Operations Research

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

Logarithmic Regret in Multisecretary and Online Linear Programs with Continuous Valuations

Robert L. Bray

To cite this article:
Robert L. Bray (2024) Logarithmic Regret in Multisecretary and Online Linear Programs with Continuous Valuations. Operations Research

Published online in Articles in Advance 17 Jul 2024

. https://doi.org/10.1287/opre.2022.0036

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

OPERATIONS RESEARCH
Articles in Advance, pp. 1–16
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Methods

Logarithmic Regret in Multisecretary and Online Linear Programs with Continuous Valuations

Robert L. Bray[a]

[a] Kellogg School of Management, Northwestern University, Evanston, Illinois 60208

Contact: r-bray@kellogg.northwestern.edu, https://orcid.org/0000-0003-2773-0663 (RLB)

Received: July 15, 2020
Revised: January 20, 2022; August 24, 2023; March 15, 2024
Accepted: May 8, 2024
Published Online in Articles in Advance: July 17, 2024

Area of Review: Stochastic Models

https://doi.org/10.1287/opre.2022.0036

Abstract. I use empirical processes to study how the shadow prices of a linear program that allocates an endowment of $n\beta \in \mathbb{R}^m$ resources to n customers behave as $n \to \infty$. I show the shadow prices (i) adhere to a concentration of measure, (ii) converge to a multivariate normal under central-limit-theorem scaling, and (iii) have a variance that decreases like $\Theta(1/n)$. I use these results to prove that the expected regret in an online linear program is $\Theta(\log n)$, both when the customer variable distribution is known upfront and must be learned on the fly. This result tightens the sharpest known upper bound from $O(\log n \log \log n)$ to $O(\log n)$, and it extends the $\Omega(\log n)$ lower bound known for single-dimensional problems to the multidimensional setting. I illustrate my new techniques with a simple analysis of a multisecretary problem.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2022.0036.

Keywords: online linear program • multisecretary problem • network revenue management • dual convergence • regret bounds • empirical process • dual convergence

1. Introduction

Caley (1875) introduced the secretary problem in the nineteenth century. The problem is to hire a secretary from n applicants that you interview sequentially. However, there's a hitch: Once you interview someone, you must decide whether to hire them before interviewing the next candidate. Therefore, you face an optimal stopping problem, with the objective being to maximize the expected capability of the secretary you hire or, equivalently, to minimize the expectation of your *regret*, the capability difference between the most competent applicant and the one you hire.

Arlotto and Gurvich (2019) studied the *multisecretary* problem, which is the same as the previous problem except with $n\beta$ posts to fill, for some $\beta \in [0,1]$. In this version of the problem, your regret is the expected capability difference between the $n\beta$ most capable applicants and the $n\beta$ applicants you hire. Arlotto and Gurvich made a striking discovery: If secretary valuations are independent and identically distributed (i.i.d.) random variables with finite support, $\{v_1, \ldots, v_k\}$, then your expected regret is uniformly bounded across $n \in \mathbb{N}$ and $\beta \in [0,1]$.

In Section 3, I study the multisecretary problem with secretary valuations drawn from the continuum $[0,1]$, rather than the finite set, $\{v_1, \ldots, v_k\}$. Specifically, I show that the expected regret lies between $(\beta/8)(1 - \beta/8)(\log(n)/2 - \log(6))$ and $(\log(n+1)+7)/8$ for all $n \geq 2^{20}\beta^{-8}$ and $\beta \in [0,1/2]$ when valuations are i.i.d. uniform random variables, and I derive mirror-image bounds for $\beta \in [1/2,1]$.[1] Furthermore, I show that the most obvious heuristic satisfies the upper regret bound.

In Section 4, I extend this $\Theta(\log n)$ regret rate to the more general online linear programming (OLP) problem of Li and Ye (2022). In this problem, you start with inventory vector $n\beta \in \mathbb{R}^m_+$, and you exchange inventory $a_t \in \mathbb{R}^m_+$ for utility $u_t \geq 0$ if you fulfill the period t customer's demand. Since none of your stocks can become negative, you must carefully husband each of your m resources. However, doing so is difficult, as you have no foreknowledge of the nature of demand; instead, you must learn the demand distribution the old-fashioned way—by serving customers.

The engine underlying my analysis of the online linear program is a set of shadow price convergence results I develop in Section 4.2. Li and Ye (2022, p. 2952) lamented that "there is still a lack of theoretical understanding of the properties of the dual optimal solutions," so I begin by characterizing their limiting behavior. I show that an online linear program's shadow prices (i) conform to a concentration of measure, (ii) converge to a multivariate normal under central-limit-theorem-like scaling, and (iii) have a covariance matrix whose elements fall like $\Theta(1/t)$. I derive these results by hemming in the shadow prices with empirical processes.

2. Related Works

2.1. Primary Antecedents

The online linear program I study in Section 4 is a multidimensional extension of the "zero-one knapsack problem" of Lueker (1998), in which you successively decide whether to add objects with random valuations and volumes to your backpack. Lueker proves that the expected regret grows like $\Theta(\log n)$ under the optimal policy, provided that the value-to-volume ratio distribution is sufficiently continuous. He establishes this bound with a proof unlike any other I have found in the literature. Specifically, he constructs lower and upper envelopes across the entire surface of the offline and online value functions. The induction required to create these bounds is painstaking because he has to weaken them just so as the inventory level departs the initial resource endowment. Extending Lueker's value-function-bounding approach to higher dimensions would have been difficult, so I tackled the multiresource version of his problem with the *compensated coupling* scheme of Vera and Banerjee (2019). Rather than construct multidimensional functional envelopes, this more modern approach adds up the *myopic regret* incurred over the inventory level's random walk.

Lueker's specification generalizes the continuous-valuation multisecretary model I study in Section 3, but it does not generalize the finite-valuation multisecretary model that Arlotto and Gurvich (2019) study. Indeed, Arlotto and Gurvich show that they can decrease the regret rate from $\Theta(\log n)$ to $O(1)$ by replacing Lueker's continuous-support secretary valuation distribution with an analogous finite-support distribution.

The mirror image of the multisecretary model is the stochastic knapsack problem of Arlotto and Xie (2020): The former has random valuations and fixed capacity consumption, and the latter has fixed valuations and random capacity consumption. Arlotto and Xie's model does not fit under the framework of Lueker (1998) because it permits an unrestricted knapsack capacity—Whereas Lueker make the backpack volume scale linearly with n, Arlotto and Xie set it to a free model parameter. They use this additional degree of freedom to show that Lueker's $O(\log n)$ upper regret bound holds universally across initial backpack capacities. However, Arlotto and Xie (2020, p. 190) do not develop a corresponding lower bound since "it is well known that the optimal policy often lacks desirable structural properties, so proving [this lower bound] is unlikely to be easy."

Jasin (2014) extends the $O(\log n)$ upper regret bound to the multivariate setting. However, Jasin only supports a finite number of consumption bundles, as he considers the network revenue management problem in which you price a set number of products (e.g., flight itineraries), each of which comprises a set number of resources (e.g., flight legs).

Li and Ye (2022) relax the finite-product assumption, allowing a given customer's resource consumption to be any number in a bounded region of $\mathbb{R}^m$. More importantly, they are the first to incorporate an unknown demand distribution. Specifically, they show that the expected regret is at most $O(\log n \log \log n)$ when the agent starts without knowing the nature of demand. However, Li and Ye (2022) do not provide a corresponding lower regret bound, so when you compare their $O(\log n \log \log n)$ bound with the prior $\Theta(\log n)$ results, you cannot help but wonder: Is revenue management with and without online learning in the same class of difficulty?

I show that they are. Specifically, I establish that the regret is $O(\log n)$ when the demand distribution is unknown and is $\Omega(\log n)$ when it is known—i.e., that it is $\Theta(\log n)$, both with and without demand distribution foreknowledge. Removing the $\log \log n$ fudge factor from Li and Ye's upper bound requires (i) more sharply characterizing the limiting behavior of the shadow prices and (ii) more tightly controlling the inventory process. Whereas Li and Ye show that the magnitude of the period-t shadow price covariance matrix is $O((\log \log t)/t)$, I show that it is $\Theta(1/t)$. Also, whereas they constrain inventories for all but the last $O(\log n \log \log n)$ periods, I constrain them for all but the last $O(1)$ periods. New methodological innovations underpin both improvements.

First, I sharpen the shadow price asymptotics by applying empirical process techniques to the subgradient of the dual linear program. Casting this subgradient as an empirical process enables me to create shadow price convergence results that hold uniformly across inventory levels. This, in turn, allows me to overcome the hopeless entanglement between the current inventory level and the current shadow price estimate.

Second, I create new techniques to constrain the inventory level's random walk. For the upper bound with a known demand distribution, I control the process with a standard martingale concentration inequality. For the upper bound with an unknown demand distribution, I split the process into martingale and drift parts. I then apply the martingale concentration inequality to the former and inductively bound the latter, showing that the inventory level being "in control" up until period $t+1$ implies that the period $(t+1)$ shadow price is "in control," which in turn means that the period t inventory level is "in control." (This induction would not have been possible without the empirical process' uniform bounds.) Finally, for the lower bound, I regulate the probability of the inventory levels spiraling out of control with the cost of splitting the offline linear program into two separate linear programs. For example, suppose you have 1,000 applicants for 100 secretarial positions and can interview all the applicants upfront; now, suppose I told you that you can only hire 10 of the last 500 candidates. This additional constraint will substantially decrease the value of your hires, with high probability. Your regret conditional on having fewer than 10 open positions with 500 remaining applicants is at least as large as the cost imposed by this offline constraint. Hence, the probability of having such a low inventory level must be sufficiently small, or the optimal policy would violate the $O(\log n)$ regret upper bound.

