r and denote $\xi^r_\tau, \overline{\xi}^r_{\to\tau}$ simply by $\xi_\tau, \overline{\xi}_{\to\tau}$ in this section.

Invoking Lemma 2, we have

$$\begin{aligned} Tr(x_T) - \phi^r_T(x_T) &\ge \mathbb{E}\Bigg[\sum_{\tau=T^\sharp}^{T}(r(x_T) - r(x^r_\tau)) \\ &\qquad + (T^\sharp - 1)(r(x_T) - r(\min\{x^r_{T^\sharp-1}, \lambda^u\}))\Bigg] \\ &= \mathbb{E}\Bigg[\sum_{\tau=T^\sharp}^{T}(r(x_T) - r(x^r_\tau)) + (T^\sharp - 1)(r(x_T) \\ &\qquad - r(x^r_{T^\sharp-1}))\Bigg] \\ &\quad + \mathbb{E}\big[(T^\sharp - 1)(r(x^r_{T^\sharp-1}) - r(\min\{x^r_{T^\sharp-1}, \lambda^u\}))\big]. \end{aligned} \tag{27}$$

By the smoothness of $r(\lambda)$ (Assumption A3) and the fact that $x^r_{T^\sharp} \le \lambda^u$ (see Equation (8)), the second term in Equation (27) is bounded by

$$\begin{aligned} &\mathbb{E}\big[(T^\sharp - 1)(r(x^r_{T^\sharp-1}) - r(\min\{x^r_{T^\sharp-1}, \lambda^u\})\big] \\ &\quad \ge -C\mathbb{E}\big[(T^\sharp - 1)(x^r_{T^\sharp-1} - \lambda^u)^+\big] \\ &\quad \ge -C\mathbb{E}\big[(T^\sharp - 1)\,|x^r_{T^\sharp-1} - x^r_{T^\sharp}|\big] \\ &\quad = -C\mathbb{E}\left[(T^\sharp - 1)\left|\frac{\xi_{T^\sharp}}{T^\sharp - 1}\right|\right] \ge -CB_\xi, \end{aligned}$$

where the last inequality holds because $|\xi_\tau| \le B_\xi$ *a.s* for all τ.

In the rest of the proof, we bound the first term in Equation (27). Expanding the difference $r(x^r_\tau) - r(x_T)$ at x_T and using the strict concavity of $r(\lambda)$ (Assumption A2), by Equation (9),

$$r(x_T) - r(x^r_\tau) = r(x_T) - r(x_T - \overline{\xi}_{\to\tau}) \ge r'(x_T)\overline{\xi}_{\to\tau} + \frac{m}{2}|\overline{\xi}_{\to\tau}|^2, \quad \forall\, \tau \ge T^\sharp$$

$$r(x_T) - r(x^r_{T^\sharp-1}) = r(x_T) - r((x_T - \overline{\xi}_{\to T^\sharp-1})^+) \ge r'(x_T)\overline{\xi}_{\to T^\sharp-1} + \frac{m}{2}|\overline{\xi}_{\to T^\sharp-1}|^2.$$

Therefore,

$$\begin{aligned} &\mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}(r(x_T) - r(x^r_\tau)) + (T^\sharp - 1)(r(x_T) - r(x^r_{T^\sharp-1}))\right] \\ &\ge r'(x_T)\mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}\overline{\xi}_{\to\tau} + (T^\sharp - 1)\overline{\xi}_{\to T^\sharp-1}\right] \\ &\quad + \frac{m}{2}\mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}|\overline{\xi}_{\to\tau}|^2 + (T^\sharp - 1)\,|\,\overline{\xi}_{\to T^\sharp-1}|^2\right]. \end{aligned} \tag{28}$$

For the first term in Equation (28), because $\overline{\xi}_\tau = \sum_{t=\tau+1}^{T}\xi_t/(t-1)$, it holds that

$$\mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}\overline{\xi}_{\to\tau} + (T^\sharp - 1)\overline{\xi}_{\to T^\sharp-1}\right] = \mathbb{E}\left[\sum_{t=T^\sharp}^{T}\xi_t\right] = 0 \tag{29}$$

by Doob's optional stopping theorem. For the second term in Equation (28), using $\overline{\xi}_{\to\tau} = \frac{\xi_T}{T-1} + \ldots + \frac{\xi_{\tau+1}}{\tau}$, we have

$$\begin{aligned} &\mathbb{E}\left[\sum_{\tau=1}^{T}|\overline{\xi}_{\to\tau}|^2\right] - \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}|\overline{\xi}_{\to\tau}|^2 + (T^\sharp - 1)\,|\,\overline{\xi}_{\to T^\sharp-1}|^2\right] \\ &\qquad = \mathbb{E}\left[\sum_{\tau=1}^{T^\sharp-2}\left(\overline{\xi}^2_{\to\tau} - \overline{\xi}^2_{\to T^\sharp-1}\right)\right] \\ &= 2\mathbb{E}\left[\sum_{\tau=1}^{T^\sharp-2}(\overline{\xi}_{\to\tau} - \overline{\xi}_{\to T^\sharp-1})\overline{\xi}_{\to T^\sharp-1}\right] + \mathbb{E}\left[\sum_{\tau=1}^{T^\sharp-2}(\overline{\xi}_{\to\tau} - \overline{\xi}_{\to T^\sharp-1})^2\right] \\ &= 2\mathbb{E}\left[\overline{\xi}_{\to T^\sharp-1}\sum_{\tau=1}^{T^\sharp-2}\sum_{j=\tau+1}^{T^\sharp-1}\frac{\xi_j}{j-1}\right] + \mathbb{E}\left[\sum_{\tau=1}^{T^\sharp-2}\left(\sum_{j=\tau+1}^{T^\sharp-1}\frac{\xi_j}{j-1}\right)^2\right]. \end{aligned} \tag{30}$$

Because $T^\sharp$ is a stopping time and $\{\xi_\tau\}$ is a martingale difference sequence, we have $\mathbb{E}[\xi_j \,|\, j < T^\sharp] = 0$ and $\mathbb{E}[\xi_j\xi_k \,|\, j < k < T^\sharp] = 0$. So we have

$$\begin{aligned} \text{Eq.}(30) &= 0 + \mathbb{E}\left[\sum_{\tau=1}^{T^\sharp\text{-}2}\sum_{j=\tau+1}^{T^\sharp\text{-}1}\left(\frac{\xi_j}{j\text{-}1}\right)^2\right] = \mathbb{E}\left[\sum_{j=2}^{T^\sharp\text{-}1}(j\text{-}1)\left(\frac{\xi_j}{j\text{-}1}\right)^2\right] \\ &\le \mathbb{E}\left[\sum_{j=2}^{T^\sharp\text{-}1}\xi^2_{j\text{-}1}\right] \le B^2_\xi\mathbb{E}[T^\sharp\text{-}2]. \end{aligned}$$

Moreover, $\mathbb{E}[\sum_{\tau=1}^{T}\overline{\xi}^2_{\to\tau}] = \mathbb{E}[\sum_{j=2}^{T}\frac{\xi^2_j}{j-1}] \ge \sigma^2\sum_{j=2}^{T}\frac{1}{j-1} \ge \sigma^2 \int_1^T u^{-1}du = \sigma^2\ln T$, so by the preceding equation and Equation (30), we have

$$\begin{aligned} \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}|\overline{\xi}_{\to\tau}|^2 + (T^\sharp - 1)\,|\overline{\xi}_{\to T^\sharp-1}|^2\right] &\ge \sigma^2\ln T - B^2_\xi\mathbb{E}[T^\sharp - 2] \\ &= \sigma^2\ln T - O(1), \end{aligned} \tag{31}$$

where the last equality holds by Lemma 1.

Combining Equations (27)–(29) and (31), we get $Tr(x_T) - \phi^r_T(x_T) \ge \frac{m\sigma^2}{2}\ln T - O(1)$, and the proof of Theorem 4 is complete.

5. Numerical Results

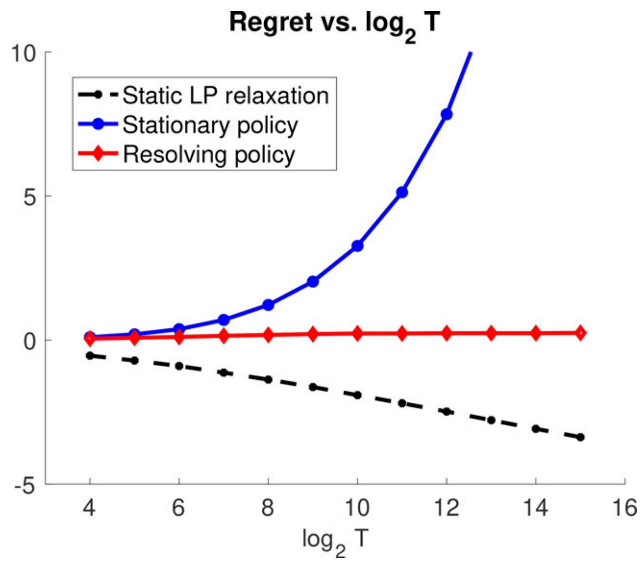
We corroborate the theoretical findings in the last section with a few simple numerical experiments. In the simulation, we assume a linear demand curve with Bernoulli demand distribution: $\mathbb{P}[d_t = 1 \mid p_t] = \alpha - \beta p_t$, $\mathbb{P}[d_t = 0 \mid p_t] = 1 - \mathbb{P}[d_t = 1 \mid p_t]$ with $p \in [0,1]$, $\alpha = 3/4$ and $\beta = 1/2$. The (normalized) initial inventory level is $x_T = 5/16$, which means that, for problem instances with T periods, the initial inventory level is $x_T T = 5T/16$. It is easy to verify that the optimal demand rate λ^u without inventory constraints is $\lambda^u = 3/8 > x_T$, and the fluid approximation suggests a $Tr(x_T) = (19/32)T = 0.59375T$ expected revenue. We select the Bernoulli demand distribution because the states of inventory levels are discrete, and therefore, the dynamic programming model can be calculated exactly.

In Table 2, we report the regret of the fluid approximation, the static policy $\pi^s : p_t \equiv f^{-1}(x_T) = 7/8$, and the resolving heuristics π^r. All regret is defined with respect to the value (expected reward) of the optimal DP pricing policy, and the regret for the fluid approximation benchmark is negative because the fluid model always upper bounds the value of any policy. Both the static policy π^s and the resolving heuristics π^r are run for each value of T ranging from $T = 2^6 = 64$ to $T = 2^{15} = 32{,}768$ to obtain an accurate estimation of their expected rewards. We plot the horizontal axis of Figure 2 in the logarithmic scale to show the regret growth rate of each policy.

As we can see from Table 2 and Figure 2, the gap between the value of the optimal policy and the value of the fluid model grows with a nearly constant slope as the number of periods T grows exponentially, which verifies the $\Omega(\ln T)$ growth rate established in Theorem 4. On the other hand, the growth of regret of the resolving heuristics π^r stagnated at $T \geq 2^{10}$ and is nearly the same for any T ranging from 2^{10} to 2^{15}. This shows the asymptotic growth of regret of π^r is far slower than $O(\ln T)$ and is compatible with the $O(1)$ regret upper bound we prove in Theorem 3.