2.2. Contemporaneous Developments

I will now discuss the noteworthy advancements that emerged since I first circulated my results. First, Balseiro et al. (2023, p. 1) produced an insightful and comprehensive survey article that organizes models corresponding to "dynamic pricing with capacity constraints, dynamic bidding with budgets, network revenue management, online matching, and order fulfillment" under a unified umbrella, "dynamic resource-constrained reward collection (DRC^2) problems." The DRC^2 framework is similar to the "online resource allocation" framework of Vera et al. (2019), except it can accommodate an infinite number of customer types. Balseiro et al. (2023, p. 8) explain that their class of problems is especially amenable to the "*certainty-equivalent principle:* replace quantities by their expected values and take the best actions given the current history." Indeed, this is how I bound the online linear program's regret, although I learned the technique from Li and Ye (2022).

Next, Jiang and Zhang (2020) extend the model of Arlotto and Xie (2020) to allow multiple servers. Specifically, they suppose that you must allocate each customer to one of m servers. They provide an $O(\log n)$ upper bound, but like Li and Ye (2022) and Arlotto and Xie (2020), do not provide a corresponding lower bound. Neither Jiang and Zhang's multiserver problem nor Li and Ye's OLP problem (which I study) generalize the other. Jiang and Zhang's framework incorporates an additional decision—which server to route a customer to—but Li and Ye's framework incorporates online learning and permits a richer set of restrictions—constraining sales with a general linear program. Moreover, Jiang and Zhang do not make the initial resource endowment scale linear with n, as Li and Ye (2022) and I do.

Wang and Wang (2022) establish an $\Omega(\log n)$ gap between the expected online value and the fluid approximation value (as opposed to the expected offline value) in the network revenue management problem of Jasin (2014). However, they only establish this result for the one-dimensional version of the problem. For the multidimensional version, they show that the optimal policy yields only $O(1)$ more expected value than the policy Jasin used.

Besbes et al. (2023) point out that the multisecretary problem's $O(\log n)$ regret may not hold if the probability density function is near zero near the acceptance-rejection threshold. Akshit Kumar explained it to me like this: If you have n applicants for $n\beta$ open positions, then the marginal applicant would have a valuation of $F^{-1}(1-\beta+\Omega_p(\sqrt{n}))$, where F is the utility cumulative distribution function. Now, if the utility probability density function equals $f(u)=|u-u^*|$ in a neighborhood of $u^*\equiv F^{-1}(1-\beta)$, then we would have $F(u)=1-\beta+\text{sign}(u-u^*)(u-u^*)^2/2$ and hence $F^{-1}(q)=u^*+\text{sign}(q-1+\beta)\sqrt{2|q-1+\beta|}$. In this case, the expected myopic regret would exceed $1/n$ because, rather than the usual $n^{-1/2}$ tolerance, we could now only discern the marginal man's utility to within a $n^{-1/4}$ tolerance: $F^{-1}(1-\beta+\Omega_p(\sqrt{n}))=u^*+\text{sign}(\Omega_p(\sqrt{n}))\sqrt{2|\Omega_p(\sqrt{n})|}=u^*+\Omega_p(n^{-1/4})$. To avoid these low-density regions, Besbes et al. create a version of the certainty-equivalent principle of Balseiro et al. (2023) that is "conservative with respect to gaps." Their algorithm steers the inventory random walk away from regions with high expected myopic regret.

Finally, Jiang et al. (2024) independently developed an $O(1/n)$ bound for the shadow price variance. They combine this dual convergence result with a technique that's similar in spirit to Besbes et al.'s "conservative with respect to gaps" to establish an $O(\log^2 n)$ regret bound for the network revenue management problem without imposing a nondegenerate fluid limit. In contrast, previous models have assumed the fluid approximation's constraints bind with pressure or are slack, with additional leeway. However, assuming extra wiggle room in the fluid model is unreasonable as it implies that some buffer stocks scale *linearly* with demand, which is a way over investment since safety stocks ought to scale with the square root of sales.

3. Multisecretary Problem

I will begin with the simple multisecretary problem to demonstrate my regret-bounding approach. Lueker (1998) has already established that this model's regret scales like $\Theta(\log n)$, but I will provide a far simpler proof, and my bounds will not have any hidden constants.

3.1. Setup

You have $n\in\mathbb{N}$ applicants for $n\beta\in\mathbb{N}$ positions, where $\beta\in[0,1/2]$. (It suffices to consider $\beta\in[0,1/2]$, because the expected regret with $n\beta$ initial open slots equals that with $n(1-\beta)$ initial open slots.[2]) You interview the candidates sequentially, starting with the nth applicant and ending with the first applicant, so that the period number corresponds with the size of the remaining candidate pool, with period $t-1$ succeeding period t. Interviewing the period t applicant reveals the utility you would get from hiring them, u_t, a standard uniform random variable independent of the other candidates' utilities. After interviewing this candidate, you must hire them on the spot or reject them for good. You seek to maximize the expected total utility from your hires. Characterizing this utility will take a few steps.

First, let v_t^b denote the utility you receive starting from period t with $tb \in \mathbb{N}$ open positions. The expectation of this variable satisfies the following Bellman equations:

$$\mathrm{E}(v_t^b) \equiv \mathrm{E}\left(\max_{x_t \in \{0,1\}} x_t u_t + \mathrm{E}(v_{t-1}^{\psi_t^b(x_t)}) \quad \text{s.t.} \quad x_t \le tb\right),$$

$$\mathrm{E}(v_0^b) \equiv 0,$$

$$\text{and} \quad \psi_t^b(a) \equiv \begin{cases} (tb-a)/(t-1) & t > 1, \\ 0 & t = 1. \end{cases}$$

I will explain the logic underlying these equations after line 1. However, the ψ_t^b function maps the fraction of applicants you can hire from period t onward, b, and your period t hiring decision, x_t, to the fraction of applicants you can hire from period $(t-1)$ onward. For example, if the period t superscript is b and you hire the period t applicant—that is, set $x_t = 1$—then the period $(t-1)$ superscript is $\underbrace{tb-1}_{\text{positions left}} \; / \; \underbrace{(t-1)}_{\text{applicants left}}$.

The previous Bellman equations specify the following optimal action:

$$\pi_t^b \equiv \arg\max_{x_t \in \{0,1\}} x_t u_t + \mathrm{E}(v_{t-1}^{\psi_t^b(x_t)}) \quad \text{s.t.} \quad x_t \le tb. \tag{1}$$

The $x_t \le tb$ constraint ensures that you do not extend a job offer if you don't have any positions available—that is, that you set $x_t = 0$ if $tb = 0$. The previous expression states that you hire the period t applicant (i.e., set $x_t = 1$) if you have a job opening (i.e., $1 \le tb$) and if the total expected utility conditional on hiring them (i.e., $u_t + \mathrm{E}(v_{t-1}^{\psi_t^b(1)})$) exceeds the total expected utility conditional on rejecting them (i.e., $\mathrm{E}(v_{t-1}^{\psi_t^b(0)})$).

Your corresponding realized value is

$$v_t^b \equiv \pi_t^b u_t + v_{t-1}^{\psi_t^b(\pi_t^b)} \quad \text{and} \quad v_0^b \equiv 0.$$

Hence, you garner value v_n^β from your n applicants and $n\beta$ positions under the expected-utility-maximizing policy. However, if you could have interviewed every applicant before extending any job offers, then you would have garnered value V_n^β, where

$$V_t^b \equiv \sum_{s=1}^{tb} h_t^s,$$

and h_t^s is the sth highest value in $\{u_t, \ldots, u_1\}$. Since the utilities follow a uniform distribution, order statistic h_t^s follows a beta$(t-s+1, s)$ distribution.

The difference between the aggregate utility received in the offline problem and that received in the online problem is your regret:

$$R_n \equiv V_n^\beta - v_n^\beta.$$

The following two propositions bound the expectation of this random variable.

Proposition 1. *The optimal policy of the multisecretary problem yields an expected regret that grows at no more than a* $\log n$ *rate:* $\mathrm{E}(R_n) \le (\log(n+1)+7)/8$, *for all* $n \in \mathbb{N}$ *and* $\beta \in [0, 1/2]$.

Proposition 2. *The optimal policy of the multisecretary problem yields an expected regret that grows at no less than a* $\log n$ *rate:* $\mathrm{E}(R_n) \ge (\beta/8)(1-\beta/8)(\log(n)/2 - \log(6))$, *for all* $n \ge 2^{20}\beta^{-8}$ *and* $\beta \in [0, 1/2]$.

These theorems provide nonasymptotic results—that is, do not rely on big-O notation. Proposition 1's finite-sample bound is especially interesting, as it highlights the near worthlessness of the value of future information. For example, suppose you have a billion applicants for 500 million jobs. In this case, your online value would be around $\underbrace{(1/2+1)/2}_{\text{value of average hire}} \cdot \underbrace{500 \text{ million}}_{\text{number of hires}} =$ 375 million and your offline value would exceed your online value by around $(\log(10^9+1)+7)/8 = 3.47$. Hence, knowing the billion worker utilities upfront increases your workforce's value by around 3.47/375 million = .00000093%.

3.2. Upper Bound

I will now prove Proposition 1 by showing that Algorithm 1 honors its bound. The proof has two parts: The first decomposes the total regret into a sum of myopic regrets, and the second shows that the expectation of the period t myopic regret is $O(1/t)$ under the myopic-regret-minimizing Algorithm 1, and hence that the expected total regret is $O(\sum_{t=1}^n 1/t) = O(\log n)$.

To derive the policy underlying Algorithm 1, suppose that you hire the period t applicant with tb available positions if and only if their valuation exceeds τ_t^b, where $\{\tau_t^b \mid t \in [n], tb \in \{0, \ldots, t\}\}$ is a collection of thresholds that have yet to be defined. These thresholds will satisfy $\tau_t^0 = 1$ and $\tau_t^1 = 0$ for all $t \in [n]$, to ensure that $b_t \in [0,1]$ for all $t \in [n]$, where

$$b_n \equiv \beta$$

$$\text{and} \quad b_{t-1} \equiv \psi_t^{b_t}(\mathbb{1}\{u_t > \tau_t^{b_t}\}).$$

In other words, you start period t with $tb_t \in \{0, \ldots, n\}$ open positions under the threshold policy. You receive corresponding value $\hat{v}_n$, where

$$\hat{v}_t \equiv \mathbb{1}\{u_t > \tau_t^{b_t}\} u_t + \hat{v}_{t-1},$$

$$\text{and} \quad \hat{v}_0 \equiv 0.$$

Since the value under Algorithm 1 cannot exceed the value under the optimal algorithm, we have $\mathrm{E}(\hat{v}_n) \le \mathrm{E}(v_n^\beta)$, which implies that that

$$\mathrm{E}(\hat{R}_n) \ge \mathrm{E}(R_n),$$
$$\text{where} \quad \hat{R}_t \equiv V_t^{b_t} - \hat{v}_t.$$

Accordingly, it will suffice to upper bound $\mathrm{E}(\hat{R}_n)$. To this end, the offline value function satisfies the following recurrence relations:

$$V_t^b = (u_t - h_{t-1}^{tb})^+ + V_{t-1}^{\psi_t^b(0)} \tag{2}$$
$$\text{and} \quad V_{t-1}^{\psi_t^b(0)} = h_{t-1}^{tb} + V_{t-1}^{\psi_t^b(1)}. \tag{3}$$

Line (2) states that if there are tb open positions, then the value of increasing the size of the applicant pool from $t-1$ to t equals the option value of replacing the tbth most capable person, out of the first $t-1$ applicants, with the period t applicant. Line (3) states that if there are $t-1$ remaining applicants then the value of increasing the number of job openings from $(t-1)\psi_t^b(1) = tb-1$ to $(t-1)\psi_t^b(0) = tb$ positions equals the value of the tbth best applicant out of these $t-1$ candidates.

Algorithm 1 (Minimize Myopic Regret)

```
1. input $n$, $\beta$, $\{u_t\}_{t=1}^n$,
2. initialize $b_n := \beta$
3. for $t$ from $n$ to $1$ do
     (a) set $x_t := \mathbb{1}\{u_t > 1 - b_t\}$
     (b) set $b_{t-1} := \psi_t^{b_t}(x_t)$
4. end for
5. output $\{x_t\}_{t=1}^n$
```

Now suppose that $u_t \le \tau_t^{b_t}$, and hence that $b_{t-1} = \psi_t^{b_t}(0)$ and $\hat{v}_t = \hat{v}_{t-1}$. In this case, (2) implies that

$$\begin{aligned}
\hat{R}_t &= V_t^{b_t} - \hat{v}_t \\
&= (u_t - h_{t-1}^{tb})^+ + V_{t-1}^{\psi_t^b(0)} - \hat{v}_{t-1} \\
&= (u_t - h_{t-1}^{tb})^+ + V_{t-1}^{b_{t-1}} - \hat{v}_{t-1} \\
&= (u_t - h_{t-1}^{tb})^+ + \hat{R}_{t-1}.
\end{aligned}$$

Next, suppose that $u_t > \tau_t^{b_t}$, and hence that $b_{t-1} = \psi_t^{b_t}(1)$ and $\hat{v}_t = u_t + \hat{v}_{t-1}$. In this case, (2) and (3) imply that

$$\begin{aligned}
\hat{R}_t &= V_t^{b_t} - \hat{v}_t \\
&= (u_t - h_{t-1}^{tb})^+ + V_{t-1}^{\psi_t^b(0)} - \hat{v}_t \\
&= (u_t - h_{t-1}^{tb})^+ + (h_{t-1}^{tb} + V_{t-1}^{\psi_t^b(1)}) - (u_t + \hat{v}_{t-1}) \\
&= (u_t - h_{t-1}^{tb})^- + \hat{R}_{t-1}.
\end{aligned}$$

Combining these two recurrence relations inductively yields

$$\hat{R}_n = r_n + \hat{R}_{n-1} = \sum_{t=1}^n r_t, \tag{4}$$

$$\begin{aligned}
\text{where} \quad r_t &\equiv \mathbb{1}\{u_t \le \tau_t^{b_t}\}(u_t - h_{t-1}^{tb_t})^+ + \mathbb{1}\{u_t > \tau_t^{b_t}\}(u_t - h_{t-1}^{tb_t})^- \\
&= (\mathbb{1}\{u_t > h_{t-1}^{tb_t}\} - \mathbb{1}\{u_t > \tau_t^{b_t}\})(u_t - h_{t-1}^{tb_t}).
\end{aligned}$$

In the previous expression, r_t is your *myopic regret*, which is the cost of your period t hiring mistake. Total regret can always be decomposed into a sum of myopic regrets.

Now, here's the key: We can integrate over u_t and h_{t-1}^{tb} when taking the expectation of r_t because these variables are independent of each other and b_t. To integrate over u_t, we use the fact that this uniform random variable satisfies $\mathrm{E}((\mathbb{1}\{u_t > h\} - \mathbb{1}\{u_t > \tau\})(u_t - h)) = h^2/2 - h\tau + \tau^2/2$ for constants h and τ. To integrate over h_{t-1}^{tb}, we use the fact that this beta$(t - tb, tb)$ random variable satisfies $\mathrm{E}(h_{t-1}^{tb}) = 1 - b$ and $\mathrm{E}(h_{t-1}^{tb})^2 = \frac{(1-b)+t(1-b)^2}{t+1}$. These properties enable us to express the expected myopic regret in terms of b_t and $\tau_t^{b_t}$:

$$\begin{aligned}
\mathrm{E}(r_t) &= \mathrm{E}(\mathrm{E}((\mathbb{1}\{u_t > h_{t-1}^{tb_t}\} - \mathbb{1}\{u_t > \tau_t^{b_t}\})(u_t - h_{t-1}^{tb_t}) \mid h_{t-1}^{tb_t} \\
&\quad = h,\ b_t = b)) \\
&= \mathrm{E}(\mathrm{E}(((h_{t-1}^{tb_t})^2/2 - h_{t-1}^{tb_t}\tau_t^{b_t} + (\tau_t^{b_t})^2/2) \mid b_t = b)) \\
&= \mathrm{E}\left(\frac{(1-b_t) + t(1-b_t)^2}{2(t+1)} - \tau_t^{b_t}(1-b_t) + (\tau_t^{b_t})^2/2\right).
\end{aligned} \tag{5}$$

I will now minimize the previous expectation by setting $\tau_t^b = 1 - b$ (as specified by Algorithm 1), in which case the expression above simplifies to

$$\mathrm{E}(r_t) = \frac{\mathrm{E}(b_t(1-b_t))}{2(t+1)}.$$

With this, we find that the regret incurred under Algorithm 1 satisfies our logarithmic bound:

$$\begin{aligned}
\mathrm{E}(R_n) &\le \mathrm{E}(\hat{R}_n) \\
&= \sum_{t=1}^n \mathrm{E}(r_t) \\
&= \sum_{t=1}^n \frac{\mathrm{E}(b_t(1-b_t))}{2(t+1)} \\
&\le \sum_{t=1}^n \sup_{b \in (0,1)} \frac{b(1-b)}{2(t+1)} \\
&= \sum_{t=1}^n \frac{1}{8(t+1)} \\
&\le (\log(n+1) + 7)/8.
\end{aligned}$$

3.3. Lower Bound

I will now prove Proposition 2. The proof has four steps. The first creates an optimal policy version of the regret decomposition derived in the last section. The decomposition is the same as before, except b_t now denotes the number of open positions under the optimal algorithm rather than under Algorithm 1. The second part of the proof shows that $\Omega(\log n)$ expected regret follows immediately from the regret decomposition, provided that there is an $\Omega(1)$ chance of b_t being bounded away from either endpoint. Finally, the third part of the proof bounds the chance of b_t being too close to one, and the fourth part bounds the chance of it being too close to zero.

To begin the proof, the objective in (1) is supermodular in x_t and u_t. Hence, Topkis's theorem implies that there exists threshold collection

$$\{\tau_t^b \mid t \in [n],\ tb \in \{0, \dots, t\}\}, \tag{6}$$

such that the optimal policy hires the period t applicant with tb available positions if and only if $u_t > \tau_t^b$. As before, these thresholds satisfy $\tau_t^0 = 1$ and $\tau_t^1 = 0$, since the optimal policy always makes exactly n job offers.