We report additional sets of numerical results in Figure 3, in which we only report the cumulative regret of the resolving heuristic compared against the benchmark of the optimal DP policy. On the left panel of Figure 3, we report the regret of the resolving heuristic with T ranging from $2^4 = 16$ to $2^{20} \approx 1{,}000{,}000$ and different $\lambda^u - x_T$ gap values. More specifically, all four curves are reported under the demand model $d = 0.75 - 0.5p$ with unconstrained optimum $\lambda^u = .375$ and (normalized) initial inventory levels $x_T \in \{0.3, 0.325, 0.35, 0.375\}$. Figure 3 clearly shows that the regret of

Figure 2. (Color online) Plots of Regret of the Fluid Model, the Static Policy π^s, and the Resolving Heuristics π^r Compared Against the Value of the Optimal DP Pricing Policy



the resolving heuristic increases as the gap between λ^u and x_T narrows, and furthermore, in the boundary case (i.e., $x_T = \lambda^u$), the regret seems to grow logarithmically in T. It would be an interesting direction of future research to formally establish the logarithmic regret for the boundary case and explore alternative policies that attain constant regret with $x_T = \lambda^u$.

On the right panel of Figure 3, we report the regret of the resolving heuristic with different a, b values in the demand model $d = a - bp$ with $a = b \in \{0.3, 0.5, 0.7, 0.9\}$. The normalized initial inventory level is fixed at $x_T = 0.1$. With different values of the slopes, the demand and revenue models exhibit different strong concavity parameter values with $r''(p) = -b/2$ and $r''(d) = -1/(2b)$. Unlike the gap $\lambda^u - x_T$, the results reported in the right panel of Figure 3 do not paint a clear picture of the role the strong concavity parameters play in the regret. Overall, intermediate values $(b = 0.5, r''(p) = -0.25, r''(d) = -1)$ seem to result in the lowest regret of the resolving heuristic.

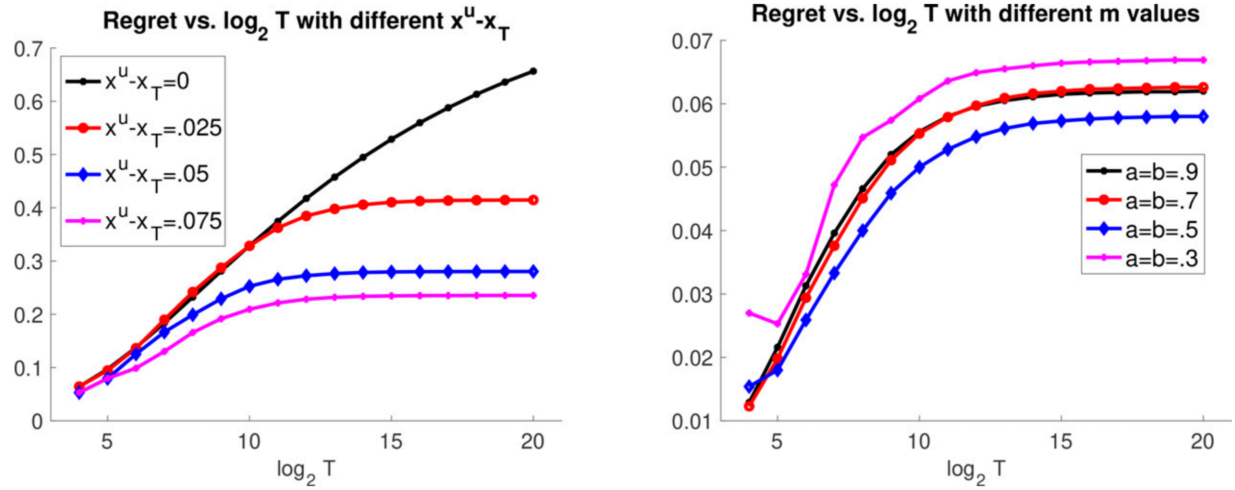
6. Extension to Network Revenue Management

In this section, we extend the analysis of the resolving heuristic to a general network revenue management

Table 2. Regret for the Fluid Model, the Optimal Static Policy π^{ps}, and the Resolving Heuristics π^{pr} Compared Against the Value of the Optimal DP Pricing Policy

$\log_2 T$	6	7	8	9	10	11	12	13	14	15
Fluid model	−0.90	−1.13	−1.37	−1.63	−1.91	−2.19	−2.48	−2.78	−3.08	−3.37
Static policy π^{ps}	0.38	0.70	1.22	2.03	3.27	5.13	7.84	11.81	17.55	25.84
Resolving policy π^{pr}	0.11	0.15	0.18	0.21	0.23	0.23	0.24	0.24	0.24	0.25

Figure 3. (Color online) Cumulative Regret of the Resolving Heuristic (Compared with the Optimal DP Value) Under Different $\lambda^u - x_T$ and m Settings



model with n products (with possibly correlated demand) and m resources. Recall the constant matrix $\mathbf{A} = [a_{ij}]_{1\le i\le m, 1\le j\le n}$ indicates that a product j requires a_{ij} units of resource i. The seller starts in period T with an initial resource inventory level $\mathbf{y}_T = \mathbf{x}_T T$, where $\mathbf{x}_T = (x_T(1), \ldots, x_T(m)) \in \mathbb{R}_+^m$ is the normalized initial inventory. At period t (i.e., t periods remaining), the seller posts a price vector $\mathbf{p}_t = (p_t(1), \ldots, p_t(n)) \in \mathcal{P} \subset \mathbb{R}_+^n$ and observes a realized demand vector $\mathbf{d}_t = \mathbf{f}(\mathbf{p}_t) + \boldsymbol{\varepsilon}_t$, where $\mathbf{f}: \mathcal{P} \to \mathcal{D}$ is the demand curve and $\boldsymbol{\varepsilon}_t \sim Q(\mathbf{p}_t)$ is a centered noise vector. Let $0 \le \mathbf{s}_t \le \mathbf{d}_t$ be the fulfilled demand defined in Equation (1) and $\mathbf{l}_t = \mathbf{d}_t - \mathbf{s}_t$ be the lost sales. The realized revenue at period t is $r_t = \langle \mathbf{p}_t, \mathbf{s}_t \rangle$.

Throughout this section, we use $\|\mathbf{x}\|$ to denote the Euclidean norm of any vector $\mathbf{x}$, and $\|\mathbf{X}\|_2$ to denote the spectral norm of any matrix $\mathbf{X}$. We extend the demand Assumptions A1–A5 in the single product setting to the multiproduct setting as follows:

B1. (Invertibility). The demand rate function $\mathbf{f}: \mathcal{P} \to \mathcal{D}$ is a bijection, where $\mathcal{P} \subset \mathbb{R}_+^n, \mathcal{D} \subset \mathbb{R}_+^n$. Let $\mathbf{f}^{-1}: \mathcal{D} \to \mathcal{P}$ denote its inverse function. Assume $\mathcal{D}$ is convex, compact, has nonempty interior, and satisfies $\mathbf{0} \in \mathcal{D}$.

B2. (Strict Concavity). The expected revenue $r(\boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \mathbf{f}^{-1}(\boldsymbol{\lambda}) \rangle$ as a function of the demand rate vector $\boldsymbol{\lambda}$ is strictly concave. That is, there exists a positive constant $m' > 0$ such that $\nabla^2 r(\boldsymbol{\lambda}) \preceq -m'\mathbf{I}_n$ for all $\boldsymbol{\lambda} \in \mathcal{D}$. The maximizer of $r(\boldsymbol{\lambda})$ is in the interior of the domain $\mathcal{D}$.

B3. (Smoothness). $r(\boldsymbol{\lambda})$ is three times continuously differentiable in the interior of $\mathcal{D}$.

B4. (Martingale Difference Sequence). Conditional on any price vector $\mathbf{p}_t \in \mathcal{P}$ ($\forall t = 1, \ldots, T$), the demand noise $\boldsymbol{\varepsilon}_t$ is independent of $\{\boldsymbol{\varepsilon}_T, \ldots, \boldsymbol{\varepsilon}_{t+1}\}$ and satisfies $\mathbb{E}[\boldsymbol{\varepsilon}_t \mid \mathcal{F}_{t+1}] = \mathbb{E}[\boldsymbol{\varepsilon}_t \mid \mathbf{p}_t] = \mathbf{0}$ *a.s.* The conditional distribution of $\boldsymbol{\varepsilon}_t$ given $\mathbf{p}_t$ is denoted by $\boldsymbol{\varepsilon}_t \sim Q(\mathbf{p}_t)$. In addition, let $\boldsymbol{\xi}_t := \mathbf{A}\boldsymbol{\varepsilon}_t$; then, $\|\boldsymbol{\xi}_t\| \le B_\xi$ *a.s.* for some constant $B_\xi < \infty$.

B5. (Bounded Wasserstein Distance). There exists a constant $L > 0$ such that, for any $\mathbf{p}, \mathbf{p}' \in \mathcal{P}$, it holds that $\mathcal{W}_2(Q(\mathbf{p}), Q(\mathbf{p}')) \le L \|\mathbf{f}(\mathbf{p}) - \mathbf{f}(\mathbf{p}')\|$, where $\mathcal{W}_2(Q, Q') := \inf_\Xi \sqrt{\mathbb{E}_{\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}' \sim \Xi}[\|\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}'\|^2]}$ is the L_2-Wasserstein distance between Q, Q' with Ξ being an arbitrary joint distribution with marginal distributions being Q and Q', respectively.

Recall that $\mathbf{x}_T \in \mathbb{R}_+^m$ is the normalized inventory level vector of the m resources at the beginning of period T, namely, the total inventory is $\mathbf{x}_T T$. The fluid approximation model is formulated as

$$\max_{\boldsymbol{\lambda} \in \mathcal{D}} \{r(\boldsymbol{\lambda}), \text{ s.t. } \mathbf{0} \le \mathbf{A}\boldsymbol{\lambda} \le \mathbf{x}_T\}.$$

At the beginning of each period $t = T, \ldots, 1$, the resolving heuristic solves

$$\max_{\boldsymbol{\lambda} \in \mathcal{D}} \{r(\boldsymbol{\lambda}), \text{ s.t. } \mathbf{0} \le \mathbf{A}\boldsymbol{\lambda} \le \mathbf{x}_t\}, \tag{32}$$

where $\mathbf{x}_t$ is the normalized inventory level at the beginning of period t. Because $\mathbf{0} \in \mathcal{D}$ by Assumption B1, Equation (32) always has a feasible solution. Let the optimal solution to (32) be $\boldsymbol{\lambda}_t$. The resolving policy sets the price vector to $\mathbf{p}_t = \mathbf{f}^{-1}(\boldsymbol{\lambda}_t)$ for period t. In addition, let $\boldsymbol{\rho}_t := \mathbf{A}\boldsymbol{\lambda}_t$ denote the expected resource consumption rate given the demand rate $\boldsymbol{\lambda}_t$.

In (32), for any given $\mathbf{x}_t \in \mathbb{R}_+^m$, the m resources are partitioned into two disjoint sets based on whether the inventory constraints are active. The *inventory-constrained* resource set $\mathcal{I}$ and the *inventory-unconstrained* resource set $\mathcal{U}$ are defined as

$$\begin{aligned} \mathcal{I} &:= \{k \in [m] : \boldsymbol{\rho}_t(k) = \mathbf{x}_t(k)\}, \\ \mathcal{U} &:= \{k \in [m] : \boldsymbol{\rho}_t(k) < \mathbf{x}_t(k)\}. \end{aligned} \tag{33}$$

As a special case, in the single product setting of Section 3 (n=m=1), we have $\mathcal{I} = \emptyset, \mathcal{U} = \{1\}$ if $x_t > \lambda^u = \arg\max_{\lambda \in [\underline{d}, \bar{d}]} r(\lambda)$ and $\mathcal{I} = \{1\}, \mathcal{U} = \emptyset$ if $x_t \le \lambda^u$.

From the definitions in Equations (32) and (33), it is clear that the sets $\mathcal{I}$ and $\mathcal{U}$ are uniquely determined by the inventory vector $\mathbf{x}_t$, which serves as the right-hand side of the fluid problem (32). We may, thus, write $\mathcal{I}(\mathbf{x}_t), \mathcal{U}(\mathbf{x}_t)$ to emphasize such dependency. This leads to a partition of $\mathbb{R}_+^m$ into 2^m subregions $\{\mathcal{S}_\mathcal{I},\ \forall \mathcal{I} \subset [m]\}$, where

$$\mathcal{S}_\mathcal{I} = \{\mathbf{x} \in \mathbb{R}_+^m : \mathcal{I}(\mathbf{x}) = \mathcal{I}\}. \tag{34}$$

We specify an additional assumption on the initial inventory level $\mathbf{x}_T$ for the network revenue management setting:

C1. The initial inventory level $\mathbf{x}_T$ is in the interior of $\mathcal{S}_\mathcal{I}$ for some $\mathcal{I} \subset [m]$. That is, given $\mathbf{x}_T \in \mathcal{S}_\mathcal{I}$, there exists a neighborhood $B_{\delta_0}(\mathbf{x}_T) = \{\mathbf{x}' \in \mathbb{R}_+^m : \|\mathbf{x}' - \mathbf{x}_T\| \le \delta_0\}$ with $\delta_0 > 0$ such that $B_{\delta_0}(\mathbf{x}_T) \subset \mathcal{S}_\mathcal{I}$.

Intuitively, Assumption C1 asserts that, when the constrained inventory level $\mathbf{x}_t$ fluctuates in a close neighborhood of $\mathbf{x}_T$, the set of active constraints in the fluid problem Equation (32) remains unchanged. In the single product case ($n = m = 1$), Assumption C1 reduces to the condition $x_T \ne \lambda^u$.

We now characterize the expected value of the optimal DP policy and the resolving heuristic π^r for the price-based network revenue management problem. Let $\phi_\tau^*(\mathbf{x}_\tau^*)$ be the expected revenue of the optimal policy π^* when there are τ time periods left with the (normalized) inventory level being $\mathbf{x}_\tau^* \in \mathbb{R}_+^m$. Similar to the single-product case, we have the following Bellman equation:

$$\phi_\tau^*(\mathbf{x}_\tau^*) = \max_{\boldsymbol{\lambda}\in\mathcal{D}} \Bigg\{ r(\boldsymbol{\lambda}) + \mathbb{E}_{\boldsymbol{\varepsilon}_\tau \sim Q(\mathbf{f}^{-1}(\boldsymbol{\lambda}))} \left[\phi_{\tau-1}^*\left(\mathbf{x}_\tau^* - \frac{\mathbf{A}(\boldsymbol{\lambda} + \boldsymbol{\varepsilon}_\tau - \mathbf{l}_t) - \mathbf{x}_\tau^*}{\tau - 1}\right) - \mathbf{f}^{-1}(\boldsymbol{\lambda})^{\mathrm{T}}\mathbf{l}_t \right] \Bigg\} \quad \forall \tau = T, \ldots, 1.$$

We denote the maximizer of the Bellman equation by $\boldsymbol{\lambda}_\tau^*$, so the optimal policy selects the price vector $\mathbf{p}_\tau = \mathbf{f}^{-1}(\boldsymbol{\lambda}_\tau^*)$. Let $\boldsymbol{\varepsilon}_\tau^*$ be the realized demand noise under this price and let $\mathbf{l}_\tau^* = (\boldsymbol{\lambda}_\tau^* + \boldsymbol{\varepsilon}_\tau^* - \tau\mathbf{x}_\tau^*)^+$ be the realized lost sales. We define $\boldsymbol{\Delta}_\tau := \mathbf{A}\boldsymbol{\lambda}_\tau^* - \mathbf{x}_\tau^*$, $\boldsymbol{\xi}_\tau^* := \mathbf{A}\boldsymbol{\varepsilon}_\tau^*$, and $\boldsymbol{\ell}_\tau^* := \mathbf{A}\mathbf{l}_\tau^*$, so the Bellman equation can be rewritten as

$$\phi_\tau^*(\mathbf{x}_\tau^*) = r(\boldsymbol{\lambda}_\tau^*) + \mathbb{E}\left[\phi_{\tau-1}^*\left(\mathbf{x}_\tau^* - \frac{\boldsymbol{\Delta}_\tau + \boldsymbol{\xi}_\tau^* - \boldsymbol{\ell}_t^*}{\tau - 1}\right) - \mathbf{f}^{-1}(\boldsymbol{\lambda}_\tau^*)^{\mathrm{T}}\mathbf{l}_t^*\right] \quad \forall \tau = T, \ldots, 1. \tag{35}$$

The resolving heuristic π^r and its expected revenue $\phi_\tau^r(\mathbf{x}_\tau^r)$, on the other hand, satisfy the following recursive equation. Recall that $\boldsymbol{\lambda}_\tau = \arg\max_{\boldsymbol{\lambda}\in\mathcal{D}}\{r(\boldsymbol{\lambda}),$ s.t. $\mathbf{0} \le \mathbf{A}\boldsymbol{\lambda} \le \mathbf{x}_\tau^r\}$ is the solution to the fluid model with $\mathbf{x}_\tau^r$ being the right-hand side and $\boldsymbol{\rho}_\tau := \mathbf{A}\boldsymbol{\lambda}_\tau$. Define $\boldsymbol{\xi}_\tau^r := \mathbf{A}\boldsymbol{\varepsilon}_\tau^r$ and $\boldsymbol{\ell}_\tau^r := \mathbf{A}\mathbf{l}_\tau^r$, where $\boldsymbol{\varepsilon}_\tau^r$ is the noise vector associated with the price vector $\boldsymbol{\lambda}_\tau$ and $\mathbf{l}_\tau^r$ is the realized lost sales. We have

$$\begin{aligned}\phi_\tau^r(\mathbf{x}_\tau^r) &= r(\boldsymbol{\lambda}_\tau) + \mathbb{E}\left[\phi_{\tau-1}^r\left(\mathbf{x}_\tau^r - \frac{\mathbf{A}(\boldsymbol{\lambda}_\tau + \boldsymbol{\varepsilon}_\tau^r) - \mathbf{x}_\tau^r + \boldsymbol{\ell}_\tau^r}{\tau - 1}\right) - \mathbf{f}^{-1}(\boldsymbol{\lambda}_\tau)^{\mathrm{T}}\mathbf{l}_t^r\right] \\ &= r(\boldsymbol{\lambda}_\tau) + \mathbb{E}\left[\phi_{\tau-1}^r\left(\mathbf{x}_\tau^r - \frac{\boldsymbol{\rho}_\tau + \boldsymbol{\xi}_\tau^r - \mathbf{x}_\tau^r + \boldsymbol{\ell}_\tau^r}{\tau - 1}\right) - \mathbf{f}^{-1}(\boldsymbol{\lambda}_\tau)^{\mathrm{T}}\mathbf{l}_t^r\right] \quad \forall \tau = T, \ldots, 1.\end{aligned} \tag{36}$$

The following theorem extends the constant regret result of the resolving heuristic in Section 3 to the multiproduct setting.

Theorem 5. *Given a demand function* $\mathbf{f}$ *and an initial inventory level* $\mathbf{x}_T \in \mathcal{S}_\mathcal{I}$ *satisfying Assumptions* B1–B5 *and* C1, *for all* $T \ge 1$, *we have*

$$\phi_T^r(\mathbf{x}_T) \ge \phi_T^*(\mathbf{x}_T) - O(1).$$

In the rest of this section, we prove Theorem 5.

6.1. Partial Optimization on a Subset of Resources

Compared with the single-product setting, the analysis for the multiple resource setting is complicated by the fact that, in Equation (32), some of the resources have binding inventory constraints, whereas other resources have nonbinding inventory constraints. We introduce some tools in this section to handle this complication.

For any vector $\mathbf{x} \in \mathbb{R}^m$ and subset $\mathcal{S} \subset [m]$, denote $\mathbf{x}(\mathcal{S}) = (\mathbf{x}(k) : k \in \mathcal{S})$ as the $|\mathcal{S}|$-dimensional subvector of $\mathbf{x}$ whose coordinates are restricted to the subset $\mathcal{S}$. The following observation follows immediately from the definition of $\mathcal{S}_\mathcal{I}$ in Equation (34), and the proof is omitted.

Lemma 3. *Suppose* $\mathbf{x} \in \mathcal{S}_\mathcal{I}$ *for some* $\mathcal{I} \subset [m]$. *Let* $\mathcal{U} = [m]\backslash\mathcal{I}$. *For any inventory vector* $\mathbf{x}' \in \mathbb{R}_+^m$ *with* $\mathbf{x}'(\mathcal{I}) = \mathbf{x}(\mathcal{I})$ *and* $\mathbf{x}'(\mathcal{U}) \ge \mathbf{x}(\mathcal{U})$, *we have* $\mathbf{x}' \in \mathcal{S}_\mathcal{I}$.

Consider a fixed $\mathbf{x}_T \in \mathcal{S}_\mathcal{I}$ with $\mathcal{I} \subset [m]$. Let $\mathcal{Z} := \{\mathbf{z} \in \mathbb{R}^{|\mathcal{I}|} : \exists \boldsymbol{\lambda} \in \mathcal{D},\ \boldsymbol{\rho} = \mathbf{A}\boldsymbol{\lambda},\ \mathbf{z} = \boldsymbol{\rho}(\mathcal{I})\}$. For every $\mathbf{z} \in \mathcal{Z}$, define $R(\mathbf{z})$ as the value of partial optimization of $r(\boldsymbol{\lambda})$ by fixing the resource levels in the set $\mathcal{I}$:

$$R(\mathbf{z}) := \max_{\boldsymbol{\lambda}\in\mathcal{D},\, \boldsymbol{\rho}\in\mathbb{R}_+^m} \{r(\boldsymbol{\lambda}),\ \text{s.t. } \mathbf{A}\boldsymbol{\lambda} = \boldsymbol{\rho},\ \boldsymbol{\rho}(\mathcal{I}) = \mathbf{z}\}. \tag{37}$$

The motivation for Equation (37) is that, in the region $\mathcal{S}_\mathcal{I}$, the solution of the resolving heuristic is only affected by the inventory levels of the *constrained resources* in $\mathcal{I}$. The following lemma establishes some useful properties of the function R. The proof can be found in Appendix A.

Lemma 4. *The function R has the following properties:*

1. *For any* $\mathbf{x}_t \in \mathcal{S}_{\mathcal{I}}$ *and* $\mathbf{z}_t = \mathbf{x}_t(\mathcal{I})$, *it holds that* $R(\mathbf{z}_t) = r(\boldsymbol{\lambda}_t)$, *where* $\boldsymbol{\lambda}_t$ *is the solution to* $\max_{\boldsymbol{\lambda}} r(\boldsymbol{\lambda})$ *subject to* $\mathbf{0} \le \mathbf{A}\boldsymbol{\lambda} \le \mathbf{x}_t$.
2. *It holds that* $R(\mathbf{z})$ *is strictly concave and three times continuously differentiable for all* $\mathbf{z} \in \mathcal{Z}$ *and satisfies* $\nabla^2 R(\mathbf{z}) \preccurlyeq -m\mathbf{I}$, $\|\nabla^3 R(\mathbf{z})\|_2 \le M$ *with some constants* $m > 0, M > 0$.
3. *Given* $\mathbf{z}, \mathbf{z}' \in \mathcal{Z}$, *let* $\boldsymbol{\lambda}, \boldsymbol{\lambda}'$ *be part of the optimal solutions in* (37). *Then,* $\|\boldsymbol{\lambda} - \boldsymbol{\lambda}'\| \le L_z \|\mathbf{z} - \mathbf{z}'\|$ *for some constant* $L_z > 0$ *uniformly on* $\mathcal{Z}$.