Now, since the optimal policy has a threshold structure, Lines (4) and (5) imply that

$$\begin{aligned}
\mathrm{E}(R_n) &= \sum_{t=1}^{n} \mathrm{E}\left(\frac{(1-b_t)+t(1-b_t)^2}{2(t+1)} - \tau_t^{b_t}(1-b_t) + (\tau_t^{b_t})^2/2\right)\\
&\geq \sum_{t=1}^{n} \mathrm{E}\left(\min_{\tau}\left(\frac{(1-b_t)+t(1-b_t)^2}{2(t+1)} - \tau(1-b_t) + \tau^2/2\right)\right)\\
&= \sum_{t=1}^{n} \frac{\mathrm{E}(b_t(1-b_t))}{2(t+1)}.
\end{aligned} \tag{7}$$

Keep in mind that b_t now characterizes the number of open positions under the optimal thresholds defined in (6):

$$\begin{aligned}
&b_n \equiv \beta\\
\text{and}\quad &b_{t-1} \equiv \psi_t^{b_t}(\mathbb{1}\{u_t > \tau_t^{b_t}\}).
\end{aligned} \tag{8}$$

Lower bounding Expression (7) will require upper bounding the probability that b_t veers too closely to either endpoint. For this, I will show that $n \geq 2^{20}\beta^{-8}$ and $\sqrt{n} \leq t \leq n/2$ imply

$$\Pr(b_t < \beta/8) \geq \Pr(b_t > 1-\beta/8) \tag{9}$$

$$\text{and}\quad \Pr(b_t < \beta/8) \leq 1/4. \tag{10}$$

Combining these bounds with Line (7) yields Proposition 2:

$$\begin{aligned}
\mathrm{E}(R_n) &\geq \sum_{t=1}^{n} \frac{\Pr(\beta/8 \leq b_t \leq 1-\beta/8)\beta/8(1-\beta/8)}{2(t+1)}\\
&= \sum_{t=1}^{n} \frac{(1-\Pr(b_t<\beta/8)-\Pr(b_t>1-\beta/8))\beta/8(1-\beta/8)}{2(t+1)}\\
&\geq \sum_{t=\lceil\sqrt{n}\rceil}^{\lfloor n/2\rfloor} \frac{(1-1/4-1/4)\beta/8(1-\beta/8)}{2(t+1)}\\
&\geq \int_{t=2\sqrt{n}}^{n/3} (\beta/8)(1-\beta/8)/(8t)dt\\
&= (\beta/8)(1-\beta/8)(\log(n)/2-\log(6)).
\end{aligned}$$

Accordingly, it will suffice to establish Lines (9) and (10). I will begin with the former because it is more straightforward. Simply put, the $\{b_t\}_{t=n}^1$ process is more likely to approach the left endpoint than the right endpoint because it starts at $\beta \leq 1/2$ and is symmetric about $1/2$.

I will now formalize this intuition with a coupling argument. First, the problem symmetry discussed in Endnote 2 implies that the acceptance thresholds satisfy

$$\tau_t^b = 1 - \tau_t^{1-b}. \tag{11}$$

Basically, this holds because one minus a uniform is also a uniform. Second, consider the following benchmark process:

$$\begin{aligned}
&\hat{b}_n \equiv 1-\beta\\
\text{and}\quad &\hat{b}_{t-1} \equiv \psi_t^{\hat{b}_t}(\mathbb{1}\{u_t > \tau_t^{\hat{b}_t}\}).
\end{aligned}$$

The $\{b_t\}_{t=n}^1$ and $\{\hat{b}_t\}_{t=n}^1$ processes cannot jump over one another, because the number of open positions can only decrease by one or remain constant in a given period. The processes couple whenever they meet, with $b_t = \hat{b}_t$ implying $b_{t-1} = \hat{b}_{t-1}$. Accordingly, $\hat{b}_t < \beta/8$ implies $b_t < \beta/8$, and hence $\Pr(\hat{b}_t < \beta/8) \leq \Pr(b_t < \beta/8)$. Third, since one minus a uniform is also a uniform, the process $\{\hat{b}_t\}_{t=n}^1$ has the same distribution as the process $\{\tilde{b}_t\}_{t=n}^1$, where

$$\begin{aligned}
&\tilde{b}_n \equiv 1-\beta\\
\text{and}\quad &\tilde{b}_{t-1} \equiv \psi_t^{\tilde{b}_t}(\mathbb{1}\{1-u_t > \tau_t^{\tilde{b}_t}\}).
\end{aligned}$$

With (11), we can rearrange these equations like this:

$$\begin{aligned}
&1-\tilde{b}_n = \beta\\
\text{and}\quad &1-\tilde{b}_{t-1} = 1-\psi_t^{\tilde{b}_t}(\mathbb{1}\{1-u_t > \tau_t^{\tilde{b}_t}\})\\
&\qquad = 1-\psi_t^{\tilde{b}_t}(\mathbb{1}\{u_t < \tau_t^{1-\tilde{b}_t}\})\\
&\qquad = \frac{t-1-t\tilde{b}_t+\mathbb{1}\{u_t < \tau_t^{1-\tilde{b}_t}\}}{t-1}\\
&\qquad = \frac{t(1-\tilde{b}_t)-\mathbb{1}\{u_t \geq \tau_t^{1-\tilde{b}_t}\}}{t-1}\\
&\qquad = \psi_t^{1-\tilde{b}_t}(\mathbb{1}\{u_t \geq \tau_t^{1-\tilde{b}_t}\}).
\end{aligned}$$

Compare this system to (8), and you will see that $1-\tilde{b}_t = b_t$, almost surely. Accordingly, $\Pr(b_t > 1-\beta/8) = \Pr(\tilde{b}_t < \beta/8) = \Pr(\hat{b}_t < \beta/8) \leq \Pr(b_t < \beta/8)$, which establishes (9).

Finally, I will establish (10). The argument has three steps. First, I establish that the regret conditional on $b_t < \beta/8$ is at least as high as the value you would get by replacing the worst $\lfloor t\beta/8\rfloor$ applicants hired before period t with the best $\lfloor t\beta/8\rfloor$ applicants rejected after period t, which is at least as high as $\lfloor t\beta/8\rfloor$ times the difference between the value of the $(tb_t + \lfloor t\beta/8\rfloor)$th best applicant interviewed after period t and the $(n\beta - tb_t - \lfloor t\beta/8 \rfloor + 1)$th best applicant interviewed before period t. Second, I use the binomial Chernoff to establish that there is at least a $1 - 1/12 - 1/12 = 5/6$ chance that the $(tb_t + \lfloor t\beta/8 \rfloor)$th best applicant interviewed after period t is at least $\beta/2$ units better than the $(n\beta - tb_t - \lfloor t\beta/8 \rfloor + 1)$th best applicant interviewed before period t. Third, I use these results to show that the optimal policy would violate the $(\log(n+1)+7)/8$ upper regret bound if the event $b_t < \beta/8$ was not sufficiently rare.

Now we will prove Line (10). Conditional on having tb_t open positions at the start of period t, the best the online policy can do is hire the best tb_t out of the last t applicants and hire the best $n\beta - tb_t$ out of the first $n - t$ applicants. Thus, the online value must satisfy

$$v_n^\beta \le \sum_{s=1}^{tb_t} h_t^s + \sum_{s=1}^{n\beta - tb_t} \underleftarrow{h}_t^s,$$

where look-back order statistic $\underleftarrow{h}_t^s$ is the sth largest value in $\{u_n, \dots, u_{t+1}\}$ (i.e., it equals h_{n-t}^s, but with the order of the applicants reversed). Furthermore, if $b_t < \beta/8$, then the offline policy could hire the best $tb_t + \lfloor t\beta/8 \rfloor$ out of the last t applicants and the best $n\beta - tb_t - \lfloor t\beta/8 \rfloor$ out of the first $n - t$ applicants. Hence, the offline value must satisfy the following when $b_t < \beta/8$:

$$V_n^\beta \ge \sum_{s=1}^{tb_t + \lfloor t\beta/8 \rfloor} h_t^s + \sum_{s=1}^{n\beta - tb_t - \lfloor t\beta/8 \rfloor} \underleftarrow{h}_t^s.$$

Differencing the last two inequalities yields the following for $b_t < \beta/8$:

$$\begin{aligned}
R_n &\ge \sum_{s=tb_t+1}^{tb_t + \lfloor t\beta/8 \rfloor} h_t^s - \sum_{s=n\beta - tb_t - \lfloor t\beta/8 \rfloor + 1}^{n\beta - tb_t} \underleftarrow{h}_t^s \\
&\ge \lfloor t\beta/8 \rfloor h_t^{tb_t + \lfloor t\beta/8 \rfloor} - \lfloor t\beta/8 \rfloor \underleftarrow{h}_t^{n\beta - tb_t - \lfloor t\beta/8 \rfloor + 1} \\
&\ge \lfloor t\beta/8 \rfloor (h_t^{\lfloor t\beta/4 \rfloor} - \underleftarrow{h}_t^{n\beta - \lfloor t\beta/4 \rfloor}) \\
&\ge \lfloor t\beta/8 \rfloor \mathbb{1}\{h_t^{\lfloor t\beta/4 \rfloor} \ge 1 - 3\beta/8\} \\
&\qquad \mathbb{1}\{\underleftarrow{h}_t^{n\beta - \lfloor t\beta/4 \rfloor} \le 1 - 7\beta/8\}(7\beta/8 - 3\beta/8) \\
&\ge \lfloor t\beta^2/16 \rfloor \mathbb{1}\{h_t^{\lfloor t\beta/4 \rfloor} \ge 1 - 3\beta/8\} \\
&\qquad \mathbb{1}\{\underleftarrow{h}_t^{n\beta - \lfloor t\beta/4 \rfloor} \le 1 - 7\beta/8\}.
\end{aligned}$$

The first line states that your regret is at least as large as the benefit you would get by replacing the worst $\lfloor t\beta/8 \rfloor$ applicants hired before period t with the best $\lfloor t\beta/8 \rfloor$ applicants rejected after period t. The second line maintains that the value of this difference is at least as large as $\lfloor t\beta/8 \rfloor$ (i.e., the number of people exchanged) times the difference between $h_t^{tb_t + \lfloor t\beta/8 \rfloor}$ (i.e., the value of the worst candidate added) and $\underleftarrow{h}_t^{n\beta - tb_t - \lfloor t\beta/8 \rfloor + 1}$ (i.e., the value of the best candidate removed). The remaining three lines use the fact that h_t^s decreases in its superscript to connect the bound with the following binomial Chernoff results: If $t \ge 48 \log(12)/\beta$, $n \ge 336 \log(12)/\beta$, and $\sqrt{n} \le t \le n/2$ then

$$\begin{aligned}
&\Pr(h_t^{\lfloor t\beta/4 \rfloor} \ge 1 - 3\beta/8) \ge 11/12 \\
\text{and} \quad &\Pr(\underleftarrow{h}_t^{n\beta - \lfloor t\beta/4 \rfloor} \le 1 - 7\beta/8) \ge 11/12,
\end{aligned}$$

and, accordingly, Proposition 1 and Bonferroni's inequality imply the following for the specified range of n and t:

$$\begin{aligned}
&(\log(n+1)+7)/8 \\
&\quad \ge \mathrm{E}(R_n) \\
&\quad \ge \lfloor t\beta^2/16 \rfloor \Pr(b_t < \beta/8 \cap h_t^{\lfloor t\beta/4 \rfloor} \ge 1 - 3\beta/8 \cap \underleftarrow{h}_t^{\lfloor (n-t)\beta \rfloor} \\
&\qquad \le 1 - 7\beta/8) \\
&\quad \ge \lfloor \sqrt{n}\beta^2/16 \rfloor (\Pr(b_t < \beta/8) + 11/12 + 11/12 - 2).
\end{aligned}$$

Finally, this inequality implies (10) when $n \ge 2^{20}\beta^{-8}$ and $\sqrt{n} \le t \le n/2$.

4. Online Linear Programming Problem

4.1. Model

I will now extend the techniques developed in the last section to the online linear program of Li and Ye (2022).[3] See Table A.1 in the appendix for a notation guide and the online supplement for the proofs.

As before, I will count backward from period $n \in \mathbb{N}$ to period 1, positioning period $t-1$ after period t. In each period, a customer arrives, and you must decide whether to fulfill their demand from your inventory. You begin in period n with initial inventory endowment $nb_n = n\beta$, for some given $\beta \in \mathbb{R}_+^m$, so that you have $e_j' b_n$ units of the jth resource budgeted for the "average" remaining period, where e_j is the unit vector indicating the jth position. If you satisfy the period–n customer then you exchange inventory bundle $a_n \in \mathbb{R}_+^m$ for utility u_n, so that you begin period $n-1$ with resource vector $b_{n-1} \equiv (nb_n - a_n)/(n-1)$ (both nb_n and a_n can take noninteger values). If, on the other hand, you reject the period n customer, then you receive no utility and lose no resources, so that you begin period $n-1$ with resource vector $b_{n-1} \equiv nb_n/(n-1)$. This pattern repeats so that $b_{t-1} \equiv (tb_t - a_t)/(t-1)$ if you satisfy the period–t customer and $b_{t-1} \equiv (tb_t)/(t-1)$ otherwise. The problem is dynamic because you do not observe variables u_t and a_t until the beginning of period t. These variables satisfy the following assumptions.

Assumption 1. *The customers are i.i.d.: vectors $\{(u_t, a_t)\}_{t=1}^n$ are drawn independently of one another, from joint distribution μ.*

Assumption 2. *The utilities and resource requirements are nonnegative: $u_1, a_1 \ge 0$ almost surely.*

Assumption 3. *The utilities have finite expectation:* $\mathrm{E}(u_1) < \infty$.

Assumption 4. *The resource requirements are bounded:* $a_1 \le \alpha$*, almost surely, for some* $\alpha \in \mathbb{R}_+^m$.

Note that u_1 can have unbounded support, whereas the other models cited in Section 2—most notably those of Lueker (1998) and Li and Ye (2022)—restrict u_1 to a finite range.

Let v_t^b denote the utility you receive from period t onwards when you begin that period with resource endowment $tb \in \mathbb{R}^m$. Since you follow the expected-utility-maximizing policy, this variable's expectation satisfies the following Bellman equations:

$$\mathrm{E}(v_t^b) \equiv \mathrm{E}\left(\max_{x_t \in \{0,1\}} x_t u_t + \mathrm{E}\big(v_{t-1}^{\psi_t^b(x_t a_t)}\big) \ \text{ s.t. } \ x_t a_t \le tb\right), \tag{12}$$

$$\mathrm{E}(v_0^b) \equiv 0,$$

$$\text{and} \quad \psi_t^b(a) \equiv \begin{cases} (tb - a)/(t-1) & t > 1, \\ 0 & t = 1. \end{cases} \tag{13}$$

To better understand this system, consider the following optimal action:

$$\pi_t^b \equiv \arg\max_{x_t \in \{0,1\}} x_t u_t + \mathrm{E}\big(v_{t-1}^{\psi_t^b(x_t a_t)}\big) \quad \text{s.t.} \quad x_t a_t \le tb. \tag{14}$$

In other words, you accept the period t customer (i.e., set $x_t = 1$) if you have inventory enough to do so (i.e., $a_t \le tb$) and if the total expected utility conditional on satisfying this customer (i.e., $u_t + \mathrm{E}(v_{t-1}^{\psi_t^b(1)})$) exceeds the total expected utility conditional on turning them away (i.e., $\mathrm{E}(v_{t-1}^{\psi_t^b(0)})$).

Under this policy you garner total value v_n^β from your initial $n\beta$ resource endowment, where

$$v_t^b \equiv \pi_t^b u_t + v_{t-1}^{\psi_t^b(\pi_t^b a_t)} \quad \text{and} \quad v_0^b \equiv 0. \tag{15}$$

However, if you could have observed all the customer attributes before deciding which ones to satisfy, then you would have garnered value V_n^β, where

$$V_t^b \equiv \max_{x \in \{0,1\}^t} \sum_{s=1}^t x_s u_s \quad \text{s.t.} \quad \sum_{s=1}^t x_s a_s \le tb. \tag{16}$$

Your regret is the difference between the utility you extract when you observe all customer variables upfront and the utility you extract when you learn these variables on the fly:

$$R_n \equiv V_n^\beta - v_n^\beta. \tag{17}$$

Our objective is to show that $\mathrm{E}(R_n) = \Theta(\log n)$ as $n \to \infty$.

Since expanding your choice set from $\{0,1\}$ to $[0,1]$ will not make you worse off, we have

$$\overline{V}_n^\beta \ge V_n^\beta,$$

$$\text{where} \quad \overline{V}_t^b \equiv \max_{x \in [0,1]^t} \sum_{s=1}^t x_s u_s \quad \text{s.t.} \quad \sum_{s=1}^t x_s a_s \le tb \tag{18}$$

$$= \min_{y \in \mathbb{R}_+^m, w \in \mathbb{R}_+^t} tb'y + \sum_{s=1}^t w_t \ \text{ s.t. } \ a_t'y + w_t \ge u_t \ \ \forall\ t, \tag{19}$$

$$= \min_{y \in \mathbb{R}_+^m} t\Lambda_t^b(y), \tag{20}$$

$$\Lambda_t^b(y) \equiv b'y + \sum_{s=1}^t \Delta_s(y)^+/t,$$

$$\text{and} \quad \Delta_t(y) \equiv u_t - a_t'y. \tag{21}$$

Line (18) is the linear programming relaxation of the integer program specified in Line (16). Accordingly, whereas V_t^b is the objective value of the offline optimization problem that gives you resource bundle $tb \in \mathbb{R}^m$ to allocate over t periods, $\overline{V}_t^b$ is the objective of the analogous problem that allows you to satisfy a fraction of a customer's demand. Line (19) is the dual of the problem given in Line (18), with y corresponding to the $\sum_{s=1}^t x_s a_s \le tb$ constraint and w corresponding to the $x_t \le 1$ constraints. Finally, we distill this dual linear program to the convex optimization problem given in Line (20) by replacing w_t with its smallest possible value, $(u_t - a_t'y)^+$. (To remember that this problem is a dual, it helps to think of Λ as an upside-down V.)

The dual problem in (20) has a not-necessarily-unique shadow price minimizer:

$$y_t^b \in \arg\min_{y \in \mathbb{R}_+^m} t\Lambda_t^b(y). \tag{22}$$

Since we initialized $b_n = \beta$, the problem in (20) converges, as $n \to \infty$, to the following deterministic fluid limit:

$$\min_{y \in \mathbb{R}_+^m} \Lambda_\infty^\beta(y) \quad \text{where} \quad \Lambda_\infty^b(y) \equiv b'y + \mathrm{E}(\Delta_1(y)^+). \tag{23}$$

The following assumption endows this limiting problem with a positive shadow price solution.

Assumption 5. *All resource constraints bind in the fluid approximation: There exists* $y_\infty^\beta \in \arg\min_{y \in \mathbb{R}_+^m} \Lambda_\infty^\beta(y)$ *such that* $y_\infty^\beta > 0$.

Extending this assumption to accommodate constraints that are strictly slack in the limit is simple. However, it is harder to accommodate constraints that just barely hold in the limit. See the recent work of Jiang et al. (2024) for an interesting analysis of the degenerate-limit case.

The final assumption is the multivariate analog of the local restriction of Lueker (1998). Lueker imposed two critical constraints on the joint distribution of (u_1, a_1): a local restriction that holds in a neighborhood of the $u_1 = a_1' y_\infty^\beta$ level set and a global restriction that holds across the entire breadth of the distribution. I will need only the former because all the tough calls lie at the margin. For example, the following assumption permits point masses in the distribution, so long as they do not abut the fluid model's accept-reject indifference curve.

Assumption 6. *There's a continuum of marginal customers that strain the resources in a linearly independent fashion: There exists a neighborhood of y_∞^β such that the Jacobian matrix $\frac{\partial}{\partial y}\mathrm{E}(\mathbb{1}\{\Delta_1(y) > 0\}a_1)$ exists, is full rank, and is continuous in y in this neighborhood.*

This assumption is more straightforward than the second-order growth condition imposed by Li and Ye (2022) and many others. Indeed, it simply states that shadow prices give us complete control over inventories. To see this, $\mathrm{E}(\mathbb{1}\{\Delta_1(y) > 0\}a_1)$ is the mean resource consumption rate when we satisfy all customers with positive surplus utility, under shadow price vector y. Accordingly, Jacobian matrix $\frac{\partial}{\partial y}\mathrm{E}(\mathbb{1}\{\Delta_1(y) > 0\}a_1)$ maps marginal shadow price changes to marginal consumption rate changes. This matrix being full rank ensures that we can control the inventory burn rate in a linearly independent fashion by fine-tuning y. For example, marginally shifting the shadow price in the direction of $(\frac{\partial}{\partial y}\mathrm{E}(\mathbb{1}\{\Delta_1(y) > 0\}a_1))^{-1}e_i$ would marginally decrease the consumption of the ith resource, without changing that of the other resources.