6.2. Stopping Time with Bounded Expectation

Recall that $\{\boldsymbol{\xi}^r_\tau\}_{\tau=1}^T = \{\mathbf{A}\boldsymbol{\varepsilon}^r_\tau\}_{\tau=1}^T$ are the stochastic noise vectors at each time period on the path of the resolving heuristic policy π^r. For any τ, define $\overline{\boldsymbol{\xi}}^r_{\to\tau} := \frac{\boldsymbol{\xi}^r_T}{T-1} + \ldots + \frac{\boldsymbol{\xi}^r_{\tau+1}}{\tau}$. Let $r_0 := \min_{k\in\mathcal{I}} \partial R(\mathbf{x}_T(\mathcal{I}))/\partial x_T(k)$. Note that $r_0 > 0$ by Assumption C1. Define $T^\sharp$ as

$$T^\sharp := \max\left\{\tau \ge 1 : \|\overline{\boldsymbol{\xi}}^r_{\to\tau-1}\| > \min\left(\delta_0, \frac{r_0}{2\|\nabla^2 R(\mathbf{x}_T(\mathcal{I}))\|_2}\right)\right\} \vee \left\lceil \frac{2\|\nabla^2 R(\mathbf{x}_T(\mathcal{I}))\|_2 B_\xi}{r_0} + 1\right\rceil, \tag{38}$$

where $\delta_0 > 0$ is the constant parameter in Assumption C1. Because $\overline{\boldsymbol{\xi}}^r_{\to t-1}$ is measurable with respect to F_t, the event $\{T^\sharp = t\}$ is adaptive to the filtration $\{\mathcal{F}_t\}$, and therefore, $T^\sharp$ is a stopping time. We note that this definition implies the inequality $\nabla R(\mathbf{x}_T(\mathcal{I})) - \nabla^2 R(\mathbf{x}_T(\mathcal{I}))\,\overline{\boldsymbol{\xi}}^r_{\to T^\sharp-1}(\mathcal{I}) \ge 0$. The definition of $T^\sharp$ also ensures that lost sales are zero for all $\tau > T^\sharp$ (see Lemma 5).

Recall that $\mathbf{x}^r_\tau$ is the normalized inventory vector of the m resources when τ time periods are remaining. The following lemma gives a characterization of $\mathbf{x}^r_\tau$ in terms of $\overline{\boldsymbol{\xi}}^r_{\to\tau}$. It also gives an upper bound on $\mathbb{E}[T^\sharp]$, similar to Lemma 1.

Lemma 5. *Under Assumptions* B1–B5 *and* C1, *the following holds for all* $\tau \ge T^\sharp$:

$$\mathbf{x}^r_\tau(\mathcal{I}) = \mathbf{x}^r_T(\mathcal{I}) - \overline{\boldsymbol{\xi}}^r_{\to\tau}(\mathcal{I});\ \mathbf{x}^r_\tau(\mathcal{U}) \ge \mathbf{x}^r_T(\mathcal{U}) - \overline{\boldsymbol{\xi}}^r_{\to\tau}(\mathcal{U}), \tag{39}$$

where $\mathcal{I} = \mathcal{I}(\mathbf{x}_T), \mathcal{U} = \mathcal{U}(\mathbf{x}_T)$ *as defined in Assumption* C1. *For* $\tau \ge T^\sharp$, *we have* $\mathbf{x}^r_\tau \in \mathcal{S}_{\mathcal{I}}$; *for , we have* $\mathbf{l}^r_\tau = \mathbf{0}$. *Furthermore, the stopping time is bounded by* $\mathbb{E}[T^\sharp] = O(1)$.

6.3. Complete Proof of Theorem 5

Given the initial inventory level $\mathbf{x}_T \in \mathcal{S}_{\mathcal{I}}$, we first upper bound the value function of the optimal DP policy by considering a relaxed problem in which the inventory of the products in $\mathcal{I}$ is $\mathbf{x}(\mathcal{I})$, but the inventory of the products in $\mathcal{U}$ is unbounded. Let $\{\mathbf{x}^*_\tau\}_{\tau=1}^T$ be the path of inventory level process for this relaxed problem. Define $\mathbf{z}^*_\tau := \mathbf{x}^*_\tau(\mathcal{I})$. Similar to Equation (35), we define $\Phi^*_\tau(\mathbf{z}^*_\tau)$ as the value function of this relaxed problem. Clearly, we have $\phi^*_T(\mathbf{x}_T) \le \Phi^*_T(\mathbf{z}_T)$ as the latter has fewer inventory constraints.

With a slight abuse of notation, we denote $\Delta_\tau(\mathcal{I}), \boldsymbol{\xi}^*_\tau(\mathcal{I}), \boldsymbol{\xi}^r_\tau(\mathcal{I}), \boldsymbol{\ell}^*_\tau(\mathcal{I})$ simply by $\Delta_\tau, \boldsymbol{\xi}^*_\tau, \boldsymbol{\xi}^r_\tau, \boldsymbol{\ell}^*_\tau$. Then, it holds that $\mathbf{z}^*_\tau = \mathbf{z}_T - \overline{\Delta}_{\to\tau} - \overline{\boldsymbol{\xi}}^*_{\to\tau} + \overline{\boldsymbol{\ell}}^*_{\to\tau}$, where

$$\overline{\Delta}_{\to\tau} = \frac{\Delta_T}{T-1} + \ldots + \frac{\Delta_{\tau+1}}{\tau},\ \overline{\boldsymbol{\xi}}^*_{\to\tau} = \frac{\boldsymbol{\xi}^*_T}{T-1} + \ldots + \frac{\boldsymbol{\xi}^*_{\tau+1}}{\tau},$$

and

$$\overline{\boldsymbol{\ell}}^*_{\to\tau} = \frac{\boldsymbol{\ell}^*_T}{T-1} + \ldots + \frac{\boldsymbol{\ell}^*_{\tau+1}}{\tau}.$$

By adapting Equation (11) to the multiproduct setting, we have

$$\phi^*_T(\mathbf{x}_T) \le \Phi^*_T(\mathbf{z}_T) \le \mathbb{E}\left[\sum_{\tau=T^\sharp}^T R(\mathbf{z}^*_\tau + \Delta_\tau) + (T^\sharp - 1)R(\mathbf{z}'_{T^\sharp-1})\right], \tag{40}$$

where

$$\mathbf{z}'_{T^\sharp-1} := \arg\max_{\mathbf{z}\in\mathcal{Z}} \{R(\mathbf{z}) \mid \mathbf{z} \le \mathbf{z}^*_{T^\sharp-1}\}.$$

Next, we analyze the inventory process under the resolving heuristic (for the original problem rather than the relaxed problem). Let $\mathbf{z}^r_\tau := \mathbf{x}^r_\tau(\mathcal{I})$. By Lemma 5, it holds that $\mathbf{z}^r_\tau = \mathbf{z}_T - \overline{\boldsymbol{\xi}}^r_{\to\tau}$ for all $\tau \ge T^\sharp$. In addition, for all $\tau \ge T^\sharp$, it holds that $r(\boldsymbol{\lambda}_\tau) = R(\mathbf{z}^r_\tau)$ by Lemma 4. By adapting Equation (12) to the multiproduct setting, we get

$$\begin{aligned}\phi^r_T(\mathbf{x}_T) &\ge \mathbb{E}\left[\sum_{\tau=T^\sharp}^T r(\boldsymbol{\lambda}_\tau)\right] - \sup_{d\in\mathcal{D}} r(d)\\ &= \mathbb{E}\left[\sum_{\tau=T^\sharp}^T R(\mathbf{z}^r_\tau)\right] - \sup_{d\in\mathcal{D}} r(d).\end{aligned} \tag{41}$$

The outline for the rest of the proof is similar to that for the single-product case in Section 4.2. The proof is again divided into three main steps. In the first step (from Equations (42)–(44)), we use Taylor expansion to bound the expected revenue difference between the optimal DP policy and the resolving policy. In the second step (from Equations (45)–(50)), the regret expression is split into five terms, and each term is simplified separately. In the third step (from Equations (51)–(53)), the five terms are combined, and the regret is eventually reduced to a function of $\mathbb{E}[T^\sharp]$, which is $O(1)$ by Lemma 5.

Generalizing the arguments from Equations (28) and (29) and using the second property of Lemma 4, we obtain

$$\begin{aligned}&\mathbb{E}\left[\sum_{\tau=T^\sharp}^T \left(R(\mathbf{z}^*_\tau + \Delta_\tau) - R(\mathbf{z}^r_\tau)\right)\right]\\ &\le \mathbb{E}\Bigg[\sum_{\tau=T^\sharp}^T (\langle \nabla R(\mathbf{0}), \overline{\boldsymbol{\ell}}^*_{\to\tau}\rangle + \langle \nabla R(\mathbf{z}_T), \Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\boldsymbol{\xi}}^\delta_{\to\tau} - \boldsymbol{\xi}^\delta_\tau\rangle\\ &\quad -(\overline{\boldsymbol{\xi}}^r_{\to\tau})^{\mathrm{T}} \nabla^2 R(\mathbf{z}_T)[\Delta_\tau - \overline{\Delta}_{\to\tau}] + (\overline{\boldsymbol{\xi}}^r_{\to\tau})^{\mathrm{T}} \nabla^2 R(z_T)[\boldsymbol{\xi}^\delta_\tau - \overline{\boldsymbol{\xi}}^\delta_{\to\tau}]\\ &\quad + \frac{M}{2}\|\overline{\boldsymbol{\xi}}^r_{\to\tau}\|^2 \|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\boldsymbol{\xi}}^\delta_{\to\tau}\| - \frac{m}{2}\|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\boldsymbol{\xi}}^\delta_{\to\tau}\|_2^2)\Bigg],\end{aligned} \tag{42}$$

where $\boldsymbol{\xi}^{\delta}_{\tau} := \boldsymbol{\xi}^{r}_{\tau} - \boldsymbol{\xi}^{*}_{\tau}$ and $\overline{\boldsymbol{\xi}}^{\delta}_{\to\tau} := \frac{\boldsymbol{\xi}^{\delta}_{T}}{T-1} + \ldots + \frac{\boldsymbol{\xi}^{\delta}_{\tau+1}}{\tau}$.

Define $\boldsymbol{\eta} = -(\mathbf{z}^{*}_{T^\sharp-1} - \mathbf{z}_{T^\sharp-1}{}')^{+}$, where $(\mathbf{z})^{+}$ denotes the element-wise positive part of a vector $\mathbf{z}$. Note that $\mathbf{z}_{T^\sharp-1}{}' = \mathbf{z}^{*}_{T^\sharp-1} + \eta$ because $\mathbf{z}_{T^\sharp-1}{}' \le \mathbf{z}^{*}_{T^\sharp-1}$. Equation (15) can be generalized to

$$\begin{aligned} R(\mathbf{z}'_{T^\sharp-1}) - R(\mathbf{z}^{r}_{T^\sharp-1}) &= R(\mathbf{z}^{*}_{T^\sharp-1} + \eta) - R(\mathbf{z}^{r}_{T^\sharp-1}) \\ &\le \langle \nabla R(\mathbf{0}), \overline{\boldsymbol{\ell}}^{*}_{\to T^\sharp-1}\rangle + \langle \nabla R(\mathbf{z}_T), \boldsymbol{\eta} - \overline{\Delta}_{\to T^\sharp-1} + \overline{\boldsymbol{\xi}}^{\delta}_{\to T^\sharp-1}\rangle \\ &\quad - (\overline{\boldsymbol{\xi}}^{r}_{\to T^\sharp-1})^{\mathrm{T}} \nabla^2 R(\mathbf{z}_T)[\boldsymbol{\eta} - \overline{\Delta}_{\to T^\sharp-1} + \overline{\boldsymbol{\xi}}^{\delta}_{\to T^\sharp-1}] \\ &\quad + \frac{M^2}{8m} \|\overline{\boldsymbol{\xi}}^{r}_{\to T^\sharp-1}\|^4 . \end{aligned} \tag{43}$$

Let $\circ$ denote the element-wise product. Subtracting Equation (41) from Equation (40) and then combining Equation (43) with Equation (42), we obtain

$$\begin{aligned} &\phi^{*}_{T}(\mathbf{x}_T) - \phi^{r}_{T}(\mathbf{x}_T) \\ &\quad \le \mathbb{E}\big[\langle \nabla R(\mathbf{z}_T), \mathbf{E}\rangle - \nabla^2 R(\mathbf{Z}_T) \circ \mathbf{B} + \nabla^2 R(\mathbf{Z}_T) \circ \mathbf{C} + \mathbf{D} \\ &\qquad + \langle \nabla R(\mathbf{0}), \mathbf{F}\rangle\big] + O(\mathbb{E}[T^\sharp]), \end{aligned} \tag{44}$$

where

$$\begin{aligned} \mathbf{E} &= \sum_{\tau=T^\sharp}^{T} [\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\boldsymbol{\xi}}^{\delta}_{\to\tau} - \boldsymbol{\xi}^{\delta}_{\tau}] \\ &\quad - (T^\sharp - 1)(\overline{\Delta}_{\to T^\sharp-1} - \overline{\boldsymbol{\xi}}^{\delta}_{\to T^\sharp-1}) + (T^\sharp - 1)\boldsymbol{\eta}, \\ \mathbf{B} &= \sum_{\tau=T^\sharp}^{T} \overline{\boldsymbol{\xi}}^{r}_{\to\tau}[\Delta_\tau - \overline{\Delta}_{\to\tau}]^{\mathrm{T}} - (T^\sharp - 1)\overline{\boldsymbol{\xi}}^{r}_{\to T^\sharp-1}\overline{\Delta}^{\mathrm{T}}_{\to T^\sharp-1} \\ &\quad + (T^\sharp - 1)\overline{\boldsymbol{\xi}}^{r}_{\to T^\sharp-1}\boldsymbol{\eta}^{\mathrm{T}}, \\ \mathbf{C} &= \sum_{\tau=T^\sharp}^{T} \overline{\boldsymbol{\xi}}^{r}_{\to\tau}[\boldsymbol{\xi}^{\delta}_{\tau} - \overline{\boldsymbol{\xi}}^{\delta}_{\to\tau}]^{\mathrm{T}} - (T^\sharp - 1)\overline{\boldsymbol{\xi}}^{r}_{\to T^\sharp-1}[\overline{\boldsymbol{\xi}}^{\delta}_{\to T^\sharp-1}]^{\mathrm{T}}, \\ \mathbf{D} &= \sum_{\tau=T^\sharp}^{T} \bigg(\frac{M}{2}\|\overline{\boldsymbol{\xi}}^{r}_{\to\tau}\|^2 \|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\boldsymbol{\xi}}^{\delta}_{\to\tau}\| - \frac{m}{2} \\ &\quad \|\Delta_\tau - \overline{\Delta}_{\to\tau} + \overline{\boldsymbol{\xi}}^{\delta}_{\to\tau}\|^2\bigg) + (T^\sharp - 1)\frac{M^2}{8m}\|\overline{\boldsymbol{\xi}}^{r}_{\to T^\sharp-1}\|^4, \\ \mathbf{F} &= \sum_{\tau=T^\sharp}^{T} \overline{\boldsymbol{\ell}}^{*}_{\to\tau} + (T^\sharp - 1)\overline{\boldsymbol{\ell}}^{*}_{\to T^\sharp-1}. \end{aligned}$$

Recall that in the definitions of $\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}$, all vectors $\Delta_\tau, \overline{\Delta}_{\to\tau}, \boldsymbol{\xi}^{\delta}_{\tau}, \overline{\boldsymbol{\xi}}^{\delta}_{\to\tau}, \overline{\boldsymbol{\xi}}^{r}_{\to\tau}, \boldsymbol{\eta}, \overline{\boldsymbol{\ell}}^{*}_{\to\tau}$ are restricted to coordinates in $\mathcal{I}$. $\mathbf{B}, \mathbf{C}$ are $|\mathcal{I}| \times |\mathcal{I}|$-dimensional matrices, $\mathbf{D}$ is a scalar, and $\mathbf{E}, \mathbf{F}$ are $|\mathcal{I}|$-dimensional vectors. Using the same calculations as in Section 4, these quantities can be reduced to

$$\mathbb{E}[\mathbf{E}] = \mathbb{E}[(T^\sharp - 1)\boldsymbol{\eta}], \tag{45}$$

$$\mathbb{E}[\mathbf{B}] = \mathbb{E}[(T^\sharp - 1)\overline{\boldsymbol{\xi}}^{r}_{\to T^\sharp-1}\boldsymbol{\eta}^{\mathrm{T}}], \tag{46}$$

$$\mathbb{E}[\mathbf{C}] = \mathbb{E}\left[-\sum_{t=T^\sharp}^{T} \frac{\boldsymbol{\xi}^{r}_{t}[\boldsymbol{\xi}^{\delta}_{t}]^{\mathrm{T}}}{t-1}\right], \tag{47}$$

$$\|\mathbf{F}\| \le \sup_{d\in\mathcal{D}} \|d\| + B_\xi. \tag{48}$$

Generalizing Equation (19), it holds that

$$\begin{aligned} &\left\|\mathbb{E}\left[-\sum_{t=T^\sharp}^{T} \frac{\boldsymbol{\xi}^{r}_{t}[\boldsymbol{\xi}^{\delta}_{t}]^{\mathrm{T}}}{t-1}\right]\right\|_2 \\ &\quad \le B_\xi \mathbb{E}\left[\sum_{t=2}^{T} \boldsymbol{I}\{T^\sharp \le t\} \frac{\sqrt{\mathbb{E}[\|\boldsymbol{\xi}^{r}_{t} - \boldsymbol{\xi}^{*}_{t}\|^2 | \mathcal{F}_{t+1}]}}{t-1}\right] \\ &\quad \le B_\xi \mathbb{E}\left[\sum_{t=2}^{T} \boldsymbol{I}\{T^\sharp \le t\} \frac{\sigma_{\max}(\mathbf{A})\sqrt{\mathbb{E}[\|\boldsymbol{\varepsilon}^{r}_{t} - \boldsymbol{\varepsilon}^{*}_{t}\|^2 | \mathcal{F}_{t+1}]}}{t-1}\right] \\ &\quad \le B_\xi \mathbb{E}\left[\sum_{t=2}^{T} \boldsymbol{I}\{T^\sharp \le t\} \frac{\sigma_{\max}(\mathbf{A}) L \|\boldsymbol{\lambda}^{*}_{t} - \boldsymbol{\lambda}_t\|}{t-1}\right] \\ &\quad \le B_\xi \mathbb{E}\left[\sum_{t=2}^{T} \boldsymbol{I}\{T^\sharp \le t\} \frac{L' \|\mathbf{z}^{*}_{t} + \Delta_t - \mathbf{z}^{r}_{t}\|}{t-1}\right], \end{aligned} \tag{49}$$

where the second inequality holds because $\|\boldsymbol{\xi}^{r}_{t} - \boldsymbol{\xi}^{*}_{t}\| \le \sigma_{\max}(\mathbf{A}) \|\boldsymbol{\varepsilon}^{r}_{t} - \boldsymbol{\varepsilon}^{*}_{t}\|$, where $\sigma_{\max}(\mathbf{A})$ is the largest singular value of $\mathbf{A}$, the third inequality uses the bound on the Wasserstein distance in Assumption B5, and the fourth inequality uses the third property in Lemma 4 with $L' = L_z L \sigma_{\max}(\mathbf{A})$. As a result, $\|\mathbb{E}[\mathbf{C}]\|_2$ can be upper bounded as

$$\|\mathbb{E}[\mathbf{C}]\|_2 \le L' B_\xi \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T} \frac{\|\Delta - \overline{\Delta}_{\to\tau} + \overline{\boldsymbol{\xi}}^{\delta}_{\to\tau}\|}{\tau - 1}\right]. \tag{50}$$

Because $\nabla R(\mathbf{x}_T(\mathcal{I})) - \nabla^2 R(\mathbf{x}_T(\mathcal{I}))\overline{\boldsymbol{\xi}}^{r}_{\to T^\sharp} \ge 0$ by the definition of $T^\sharp$, combining Equation (44) with Equations (45), (46), and (50) and using the same derivation that leads to Equation (22), we have

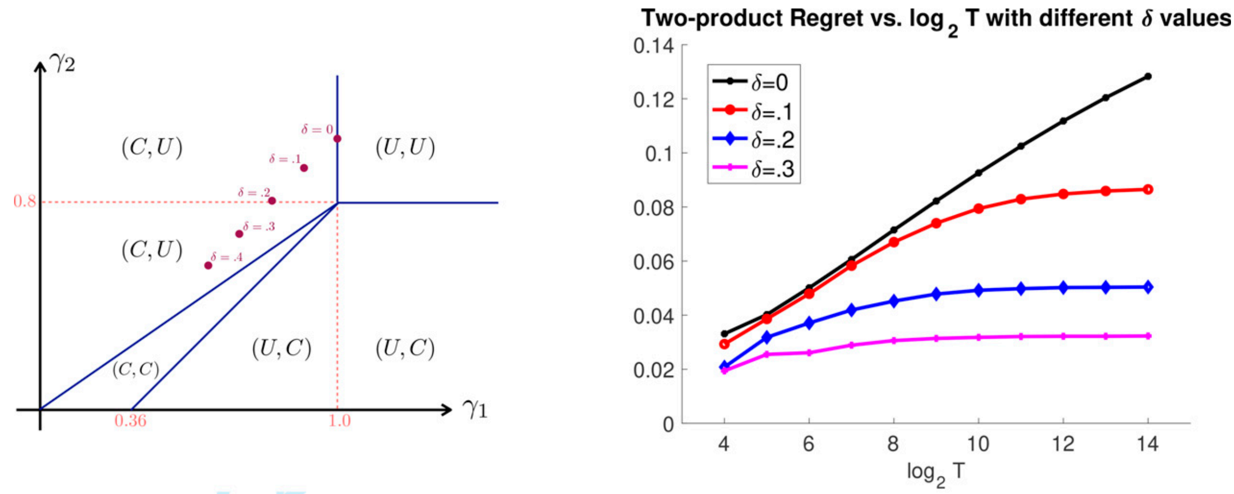
$$\begin{aligned} &\mathbb{E}\big[\langle \nabla R(\mathbf{z}_T), \mathbf{E}\rangle - \nabla^2 R(\mathbf{Z}_T) \circ \mathbf{B} + \nabla^2 R(\mathbf{Z}_T) \circ \mathbf{C} + \mathbf{D} \\ &\quad + \langle \nabla R(\mathbf{0}), \mathbf{F}\rangle\big] \\ &\le \mathbb{E}\bigg[(T^\sharp - 1)\frac{M^2}{8m}\|\overline{\boldsymbol{\xi}}^{r}_{\to T^\sharp-1}\|^4 + \sum_{\tau=T^\sharp}^{T} \frac{1}{m}\bigg(\bigg(\|\nabla^2 R(\mathbf{z}_T)\|_2 \frac{L' B_\xi}{\tau - 1}\bigg)^2 \\ &\quad + \frac{M^2}{4}\|\overline{\boldsymbol{\xi}}^{r}_{\to\tau}\|^4\bigg)\bigg] \\ &\quad + \|\nabla R(\mathbf{0})\| \bigg(\sup_{d\in\mathcal{D}} \|d\| + B_\xi\bigg). \end{aligned} \tag{51}$$

To complete the proof, we upper bound each term in Equation (51). First, it is easy to verify that

$$\begin{aligned} \sum_{\tau=T^\sharp}^{T} \bigg(\frac{\|\nabla^2 R(\mathbf{z}_T)\|_2 L' B_\xi}{\tau - 1}\bigg)^2 &\le (\|\nabla^2 R(\mathbf{z}_T)\|_2 L' B_\xi)^2 \sum_{j=1}^{T-1} \frac{1}{j^2} \\ &\le 2(\|\nabla^2 R(\mathbf{z}_T)\|_2 L' B_\xi)^2 \quad a.s. \end{aligned} \tag{52}$$

We next focus on the terms involving $\|\overline{\boldsymbol{\xi}}^{r}_{\to\tau}\|^4$. Note that $\{\|\overline{\boldsymbol{\xi}}^{r}_{\to\tau-1}\|^4\}$ is a submartingale adapted to the filtration $\{\mathcal{F}_\tau\}^{T}_{\tau=1}$. Let $S_t = \sum^{T}_{\tau=t}(t-1)(\|\overline{\boldsymbol{\xi}}^{r}_{\to\tau-1}\|^4 -$

Figure 4. (Color online) Inventory State Partition and Numerical Results for a 2-Product, 2-Resource Setting



Notes. Left: Partition of inventory states according to the resources being constrained "C" or unconstrained "U." Right: Regret versus $\log_2 T$ for the two-product, two-resource example with initial inventory levels being $(1-\delta)T$ for both resources.