Here's a simple sufficient condition that implies Assumption 6.

Example 1. Suppose that given a_1, utility u_1 has bounded and continuous conditional density function $g(u_1|a_1)$, which almost surely satisfies $g(a_1' y_\infty^\beta | a_1) > 0$. Furthermore, suppose that $\mathrm{E}(a_1 a_1')$ is nonsingular. □

The following lemma is equivalent to Assumption 6, so you can consider it an alternative assumption.

Lemma 1. *The limiting problem's second derivative is positive and continuous at its minimizer: Hessian matrix $\ddot{\Lambda}_\infty(y) \equiv \frac{\partial^2}{\partial y^2}\Lambda_\infty^b(y) = -\frac{\partial}{\partial y}\mathrm{E}(\mathbb{1}\{\Delta_1(y) > 0\}a_1)$ exists, is positive definite (and hence full rank), and its elements are continuous in y in a neighborhood of y_∞^β.*

Combining Lemma 1 with Assumption 5 yields the following sister lemma via the implicit function theorem.

Lemma 2. *Limiting shadow prices are locally differentiable in the resource vector: If b is sufficiently close to β, then Λ_∞^b has a unique minimizer, $y_\infty^b > 0$, which is continuously differentiable—and hence Lipschitz continuous—in b, with $\frac{\partial}{\partial b} y_\infty^b = -\ddot{\Lambda}_\infty(y_\infty^b)^{-1}$.*

Together, Lemmas 1 and 2 imply that $\ddot{\Lambda}_\infty(y_\infty^b)$—the Hessian matrix of Λ_∞^b at its minimum—is continuous in b in a neighborhood of β. Accordingly, $\{\omega_i^b\}_{i\in[m]}$ and $\{\sigma_i^b\}_{i\in[m]}$ are likewise continuous in b, where ω_i^b is an eigenvector of $\ddot{\Lambda}_\infty(y_\infty^b)$ with eigenvalue σ_i^b. Furthermore, since $\ddot{\Lambda}_\infty(y_\infty^b)$ is positive definite, we can take $\{\omega_i^b\}_{i\in[m]}$ to be orthonormal and take $\{\sigma_i^b\}_{i\in[m]}$ to be real numbers that satisfy $\sigma_1^b \geq \cdots \geq \sigma_m^b > 0$ (provided that b is sufficiently close to β).

Lemma 1 also implies that

$$\dot{\Lambda}_\infty^b(y) \equiv \frac{\partial}{\partial y}\Lambda_\infty^b(y) = b - \mathrm{E}(\mathbb{1}\{\Delta_1(y) > 0\}a_1) \tag{24}$$

exists and is continuous in y a neighborhood of y_∞^β. Unfortunately, the finite analog, Λ_t^b, is not always differentiable, but when it is, its gradient equals subgradient

$$\dot{\Lambda}_t^b(y) \equiv b - \sum_{s=1}^{t} \mathbb{1}\{\Delta_s(y) > 0\}a_s/t. \tag{25}$$

Our model is now fully characterized. Thus, we are now ready for the primary results.

Theorem 1. *The optimal policy of the online linear program without distribution learning yields an expected regret that grows at no more than a $\log n$ asymptotic rate: $\mathrm{E}(R_n) = O(\log n)$ as $n \to \infty$ when distribution μ is known to the decision maker.*

Theorem 2. *The optimal policy of the online linear program with distribution learning yields an expected regret that grows at no more than a $\log n$ asymptotic rate: $\mathrm{E}(R_n) = O(\log n)$ as $n \to \infty$ when distribution μ is unknown to the decision maker.*

Theorem 3. *The optimal policy of the online linear program without distribution learning yields an expected regret that grows at no less than a $\log n$ asymptotic rate: $\mathrm{E}(R_n) = \Omega(\log n)$ as $n \to \infty$ when distribution μ is known to the decision maker.*

Corollary 1. *The optimal policy of the online linear program with distribution learning yields an expected regret that grows at no less than a $\log n$ asymptotic rate: $\mathrm{E}(R_n) = \Omega(\log n)$ as $n \to \infty$ when distribution μ is unknown to the decision maker.*

Because knowing μ will not decrease your regret, Corollary 1 follows immediately from Theorem 3, and Theorem 1 follows immediately from Theorem 2. However, I do not call Theorem 1 a corollary because I provide an independent proof for it. Indeed, I will use the proof of Theorem 1 as a stepping stone to the proof of Theorem 2.

Also, the single-dimensional results of Section 3 and Lueker (1998) imply none of the previous multidimensional results—The previous findings establish that an online linear program *can* exhibit $\log n$ regret but not that it *must* do so. Naturally, the regret could be larger for the "harder" online linear program, but the regret can also be *smaller*. Indeed, although an additional restriction cannot increase the objective value, it can decrease the regret by burdening the offline problem more than the online problem. For instance, Example 2 illustrates that adding a second constraint to the multisecretary problem can reduce its expected regret from $\Theta(\log n)$ to $o(1)$, and Example 3 illustrates that adding a second constraint to two copies of the stochastic knapsack problem of Arlotto and Xie (2020) can intertwine the problem instances in a manner that reduces their combined regrets from $\Theta(\log n)$ to $O(1)$. Hence, some constraints negate the $\log n$ regret rate; I must prove that all such negating constraints violate our assumptions.

Example 2. Consider the multisecretary problem of Section 3.1, but with an additional payroll budget constraint: Now, in addition to the $n\beta$ available positions, you also start with $n\beta/2$ dollars, which you use to pay your workforce. The period t applicant commands wage u_t, so the applicants all yield the same bang for the buck. By design, you will almost certainly run out of money before you fill all the positions when n is large, both under the optimal online and offline policies. Hence, only your payroll budget constraint is relevant as $n \to \infty$. However, you will never regret how you spend this budget because every dollar yields the same marginal utility. Accordingly, the regret must go to zero, almost surely, as $n \to \infty$. □

Example 3. Suppose you encounter a stream of n identical items. Since they are all the same, a stochastic knapsack problem involving these n items would yield zero regret. Now, imagine that each item consists of two components, A and B, each valued at one dollar. Assume the volume of the A component is a uniform random variable, and the volume of the B component is one minus the volume of the A component, making it also a uniform random variable. Also, suppose you have two backpacks: bag A for storing A components and bag B for storing B components. If both backpacks have a capacity of $n/4$, then you will face A and B stochastic knapsack problems, both of which are expected to yield $\Theta(\log n)$ regret, according to the results of Lueker (1998) and Arlotto and Xie (2020). Now, introduce a constraint that prohibits packing only one component from an item. Compelling you to pack both components or neither effectively reverts the problem to the scenario in which all items have equal value. For example, an item with an attractive A component will have a commensurately unattractive B component. In this case, the regret will be proportional to the unused space in one backpack when the other is filled, a quantity that has $O(1)$ expectation under Algorithm 2, per Corollary 4. Adding the restriction lowers the regret by replacing the component-level selection with item-level selection. Although there is plenty of variation in component valuations, there is essentially no variation in item valuations due to the perfectly negative correlation between the components' volumes. □

4.2. Dual Convergence Results

Everything boils down to shadow prices, so we can only make progress once we understand how y_t^b converges to y_∞^b. I will thus begin the analysis by presenting four propositions that crisply characterize the shadow prices' limiting behavior.

Proposition 3. *There exists* $\delta > 0$ *such that* $\sqrt{t}(y_t^b - y_\infty^b) \xrightarrow{d} \mathcal{N}(0, \Sigma^b)$ *for all* $b \in B_\delta(\beta)$, *where* $\Sigma^b \equiv \ddot{\Lambda}_\infty(y_\infty^b)^{-1}\text{Cov}(\mathbb{1}\{\Delta_1\ (y_\infty^b) > 0\}a_1)\ddot{\Lambda}_\infty(y_\infty^b)^{-1}$ *is full rank and continuous in* $b \in B_\delta(\beta)$.

In the previous proposition, $B_\delta(\beta)$ denotes the ball of radius δ about β. However, do not dwell on these technical δ balls; instead, direct your attention to this: $\sqrt{t}(y_t^b - y_\infty^b) \xrightarrow{d} \mathcal{N}(0, \Sigma^b)$. It's hard to believe, but it seems the basic fact that the shadow prices of a stochastic linear program converge to a multivariate normal was previously unknown.

Unfortunately, this proposition proved less helpful than I had hoped because the rate of convergence could depend on b—that is, the magnitude of t required to ensure that $\sqrt{t}(y_t^b - y_\infty^b) \approx \mathcal{N}(0, \Sigma^b)$ could be unbounded in any neighborhood of β. Unfortunately, this will not do because I will need to invoke my convergence results at a random value of b_t. Hence, rather than Proposition 3, I will use the following results, which control the limiting shadow prices *uniformly* across $b \in B_\delta(\beta)$.

Proposition 4. *There exists* $\delta > 0$ *such that* $\text{E}(\sup_{b \in B_\delta(\beta)} \|y_t^b - y_\infty^b\|^2) = O(1/t)$.

Proposition 5. *There exists* $\delta > 0$ *such that* $\text{E}(\inf_{b \in B_\delta(\beta)} \|y_t^b - y_\infty^b\|^2) = \Omega(1/t)$.

Corollary 2. *There exists* $\delta > 0$ *such that the covariance matrix of* y_t^b *has a* $\Theta(1/t)$ *spectral norm, for all* $b \in B_\delta(\beta)$.