$\|\bar{\boldsymbol{\xi}}^r_{\to\tau}\|^4)$; then, $\{S_t\}$ is also a submartingale. Because $T^\sharp$ is a stopping time, we have

$$\mathbb{E}\left[(T^\sharp-1)\,\|\bar{\boldsymbol{\xi}}^r_{\to T^\sharp-1}\|^4+\sum_{\tau=T^\sharp}^{T}\|\bar{\boldsymbol{\xi}}^r_{\to\tau}\|^4\right]=\mathbb{E}[S_{T^\sharp}]\le\mathbb{E}[S_2]$$
$$=\mathbb{E}\left[\sum_{\tau=1}^{T}\|\bar{\boldsymbol{\xi}}^r_{\to\tau}\|^4\right].$$

It is easy to verify that

$$\mathbb{E}[\|\bar{\boldsymbol{\xi}}^r_{\to t}\|^4]\le\sum_{j,k>t}\frac{\mathbb{E}[\|\boldsymbol{\xi}^r_j\|^2\|\boldsymbol{\xi}^r_k\|^2]}{(j-1)^2(k-1)^2}\le B_\xi^4\left(\sum_{j>t}\frac{1}{(j-1)^2}\right)^2\le\frac{4B_\xi^4}{t^2}.$$

Subsequently,

$$\mathbb{E}\left[\frac{M^2}{8m}(T^\sharp-1)\,\|\bar{\boldsymbol{\xi}}^r_{\to T^\sharp-1}\|^4+\frac{M^2}{4m}\sum_{\tau=T^\sharp}^{T}\|\bar{\boldsymbol{\xi}}^r_{\to\tau}\|^4\right]$$
$$\le\frac{M^2}{4m}\sum_{\tau=1}^{\mathbb{T}}\frac{4B_\xi^4}{\tau^2}\le\frac{2M^2B_\xi^4}{m}. \tag{53}$$

Finally, combining Equations (42), (44), (51), and (53) and using Lemma 5, we obtain

$$\phi_T^*(\mathbf{x}_T)-\phi_T^r(\mathbf{x}_T)\le\frac{2M^2B_\xi^4}{m}+\frac{2(\|\nabla^2R(\mathbf{z}_T)\|_2L'B_\xi)^2}{m}$$
$$+\|\nabla R(\mathbf{0})\|\left(\sup_{d\in\mathcal{D}}\|d\|+B_\xi\right)+O(\mathbb{E}[T^\sharp])=O(1),$$

which completes the proof of Theorem 5.

6.4. Numerical Results

We provide a numerical example with two products and two resources ($n=m=2$). We assume that product 1 requires two units of the first resource and one unit of the second resource, whereas product 2 requires one unit of the first resource and two units of the second resource. Both resources have initial inventory levels of $(1-\delta)T$, where T is the total number of selling periods and $\delta\in\{0,0.1,0.2,0.3\}$ is a parameter that varies in the experiment. At each time t, the seller could choose a demand rate vector $x=(x_1,x_2)\in[0,1]^2$, and the corresponding prices are given by $p_1(x_1)=0.4-0.5x_1$, $p_2(x_2)=0.2-0.5x_2$.

In the left panel of Figure 4, we show the partition of inventory states based on which resource constraints are active. More specifically, each point (γ_1,γ_2) on the plane corresponds to a problem instance for which the initial inventory level of the first resource is γ_1T and the initial inventory level of the second resource is γ_2T. The plane is partitioned into four regions. The region (U,U) means that inventory constraints of neither resource are active at the beginning of the first selling period. The region (C,U) means that the inventory constraint for the first resource is active but not the constraint for the second resource and vice versa for (U,C). The region (C,C) means that both inventory constraints are active. It is easy to verify that, with $\delta=0$, the initial normalized inventory $(\gamma_1,\gamma_2)=(1,1)$ lies on the boundary between the (U,U) and the (C,U) regions, which violates Assumption C1. For $0<\delta<1$, the initial normalized inventory $(1-\delta,1-\delta)$ belongs to the interior of the (C,U) region and, thus, satisfies Assumption C1.

In the right panel of Figure 4, we report the regret of this two-product, two-resource example under different initial inventory levels. We observe that, when $\delta=0$ (Assumption C1 does not hold), the cumulative regret of the resolving heuristics is unbounded and seems to scale logarithmically with time horizon T.

On the other hand, with $\delta \in \{0.1, 0.2, 0.3\}$ (Assumption C1 holds), the regret is bounded for sufficiently large time horizons. This numerical result validates our theoretical findings and is also qualitatively similar to our numerical results for the single-product, single-resource example in Section 5.

7. Conclusion

In this paper, we analyze a simple resolving heuristic for the classic price-based revenue management problem with either a single or multiple products. The heuristic resolves the fluid model in each period to reset the prices.

We establish two complementary theoretical results. First, the resolving heuristic attains $O(1)$ regret compared against the value of the optimal policy. The $O(1)$ regret depends on the shape of the demand function as well as how close the initial inventory level is to certain "boundaries." Going forward, an open question is whether the boundary condition can be removed. Our numerical experiment shows that the resolving heuristic considered in this paper may not have $O(1)$ regret when the initial inventory is on the boundary, so the pricing algorithm needs to be modified for the boundary case.

Second, we show that there exists an $\Omega(\ln T)$ gap between the value of the optimal policy and the value of the fluid model. For that reason, our regret analysis does not use the fluid model as a benchmark; the proof directly compares the value of the optimal policy and that of the heuristic. An interesting future direction is to find a different benchmark that is within $O(1)$ of the optimal value, which may help simplify our proof.

Appendix A. Additional Proofs

Proof of Proposition 1. For any $1 \le t \le T$, let $x_t = y_t/t$ be the normalized inventory level in period t. The static pricing policy commits to the price $p_t = f^{-1}(\lambda^u)$ in every period. If $x_t > \lambda^u$, the resolving heuristic also selects the price $p_t = f^{-1}(\lambda^u)$. Let π denote either the static pricing policy π^s or the resolving heuristic π^r. Let $\mathcal{E}_t = \{\sup_{t+1\le t'\le T}\sum_{\tau=t'}^{T}\xi_\tau \le T(x_T - \lambda^u)\}$ be the event that the inventory level never drops below λ_u from the start to the beginning of period t. Let d_t and $s_t = \max\{d_t, y_t\}$ be the realized and satisfied demand at period t, respectively. We have

$$R^\pi(T, x_T) \ge \mathbb{E}\left[\sum_{t=1}^{T} p_t s_t \mathbf{1}\{\mathcal{E}_t\}\right] = \mathbb{E}\left[\sum_{t=1}^{T} p_t d_t \mathbf{1}\{\mathcal{E}_t\}\right] - \mathbb{E}\left[\sum_{t=1}^{T} p_t (d_t - s_t)\mathbf{1}\{\mathcal{E}_t\}\right]. \tag{A.1}$$

Because $d_t = \lambda^u + \xi_t$ under event $\mathcal{E}_t \in \mathcal{F}_{t+1}$ and $E[\xi_t \mid \mathcal{F}_{t+1}] = 0$ by Assumption A4, the first term in Equation (A.1) is equal to

$$\mathbb{E}\left[\sum_{t=1}^{T} p_t d_t \mathbf{1}\{\mathcal{E}_t\}\right] = \mathbb{E}\left[\sum_{t=1}^{T} f^{-1}(\lambda^u)\mathbf{1}\{\mathcal{E}_t\}\mathbb{E}[d_t \mid \mathcal{F}_{t+1}]\right] = \mathbb{E}\left[\sum_{t=1}^{T} f^{-1}(\lambda^u)\mathbf{1}\{\mathcal{E}_t\}\lambda^u\right] = \mathbb{E}\left[\sum_{t=1}^{T} r^u \mathbf{1}\{\mathcal{E}_t\}\right].$$

The second term in Equation (A.1) measures the amount of revenue loss resulting from partial demand fulfillment. The event of partial demand fulfillment can happen at most once during the entire horizon because the inventory level is zero afterward. By the bound on demand noises ξ_t in Assumption A4, we have

$$\mathbb{E}\left[\sum_{t=1}^{T} p_t (d_t - s_t)\mathbf{1}\{\mathcal{E}_t\}\right] \le f^{-1}(\lambda^u)(\lambda^u + B_\xi).$$

Therefore,

$$R^\pi(T, x_T) \ge \mathbb{E}\left[\sum_{t=1}^{T} r^u \mathbf{1}\{\mathcal{E}_t\}\right] - f^{-1}(\lambda^u)(\lambda^u + B_\xi) = r^u\left(T - \sum_{t=1}^{T}\mathbb{P}[\mathcal{E}_t^c]\right) - f^{-1}(\lambda^u)(\lambda^u + B_\xi).$$

By Doob's martingale inequality, for any $1 \le t \le T$, we have

$$\mathbb{P}[\mathcal{E}_t^c] = \mathbb{P}\left[\sup_{t+1\le t'\le T}\sum_{\tau=t'}^{T}\xi_\tau > T(x_T - \lambda^u)\right] \le \frac{\sum_{\tau=t+1}^{T}\mathbb{E}[\xi_\tau^2]}{T^2(x_T - \lambda^u)^2} \le \frac{(T-t)B_\xi^2}{T^2(x_T - \lambda^u)^2} \le \frac{B_\xi^2}{T(x_T - \lambda^u)^2}.$$

Thus, $R^\pi(T, x_T) \ge r^u[T - (B_\xi^2)/(x_T - \lambda^u)^2] - f^{-1}(\lambda^u)(\lambda^u + B_\xi) = Tr^u - O(1)$. The proof is complete by noting that $Tr^u \ge R^{\pi^*}(T, x_T)$ using Theorem 1. □

Proof of Lemma 1. Let $\gamma := \min\{x_T - \underline{d}, \lambda^u - x_T, r'(x_T)/|2r''(x_T)|\}$. By the definition from Equation (8), for $\tau > 2|r''(x_T)|B_\xi/r'(x_T) + 1$, we have

$$\mathbb{P}[T^\sharp \ge \tau] = \mathbb{P}\left[\sup_{\tau-1\le t\le T}|\overline{\xi}^r_{\to t}| > \gamma\right] \le \frac{\mathbb{E}[|\overline{\xi}^r_{\to\tau-1}|^4]}{\gamma^4},$$

which follows from Doob's martingale inequality by noting that $\{\xi_\tau/\tau\}$ is a martingale difference sequence (see Assumption A4). Recall that $\overline{\xi}^r_{\to t} = \frac{\xi_T^r}{T-1} + \ldots + \frac{\xi_{t+1}^r}{t}$ with $|\xi_t^r| \le B_\xi$ *a.s.* For $\tau > 2|r''(x_T)|B_\xi/r'(x_T) + 1$, it is easily verified that

$$\mathbb{E}[|\overline{\xi}^r_{\to\tau-1}|^4] = \sum_{\tau\le j,k\le T}\frac{\mathbb{E}[|\xi_j^r|^2|\xi_k^r|^2]}{(j-1)^2(k-1)^2} \le \left(\sum_{\tau\le j\le T}\frac{B_\xi^2}{(j-1)^2}\right)\left(\sum_{\tau\le k\le T}\frac{B_\xi^2}{(k-1)^2}\right) \le \frac{4B_\xi^4}{(\tau-1)^2}.$$

Using the equality $\mathbb{E}[T^\sharp] = \sum_{\tau=1}^{T}\mathbb{P}[T^\sharp \ge \tau]$, we have

$$\mathbb{E}[T^\sharp] \le \left\lceil \frac{2|r''(x_T)|B_\xi}{r'(x_T)} \right\rceil + 1 + \sum_{\tau=3}^{T}\frac{4B_\xi^4}{\gamma^4(\tau-1)^2} \le \frac{2|r''(x_T)|B_\xi}{r'(x_T)} + 2 + \frac{4B_\xi^4}{\gamma^4}. \quad \square$$

Proof of Lemma 2. For the optimal DP policy, by Equation (5), the revenue collected in period τ is $r(x_\tau^* + \Delta_\tau) +$

$f^{-1}(x_\tau^* + \Delta_\tau)(\xi_\tau^* - \ell_\tau^*)$. At the beginning of period $T^\sharp - 1$, the remaining inventory level is $x_{T^\sharp-1}^*(T^\sharp - 1)$. Therefore,

$$\begin{aligned}\phi_T^*(x_T) &= \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}[r(x_\tau^*+\Delta_\tau)+f^{-1}(x_\tau^*+\Delta_\tau)(\xi_\tau^*-\ell_\tau^*)]+\phi_{T^\sharp-1}^*(x_{T^\sharp-1}^*)\right]\\ &\le \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}[r(x_\tau^*+\Delta_\tau)+f^{-1}(x_\tau^*+\Delta_\tau)\xi_\tau^*]+\phi_{T^\sharp-1}^*(x_{T^\sharp-1}^*)\right]\\ &= \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}r(x_\tau^*+\Delta_\tau)+\phi_{T^\sharp-1}^*(x_{T^\sharp-1}^*)\right],\end{aligned}$$

where the second equality holds because $\mathbb{E}[\sum_{\tau=T^\sharp}^T f^{-1}(x_\tau^* + \Delta_\tau)\xi_\tau^*] = 0$ by applying Doob's optional stopping theorem. Similarly, for the resolving policy, the cumulative revenue collected in periods $T, T-1, \ldots, T^\sharp$ is $\sum_{\tau=T^\sharp}^T [r(x_\tau^r) + f^{-1}(x_\tau^r)\xi_\tau^r]$, so we have

$$\begin{aligned}\phi_T^r(x_T^r) &\le \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}[r(x_\tau^r)+f^{-1}(x_\tau^r)\xi_\tau^r]+\phi_{T^\sharp-1}^r(x_{T^\sharp-1}^r)\right]\\ &= \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T}r(x_\tau^r)+\phi_{T^\sharp-1}^r(x_{T^\sharp-1}^r)\right].\end{aligned}$$

Recall that the resolving policy does not incur lost sales before period $T^\sharp + 1$ by the definition in Equation (8). Because $r(x) \le r^u$ for any x, we have

$$\phi_T^r(x_T^r) \ge \mathbb{E}\left[\sum_{\tau=T^\sharp+1}^{T} r(x_\tau^r)\right] \ge \mathbb{E}\left[\sum_{\tau=T^\sharp}^{T} r(x_\tau^r)\right] - r^u.$$

To complete the proof, note that, for any admissible policy π, we have $0 \le \phi_{T^\sharp-1}^\pi(x_{T^\sharp-1}) \le (T^\sharp - 1)r(\min\{x_{T^\sharp-1}, \lambda^u\})$ by Theorem 1. □

Proof of Lemma 4. The first property follows immediately from the definition of $R(\mathbf{z})$ in Equation (37) and the definition of $\mathcal{S}_\mathcal{I}$ in Equation (34) as

$$\begin{aligned}R(\mathbf{z}_t) &= \max_{\boldsymbol\lambda\in\mathcal{D}}\{r(\boldsymbol\lambda), \text{ s.t. } \mathbf{A}\boldsymbol\lambda = \boldsymbol\rho,\ \boldsymbol\rho(\mathcal{I}) = \mathbf{x}_t(\mathcal{I}),\ \boldsymbol\rho(\mathcal{U}) \le \mathbf{x}_t(\mathcal{U})\}\\ &= r(\boldsymbol\lambda_t).\end{aligned}$$

To prove the second property, consider the Lagrangian $\mathcal{L}(\boldsymbol\lambda, \boldsymbol\mu, \mathbf{z}) := r(\boldsymbol\lambda) + \boldsymbol\mu^{\mathrm T}(\mathbf{z} - \mathbf{A}_\mathcal{I}\boldsymbol\lambda)$ of (37), where $\mathbf{A}_\mathcal{I}$ consists of the subset of rows of $\mathbf{A}$ corresponding to those resources in $\mathcal{I}$. Because $r(\boldsymbol\lambda)$ is concave, $\mathcal{D}$ is a convex set with nonempty interior by Assumption B1, and the optimization problem (37) satisfies Slater's condition, $\mathcal{L}$ has a saddle point $(\boldsymbol\lambda^*(\mathbf{z}), \boldsymbol\mu^*(\mathbf{z}))$ satisfying

$$\begin{aligned}R(\mathbf{z}) &= \max_{\boldsymbol\lambda\in\mathcal{D}}\min_{\boldsymbol\mu\in\mathbb{R}^{|\mathcal{I}|}} \mathrm{L}(\boldsymbol\lambda,\boldsymbol\mu,\mathbf{z}) = \min_{\boldsymbol\mu\in\mathbb{R}^{|\mathcal{I}|}}\max_{\boldsymbol\lambda\in\mathcal{D}} \mathcal{L}(\boldsymbol\lambda,\boldsymbol\mu,\mathbf{z})\\ &= \mathcal{L}(\boldsymbol\lambda^*(\mathbf{z}),\boldsymbol\mu^*(\mathbf{z}),\mathbf{z}).\end{aligned}$$

Note that the saddle point $(\boldsymbol\lambda^*(\mathbf{z}), \boldsymbol\mu^*(\mathbf{z}))$ is unique because $r(\boldsymbol\lambda)$ is strictly convex by Assumption B2. By the envelop theorem for saddle point problems (Milgrom and Segal 2002, theorem 5), if the saddle point is unique for every $\mathbf{z}$, the function $R(\mathbf{z})$ is differentiable in the interior of $\mathcal{Z}$ with

$$\nabla R(\mathbf{z}) = \boldsymbol\mu^*(\mathbf{z}). \tag{A.2}$$

By the Karush–Kuhn–Tucker conditions of (37), we have $\nabla r(\boldsymbol\lambda^*(\mathbf{z})) = \mathbf{A}_\mathcal{I}^{\mathrm T}\mu^*(\mathbf{z})$ and $\mathbf{A}_\mathcal{I}\boldsymbol\lambda^*(\mathbf{z}) = \mathbf{z}$. Suppose $\mathrm{rank}(\mathbf{A}_\mathcal{I}) = k$. Let $\mathbf{P} \in \mathbb{R}^{(n-k)\times n}$ be a matrix whose rows are linearly independent and orthogonal to the rows of $\mathbf{A}_\mathcal{I}$, so $\mathbf{P}\nabla r(\boldsymbol\lambda^*(\mathbf{z})) = \mathbf{P}\mathbf{A}_\mathcal{I}^{\mathrm T}\mu^*(\mathbf{z}) = 0$. By Assumption B2, the Hessian $\nabla^2 r(\boldsymbol\lambda^*(\mathbf{z}))$ is invertible, so $\mathbf{P}\nabla^2 r(\boldsymbol\lambda^*(\mathbf{z}))$ has rank $n - k$. Now, consider the system of equations made of $\mathbf{P}\nabla r(\boldsymbol\lambda^*(\mathbf{z})) = 0$ and $\mathbf{A}_\mathcal{I}\boldsymbol\lambda^*(\mathbf{z}) = \mathbf{z}$, which determines $\boldsymbol\lambda^*$ as an implicit function of $\mathbf{z}$. By the implicit function theorem, because the Jacobian matrix with respect to $\boldsymbol\lambda^*$ is invertible, $\boldsymbol\lambda^*(\mathbf{z})$ is continuously differentiable. We denote its derivative by $\nabla\boldsymbol\lambda^*(\mathbf{z})$, which satisfies $\mathbf{A}_\mathcal{I}\nabla\boldsymbol\lambda^*(\mathbf{z}) = \mathbf{I}$. Note that this implies that $\mathbf{A}_\mathcal{I}$ has linearly independent rows and $\nabla\boldsymbol\lambda^*(\mathbf{z})$ has linearly independent columns.

Because $\nabla R(\mathbf{z}) = \boldsymbol\mu^*(\mathbf{z})$ by Equation (A.2) and $\nabla r(\boldsymbol\lambda^*(\mathbf{z})) = \mathbf{A}_\mathcal{I}^{\mathrm T}\mu^*(\mathbf{z})$, we have

$$\mathbf{A}_\mathcal{I}^{\mathrm T}\nabla R(\mathbf{z}) = \mathbf{A}_\mathcal{I}^{\mathrm T}\boldsymbol\mu^*(\mathbf{z}) = \nabla r(\boldsymbol\lambda^*(\mathbf{z})).$$

As $\boldsymbol\lambda^*(\mathbf{z})$ is differentiable, using the chain rule, we get

$$\mathbf{A}_\mathcal{I}^{\mathrm T}\nabla^2 R(\mathbf{z}) = \nabla^2 r(\boldsymbol\lambda^*(\mathbf{z}))\nabla\boldsymbol\lambda^*(\mathbf{z}).$$

Multiplying both sides by $\nabla\boldsymbol\lambda^*(\mathbf{z})^{\mathrm T}$ and using $\mathbf{A}_\mathcal{I}\nabla\boldsymbol\lambda^*(\mathbf{z}) = \mathbf{I}$, we have

$$\nabla^2 R(\mathbf{z}) = (\mathbf{A}_\mathcal{I}\nabla\boldsymbol\lambda^*(\mathbf{z}))^{\mathrm T}\nabla^2 R(\mathbf{z}) = \nabla\boldsymbol\lambda^*(\mathbf{z})^{\mathrm T}\nabla^2 r(\boldsymbol\lambda^*(\mathbf{z}))\nabla\boldsymbol\lambda^*(\mathbf{z}).$$

Recall that $\nabla^2 r(\boldsymbol\lambda^*(\mathbf{z}))$ is negative definite by Assumption B2. Because $\nabla\boldsymbol\lambda^*(\mathbf{z})$ has linearly independent columns, the Hessian $\nabla^2 R(\mathbf{z})$ is negative definite, which implies that $R(\mathbf{z})$ is strictly concave.