Positioning the $\sup_{b \in B_\delta(\beta)}$ and $\inf_{b \in B_\delta(\beta)}$ terms inside of the expectations makes these results especially strong. We'll need this extra strength to bound the regret when distribution μ is unknown, in which case shadow prices and inventory vectors become tangled.[4]

Proposition 4 is a stronger version of the first theorem of Li and Ye (2022), which states that $\text{E}(\|y_t^b - y_\infty^b\|^2) = O((\log\log t)/t)$. I had to shave off the repeated logarithms to derive a sharp $\log n$ upper bound. I did so with a new approach. I first bounded the difference between y_t^b and y_∞^b with the difference between the limiting gradient, $\dot{\Lambda}_\infty^b(\cdot)$, and its finite analog, $\dot{\Lambda}_t^b(\cdot)$, evaluated at the shadow price midway point, $\hat{y}_t^b \equiv (y_t^b + y_\infty^b)/2$. However, $\hat{y}_t^b$ is difficult to work with, so I then bounded the expected value of $\|\dot{\Lambda}_t^b(\hat{y}_t^b) - \dot{\Lambda}_\infty^b(\hat{y}_t^b)\|^2$

with the expected value of $\sup_{y \in B_{2\epsilon}(y^\beta_\infty)} \|\dot{\Lambda}^b_t(y) - \dot{\Lambda}^b_\infty(y)\|^2$. Finally, I bounded the expected value of this supremum with a classic empirical processes result.

I also used empirical processes to prove Proposition 5, which will permit the corresponding lower regret bound. Specifically, I establish this result by showing that $\sqrt{t}(y^b_t - y^b_t)$ is near $\gamma \in \mathbb{R}^m$ if $\sqrt{t}(\dot{\Lambda}^\beta_t(y) - \dot{\Lambda}^\beta_\infty(y))$ is near $\ddot{\Lambda}_\infty(y^\beta_\infty)\gamma$ for all y in a neighborhood of y^β_∞, and this latter condition holds because the mapping $(j, y) \longmapsto \sqrt{t}e'_j(\dot{\Lambda}^b_t(y) - \dot{\Lambda}^b_\infty(y))$ converges to a sufficiently well-behaved Gaussian process, indexed by y and j.

While the previous propositions establish that our shadow price variances fall linearly with t, the following proposition and corollary show that their tails fall exponentially with t.

Proposition 6. *For all $p \geq 0$, there exist $\delta, C > 0$ such that $\mathrm{E}(\sup_{b \in B_\delta(\beta)} \mathbb{1}\{y^b_t \notin B_\epsilon(y^b_\infty)\} \|y^b_t - y^b_\infty\|^p) \leq \exp(-C\epsilon^2 t)$ for all sufficiently small $\epsilon > 0$ and sufficiently large t.*

Corollary 3. *There exist $\delta, C > 0$ such that $\Pr(\sup_{b \in B_\delta(\beta)} \|y^b_t - y^b_\infty\| > \epsilon) \leq \exp(-C\epsilon^2 t)$ for all sufficiently small $\epsilon > 0$ and sufficiently large t.*

Whereas the third proposition of Li and Ye (2022) establishes a concentration of measure for random subgradient $\dot{\Lambda}^b_t(y^b_\infty)$, Corollary 3 establishes a concentration of measure for random shadow price y^b_t. This latter result is far harder to prove because y^b_t is not a sum of i.i.d. random variables, unlike $\dot{\Lambda}^b_t(y^b_\infty)$. I establish the shadow price concentration of measure by projecting the shadow prices onto the subgradient of the dual value function at many points. These projections yield inequalities that describe a small box around y^b_t and y^b_∞. This box has random faces, so its walls do not meet at 90-degree angles. Still, the angles exhibit a concentration of measure, so the probability that the wall's fluctuations undermine the box's integrity falls exponentially fast with t. More specifically, because y^b_t is a minimizer, it must satisfy subgradient constraint $(y^b_t - y^b_\infty - \eta k \omega^b_j)' \dot{\Lambda}^b_t(y^b_\infty + \eta k \omega^b_j) \leq 0$ for all $j \in [m]$ and $k \in \{-1, 1\}$. These inequalities position y^b_t in the intersection of $2m$ half-spaces. Unfortunately, these half-spaces are random because $\dot{\Lambda}^b_t$ is stochastic. However, $\dot{\Lambda}^b_t(y^b_\infty + \eta k \omega^b_j)$ concentrates about $\eta k \sigma^b_j \omega^b_j$, so the set of points that satisfy $(y^b_t - y^b_\infty - \eta k \omega^b_j)' \dot{\Lambda}^b_t(y^b_\infty + \eta k \omega^b_j) \leq 0$ for all $j \in [m]$ and $k \in \{-1, 1\}$ resemble those that satisfy $(y^b_t - y^b_\infty - \eta k \omega^b_j)' \eta k \sigma^b_j \omega^b_j \leq 0$ for all $j \in [m]$ and $k \in \{-1, 1\}$. This latter set of points forms a perfect cube around y^b_∞. Hence, our initial subgradient constraints situate y^b_t in a wonky cube about y^b_∞, with off-kilter faces.

4.3. Upper Bound with Known Distribution

I will now prove Theorem 1 by showing that Algorithm 2 honors its $O(\log n)$ bound. I will begin by showing that the inventory levels follow a martingale under this algorithm. This martingale property concentrates the distribution of b_t to the small neighborhood of β for which our lemmas apply. Next, I will express the values obtained under Algorithm 2 and those obtained under the optimal algorithm with Bellman-style recurrence relations. I will then combine these recurrence relations to create an analogous regret recurrence relation, which I will unravel to create a corresponding regret recurrence relation. Finally, I will bound this decomposition's myopic regret with our shadow price convergence results.

Algorithm 2 satisfies the period t customer if and only if (i) there is inventory enough to do so (i.e., $tb_t \geq a_t$) and (ii) the customer has positive surplus utility under the fluid-approximation shadow prices (i.e., $\Delta_t(y^{b_t}_\infty) > 0$). Under this policy, the inventory vector follows a martingale: for $t > 1$, $b_t \geq \alpha/t$, and b_t sufficiently close to β, we have

$$\begin{aligned}
\mathrm{E}(b_{t-1} \mid b_t) &= \mathrm{E}(\psi^{b_t}_t(x_t a_t) \mid b_t) \\
&= (tb_t - \mathrm{E}(\mathbb{1}\{\Delta_t(y^{b_t}_\infty) > 0\} a_t \mid b_t))/(t-1) \\
&= (tb_t - b_t + \dot{\Lambda}^{b_t}_\infty(y^{b_t}_\infty))/(t-1) \\
&= b_t + \dot{\Lambda}^{b_t}_\infty(y^{b_t}_\infty)/(t-1) \\
&= b_t.
\end{aligned}$$

This martingale property implies the following, via the Azuma–Hoeffding inequality.

Lemma 3. *The inventory vector abides by a concentration of measure under Algorithm 2: For all $\delta > 0$, there exists $C > 0$ such that $\Pr(b_t \notin B_\delta(\beta)) \leq \exp(-Ct)$, for all sufficiently large t.*

This result is stronger than one Li and Ye (2022) used. To see this, let $\tau(\delta)$ represent the first time that b_t leaves $B_\delta(\beta)$:

$$\tau(\delta) \equiv \begin{cases} 0 & \{b_t \mid t \in [n]\} \subset B_\delta(\beta), \\ \max\{t \mid b_t \notin B_\delta(\beta)\} & \text{otherwise.} \end{cases} \tag{26}$$

Li and Ye proved that their algorithm yields $\mathrm{E}(\tau(\delta)) = O(\log n \log \log n)$—that is, that it constrains the resource vector for all but the last $O(\log n \log \log n)$ periods. However, I could not use this $O(\log n \log \log n)$ result to derive a $O(\log n)$ regret bound, so I had to sharpen their finding. As the following corollary explains, I managed to tighten it to $O(1)$.

Corollary 4. *The time remaining after the resource vector leaves a given neighborhood of β is asymptotically independent of n, under Algorithm 2: $\mathrm{E}(\tau(\delta)) = O(1)$ as $n \to \infty$, for all $\delta > 0$.*

Algorithm 2 (Martingale Controler)

```
1. input $n$, $\beta$, $\{u_t\}_{t=1}^n$, $\{a_t\}_{t=1}^n$, $\mu$
2. initialize $b_n := \beta$
```

```
3. for t from n to 1 do
    (a) set $x_t := \mathbb{1}\{\Delta_t(y_\infty^{b_t}) > 0\}\mathbb{1}\{tb_t \ge a_t\}$
    (b) set $b_{t-1} := \psi_t^{b_t}(x_t a_t)$
4. end for
5. output $\{x_t\}_{t=1}^n$
```

Since the optimal policy is no worse than our martingale policy, we have

$$\mathrm{E}(\hat{R}_n) \ge \mathrm{E}(R_n), \tag{27}$$

where $\hat{R}_t \equiv \overline{V}_t^{b_t} - \hat{v}_t$ and $\hat{v}_t$ is the value collected by Algorithm 2 after period t:

$$\hat{v}_t \equiv \mathbb{1}\{\Delta_t(y_\infty^{b_t}) > 0\}\mathbb{1}\{tb_t \ge a_t\}u_t + \hat{v}_{t-1}, \quad \text{and} \quad \hat{v}_0 \equiv 0. \tag{28}$$

Accordingly, it will suffice to show that $\mathrm{E}(\hat{R}_n) = O(\log n)$. As in the multisecretary case, I will bound this benchmark regret by decomposing it into a sum of myopic regrets.