Next, we establish the smoothness condition of $R(\mathbf{z})$. Recall that $\boldsymbol\lambda^*(\mathbf{z})$ satisfies $\mathbf{P}\nabla r(\boldsymbol\lambda^*(\mathbf{z})) = 0$ and $\mathbf{A}_\mathcal{I}\boldsymbol\lambda^*(\mathbf{z}) = \mathbf{z}$; the Jacobian matrix with respect to $\boldsymbol\lambda^*$ is invertible. Because ∇r is twice continuously differentiable by Assumption B3, by the implicit function theorem for the C^2 class (Krantz and Parks 2012, theorem 3.3.1), $\boldsymbol\lambda^*(\mathbf{z})$ is also twice continuously differentiable, and hence, $\nabla^2 R(\mathbf{z}) = \nabla\boldsymbol\lambda^*(\mathbf{z})^{\mathrm T}\nabla^2 r(\boldsymbol\lambda^*(\mathbf{z}))\nabla\boldsymbol\lambda^*(\mathbf{z})$ is continuously differentiable. Because the domain $\mathcal{Z}$ is compact by Assumption B1, $\nabla^3 R(\mathbf{z})$ is uniformly bounded in $\mathcal{Z}$.

Finally, the third property of the lemma regarding the Lipschitz continuity of $\boldsymbol\lambda^*(\mathbf{z})$ is straightforward because the domain $\mathcal{Z}$ is compact, and we show that $\boldsymbol\lambda^*(\mathbf{z})$ is continuously differentiable in $\mathbf{z}$. □

Proof of Lemma 5. We first prove Equation (39) by induction. The base case of $\tau = T$ clearly holds because $\mathbf{x}_\tau = \mathbf{x}_T$, which belongs to $\mathcal{S}_\mathcal{I}$ by Assumption C1. Suppose Equation (39) holds at period τ. Because $\tau \ge T^\sharp$, it holds that $\|\overline{\boldsymbol\xi}^r_{\to\tau}\| \le \delta_0$, and therefore, by Assumption C1 and Lemma 3. Let $\boldsymbol\lambda_\tau = \arg\max_{\boldsymbol\lambda\in\mathcal{D}}\{r(\boldsymbol\lambda), \text{ s.t. } \mathbf{A}\boldsymbol\lambda \le \mathbf{x}_\tau^r\}$ be the demand rate chosen by the resolving heuristic at period τ and let $\boldsymbol\rho_\tau = \mathbf{A}\boldsymbol\lambda_\tau$. Because $\mathbf{x}_\tau^r \in \mathcal{S}_\mathcal{I}$, we have $\boldsymbol\rho_\tau(\mathcal{I}) = \mathbf{x}_\tau^r(\mathcal{I})$ and $\boldsymbol\rho_\tau(\mathcal{U}) < \mathbf{x}_\tau^r(\mathcal{U})$. Therefore,

$$\begin{aligned}\mathbf{x}_{\tau-1}^r(\mathcal{I}) &= \mathbf{x}_\tau^r(\mathcal{I}) - \frac{\boldsymbol\rho_\tau(\mathcal{I}) - \mathbf{x}_\tau^r(\mathcal{I}) + \boldsymbol\xi_\tau^r(\mathcal{I})}{\tau-1} = \mathbf{x}_\tau^r(\mathcal{I}) - \frac{\boldsymbol\xi_\tau^r(\mathcal{I})}{\tau-1}\\ &= \mathbf{x}_T(\mathcal{I}) - \overline{\boldsymbol\xi}^r_{\to\tau-1}(\mathcal{I});\\ \mathbf{x}_{\tau-1}^r(\mathcal{U}) &= \mathbf{x}_\tau^r(\mathcal{U}) - \frac{\boldsymbol\rho_\tau(\mathcal{U}) - \mathbf{x}_\tau^r(\mathcal{U}) + \boldsymbol\xi_\tau^r(\mathcal{U})}{\tau-1} \ge \mathbf{x}_\tau^r(\mathcal{U}) - \frac{\boldsymbol\xi_\tau^r(\mathcal{U})}{\tau-1}\\ &\ge \mathbf{x}_T(\mathcal{U}) - \overline{\boldsymbol\xi}^r_{\to\tau-1}(\mathcal{U}).\end{aligned}$$

Furthermore, if $\tau > T^\sharp$, we have $\mathbf{x}_{\tau-1}^r \ge \mathbf{0}$ by the definition of $\overline{\boldsymbol\xi}^r_{\to\tau-1}$ and $T^\sharp$, which implies that lost sales do no occur in period τ.

Next, we prove the upper bound on $\mathbb{E}[T^\sharp]$. Recall that $\boldsymbol{\xi}_T^r = \mathbf{A}\varepsilon_\tau^r$. By Assumption B4, $\{\|\overline{\boldsymbol{\xi}}_{\to\tau}^r\|^2\}$ is a submartingale and $\|\boldsymbol{\xi}_\tau^r\| \le B_\xi$ *a.s.* Using the identical proof by Doob's martingale inequality in Lemma 1, we have $\mathbb{E}[T^\sharp] = O(1)$. □

Appendix B. The HO Benchmark

The HO benchmark is used to analyze resolving algorithms for quantity-based network revenue management (Reiman and Wang 2008, Bumpensanti and Wang 2020). Because, in quantity-based network revenue management, the demand rates are not affected by the (adaptively chosen) prices, the formulation in Bumpensanti and Wang (2020) is not directly applicable to our setting. Instead, we formulate an HO benchmark following the strategy in Vera et al. (2021), which also considers price-based revenue management with a finite subset of prices.

Definition B.1 (The HO Benchmark). For any p, define random variable $D_T(p) := \sum_{t=1}^T d_t$ as the total realized demand with fixed price $p_t \equiv p$. A policy π is HO-admissible if, at time t, the price decision p_t depends only on $\{p_{t'}, x_{t'}, d_{t'}\}_{t'>t}$ and $\{D_T(p)\}_{p\in[\underline{p},\overline{p}]}$. The HO benchmark $R^{\mathrm{HO}}(T, x_0)$ is defined as the expected revenue of the optimal HO-admissible policy π.

At a high level, the HO benchmark equips a policy with the knowledge of the total realized demand for *each* hypothetical fixed price $p \in [\underline{p}, \overline{p}]$ in hindsight. Clearly, such policies are more powerful than an ordinary admissible policy, which only knows the *expected* demand but not the realized demand for a specific price p.

Our next proposition shows that the HO benchmark $R^{\mathrm{HO}}(T, x_T)$ has a constant gap compared against the fluid model value $Tp^*f(p^*)$ in the single-product setting. Hence, it also has an $\Omega(\ln T)$ gap from the resolving heuristic and the optimal DP solution. The conclusion in Theorem 4 then holds with $Tr(x_T)$ replaced by $R^{\mathrm{HO}}(T, x_T)$.

Proposition B.1. *For any $x_T \in (\underline{d}, \lambda^u)$, it holds that $R^{\mathrm{HO}}(T, x_T) \ge Tr(x_T) - O(1)$, where $y_T = x_T T$.*

Proof of Proposition B.1. Consider a single-product setting in which $\xi_T, \ldots, \xi_1$ are independent and identically distributed and inventory is constrained $(x_T < \lambda^u)$. It is clear that knowing the aggregate demand $\{D_T(p)\}_{p\in[\underline{p},\overline{p}]}$ is equivalent to knowing $\overline{\xi} = \frac{1}{T}\sum_{t=1}^T \xi_t$ because $D_T(p) = T(f(p) + \overline{\xi})$ for all p. Given the knowledge of $\overline{\xi}$, the HO benchmark sets the price to $p = f^{-1}(x_T - \overline{\xi})$ for all periods so that the realized total demand $T(f(p) + \overline{\xi})$ exactly matches the inventory level $x_T T$. Let $g := f^{-1}$ be the inverse demand function. Recall that we assume g is twice differentiable in $(\underline{d}, \overline{d})$ with $|g''(d)| \le C$ (Assumption A3). By applying the Taylor expansion of g because $\mathbb{E}[\overline{\xi}] = 0$ and $\mathbb{E}[\overline{\xi}^2] = O(1/T)$, the expected regret of the HO benchmark can be bounded as

$$T\mathbb{E}_{\overline{\xi}}\left[x_T g(x_T - \overline{\xi})\right] - Tx_T g(x_T) \ge Tx_T \mathbb{E}_{\overline{\xi}}\left[-g'(x_T)\overline{\xi} - \frac{C}{2}\overline{\xi}^2\right]$$
$$= -\frac{x_T C}{2} T\mathbb{E}[\overline{\xi}^2] = -O(1).$$

Because $Tr(x_T) = Tx_T g(x_T)$, the proof is complete. □

References

Arlotto A, Gurvich I (2019) Uniformly bounded regret in the multisecretary problem. *Stochastic Systems* 9(3):231–260.
Atar R, Reiman MI (2012) Asymptotically optimal dynamic pricing for network revenue management. *Stochastic Systems* 2(2): 232–276.
Aviv Y, Pazgal A (2005) Dynamic pricing of short life-cycle products through active learning. Working paper, Olin School Business, Washington University, St. Louis, MO.
Besbes O, Zeevi A (2009) Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Oper. Res.* 57(6):1407–1420.
Besbes O, Zeevi A (2012) Blind network revenue management. *Oper. Res.* 60(6):1537–1550.
Broder J, Rusmevichientong P (2012) Dynamic pricing under a general parametric choice model. *Oper. Res.* 60(4):965–980.
Bumpensanti P, Wang H (2020) A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Sci.* 66(7):2993–3009.
Chen Y, Farias VF (2013) Simple policies for dynamic pricing with imperfect forecasts. *Oper. Res.* 61(3):612–624.
Cooper WL (2002) Asymptotic behavior of an allocation policy for revenue management. *Oper. Res.* 50(4):720–727.
den Boer AV, Zwart B (2015) Dynamic pricing and learning with finite inventories. *Oper. Res.* 63(4):965–978.
Ferreira KJ, Simchi-Levi D, Wang H (2018) Online network revenue management using Thompson sampling. *Oper. Res.* 66(6): 1586–1602.
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.
Krantz SG, Parks HR (2012) *The Implicit Function Theorem: History, Theory, and Applications* (Springer Science & Business Media, Boston).
Lei YM, Jasin S, Sinha A (2014) Near-optimal bisection search for nonparametric dynamic pricing with inventory constraint. Working paper, Ross School of Business, University of Michigan, Ann Arbor.
Ma W, Simchi-Levi D (2020) Algorithms for online matching, assortment, and pricing with tight weight-dependent competitive ratios. *Oper. Res.* 68(6):1787–1803.
Ma Y, Rusmevichientong P, Sumida M, Topaloglu H (2020) An approximation algorithm for network revenue management under nonstationary arrivals. *Oper. Res.* 68(3):834–855.
Maglaras C, Meissner J (2006) Dynamic pricing strategies for multiproduct revenue management problems. *Manufacturing Service Oper. Management* 8(2):136–148.
Milgrom P, Segal I (2002) Envelope theorems for arbitrary choice sets. *Econometrica* 70(2):583–601.
Reiman MI, Wang Q (2008) An asymptotically optimal policy for a quantity-based network revenue management problem. *Math. Oper. Res.* 33(2):257–282.

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.

Vera A, Banerjee S (2021) The Bayesian prophet: A low-regret framework for online decision making. *Management Sci.* 67(3): 1368–1391.

Vera A, Banerjee S, Gurvich I (2021) Online allocation and pricing: Constant regret via bellman inequalities. *Oper. Res.* 69(3):821–840.

Wang Y, Chen B, Simchi-Levi D (2021) Multi-modal dynamic pricing. *Management Sci.* 67(10):6136–6152.

Wang Z, Deng S, Ye Y (2014) Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Oper. Res.* 62(2):318–331.

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

Yining Wang is an assistant professor in the department of information systems and operations research at Warrington College of Business, University of Florida. His research interests include statistical machine learning, optimization, reinforcement learning, and their applications to data-driven revenue and operations management, such as assortment optimization and dynamic pricing.

He Wang is Colonel John B. Day Early Career Professor and an assistant professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. His research focuses on the interface between machine learning and operations management; he develops data-driven methods for applications in transportation, supply chain, dynamic pricing, and marketplace design.