Lemma 4. *The benchmark regret under Algorithm 2 can be upper bounded by a sum of approximate myopic regrets: There exists sufficiently small $\delta > 0$ such that*

$$\hat{R}_n \le \sum_{t=1}^n r_t,$$

$$\text{where} \quad r_t \equiv \mathbb{1}\{b_t \notin B_{\delta/2}(\beta)\}\sum_{s=1}^t u_s + \mathbb{1}\{b_t \in B_{\delta/2}(\beta)\}\mathbb{1}\{\Delta_t(y_\infty^{b_t}) > 0\}\Delta_t(y_{t-1}^{b_{t-1}})^- + \mathbb{1}\{b_t \in B_{\delta/2}(\beta)\}\mathbb{1}\{\Delta_t(y_\infty^{b_t}) \le 0\}\Delta_t(y_{t-1}^{b_{t-1}})^+.$$

The indicator variables in the definition of r_t characterize whether Algorithm 2 specifies satisfying the period t customer and whether b_t lies in the $(\delta/2)$-ball of β. (I use the $(\delta/2)$-ball rather than the δ-ball, because $b_t \in B_{\delta/2}(\beta)$ implies $b_{t-1} \in B_\delta(\beta)$ when t is large.)

Finally, combining the preceding lemma with the following lemma yields Theorem 1.

Lemma 5. *The approximate period-t myopic regret under Algorithm 2 is $O(1/t)$ in expectation: there exists $C > 0$ such that $r_t \le C/t$.*

To control the first term of the myopic regret, I use the fact that $\mathrm{E}(\sum_{s=1}^t u_s)$ increases linearly in t, whereas $\Pr(b_t \notin B_{\delta/2}(\beta))$ falls exponentially, by Lemma 3. To control the second term, I bound $\mathrm{E}(\mathbb{1}\{\Delta_t(y_\infty^{b_t}) > 0\}\Delta_t(y_{t-1}^{b_{t-1}})^-)$ in terms of $\mathrm{E}(\sup_{b\in B_\delta(\beta)}\|y_{t-1}^b - y_\infty^b\|^2)$, $\mathrm{E}(\sup_{b\in B_\delta(\beta)}\mathbb{1}\{y_{t-1}^b \notin B_\epsilon(y_\infty^b)\}\|y_{t-1}^b - y_\infty^b\|)$, and $\Pr(\sup_{b\in B_\delta(\beta)}\|y_t^b - y_\infty^b\| > \epsilon)$, and then apply Propositions 4 and 6 and Corollary 3. Finally, I control the third term in a similar manner.

4.4. Upper Bound with Unknown Distribution

I will now prove Theorem 2 by showing that Algorithm 3 honors its $O(\log n)$ bound. The only difference between Algorithms 2 and 3 is that the former uses limiting shadow price $y_\infty^{b_t}$, which requires knowledge of μ, whereas the latter uses *look-back shadow price* $\underset{\leftarrow}{y}_t^{b_t}$, which is an estimate of $y_\infty^{b_t}$ given the data observed up until period $t+1$. More specifically $\underset{\leftarrow}{y}_t^{b_t}$ is a minimizer of the backward-looking problem

$$\underset{\leftarrow}{\Delta}_t^b(y) \equiv b'y + \sum_{s=t+1}^n \Delta_s(y)^+/(n-t). \tag{29}$$

Algorithm 3 incorporates learning, as shadow price estimate $\underset{\leftarrow}{y}_t^{b_t}$ starts hopelessly crude and ends finely tuned.

Our shadow price convergence results hold for look-back shadow prices but with $(n-t)$-period scaling rather than t-period scaling. For example, Proposition 4 implies that $\mathrm{E}(\mathbb{1}\{b_t \in B_\delta(\beta)\}\|\underset{\leftarrow}{y}_t^{b_t} - y_\infty^{b_t}\|^2) = O(1/(n-t))$. (This would *not* be the case if the proposition positioned the $\sup_{b\in B_\delta(\beta)}$ term outside of the expectation because b_t correlates with the random map $b \longmapsto \underset{\leftarrow}{y}_t^b$.)

Algorithm 3 (Estimate then Forecast)

```
1. input $n$, $\beta$, $\{u_t\}_{t=1}^n$, $\{a_t\}_{t=1}^n$
2. initialize $b_n := \beta$
3. for t from n to 1 do
    (a) set $x_t := \mathbb{1}\{\Delta_t(\underset{\leftarrow}{y}_t^{b_t}) > 0\}\mathbb{1}\{tb_t \ge a_t\}$
    (b) set $b_{t-1} := \psi_t^{b_t}(x_t a_t)$
4. end for
5. output $\{x_t\}_{t=1}^n$
```

Although the inventory vector does not follow a martingale under Algorithm 3, as it does under Algorithm 2, we can still control its trajectory for all but $O(1)$ periods, as the following results establish.

Lemma 6. *The inventory vector abides by a concentration of measure under Algorithm 3: For all $\delta > 0$, there exists $C > 0$ such that $\Pr(b_t \notin B_\delta(\beta)) \le \exp(-C\min(t, \sqrt{n}))$, for all sufficiently large $t \le n$.*

Corollary 5. *The time remaining after the resource vector leaves a given neighborhood of β is asymptotically independent of n under Algorithm 3: $\mathrm{E}(\tau(\delta)) = O(1)$ as $n \longrightarrow \infty$, for all $\delta > 0$.*

The critical insight underlying Lemma 6 is that b_t cannot escape $B_{\delta/2}(\beta)$ in less than $\Omega(n)$ time and hence without first generating an $\Omega(n)$-sized sample of training data. This means that by the time the $\{b_t\}_{t=n}^1$ process has made it halfway out of $B_\delta(\beta)$—that is, departed $B_{\delta/2}(\beta)$—our look-back shadow prices are accurate enough to (almost) guarantee that it cannot traverse the second half. This property enables us to restrict attention to the periods with accurate look-back shadow prices (i.e., periods after time $\tau(\delta/2)$).

However, controlling the evolution of $\{b_t\}_{t=n}^1$ is difficult even when look-back shadow prices are accurate. The problem is that, although b_t is independent of the mapping $b \longmapsto y_t^b$, it is not independent of the mapping

$b \longmapsto \underset{\leftarrow}{y}_t^b$. Indeed, the inventory vectors and look-back shadow prices intertwine in a complex dance. To extricate b_t from this pas de deux, I decompose it into three parts: $b_{\tau(\delta/2)+1}$, $\sum_{s=t}^{\tau(\delta/2)} b_s - \mathrm{E}(b_s \mid b_{s+1})$, and $\sum_{s=t}^{\tau(\delta/2)} \mathrm{E}(b_s \mid b_{s+1}) - b_{s+1}$. By definition, the first part is within $\delta/2$ of β. The second part follows a martingale and thus concentrates around zero. The third part is small, provided that $\underset{\leftarrow}{y}_s^{b_s}$ is near $y_\infty^{b_s}$, for all $s \in \{t+1, \ldots, \tau(\delta/2)+1\}$. Crucially, $\overset{\leftarrow}{\tau}(\delta/2)$ will be small enough to ensure that this holds with high probability, provided that b_s is near β for all $s \in \{t+1, \ldots, \tau(\delta/2)+1\}$. Thus, I can inductively establish the result: b_s being near β for $s \in \{t+1, \ldots, \tau(\delta/2)+1\}$ implies that $\underset{\leftarrow}{y}_s^{b_s}$ is near $y_\infty^{b_s}$ for $s \in \{t+1, \ldots, \tau(\delta/2)+1\}$, which implies that $\sum_{s=t}^{\tau(\delta/2)} \mathrm{E}(b_s \mid b_{s+1}) - b_{s+1}$ is small, which implies that b_t is near β.

Having reigned in our inventory vectors, we are now ready to decompose regret benchmark

$$\hat{R}_t \equiv \overline{V}_t^{b_t} - \hat{v}_t, \tag{30}$$

where $\hat{v}_t$ now denotes the value collected after period t under Algorithm 3:

$$\hat{v}_t \equiv \mathbb{1}\{\Delta_t(\underset{\leftarrow}{y}_t^{b_t}) > 0\}\mathbb{1}\{tb_t \geq a_t\}u_t + \hat{v}_{t-1}, \text{ and } \hat{v}_0 \equiv 0. \tag{31}$$

Lemma 7. *The benchmark regret under Algorithm 3 can be upper bounded by a sum of approximate myopic regrets: There exists sufficiently small $\delta > 0$ such that*

$$\hat{R}_n \leq \sum_{t=1}^{n} r_t,$$

$$\textit{where} \quad r_t \equiv \mathbb{1}\{b_t \notin B_{\delta/2}(\beta)\}\sum_{s=1}^{t} u_s + \mathbb{1}\{b_t \in B_{\delta/2}(\beta)\}\mathbb{1}\{\Delta_t(\underset{\leftarrow}{y}_t^{b_t}) > 0\}\Delta_t(y_{t-1}^{b_{t-1}})^- + \mathbb{1}\{b_t \in B_{\delta/2}(\beta)\}\mathbb{1}\{\Delta_t(\underset{\leftarrow}{y}_t^{b_t}) \leq 0\}\Delta_t(y_{t-1}^{b_{t-1}})^+.$$

Combining the preceding lemma with the following lemma yields Theorem 2.

Lemma 8. *The approximate period t myopic regret under Algorithm 3 is $O(1/t) + O(1/(n-t))$ in expectation: There exists $C > 0$ such that $\mathrm{E}(r_t) \leq C/t + C/(n-t)$, for all $n \in \mathbb{N}$ and $t \leq n$.*

This lemma is the same as Lemma 5, except now both the $O(1/\sqrt{t})$ errors between $y_t^{b_t}$ and $y_\infty^{b_t}$ and the $O(1/\sqrt{n-t})$ errors between $\underset{\leftarrow}{y}_t^{b_t}$ and $y_\infty^{b_t}$ contribute to your regret.

4.5. Lower Bound with Known Distribution

We will now prove Theorem 3. To reiterate, the results of Section 3.3 do not make this analysis redundant: Whereas we previously established $\Omega(\log n)$ regret for one specific OLP—the multisecretary problem—we now establish $\Omega(\log n)$ regret for *all* OLPs. In other words, Section 3.3 illustrates that the expected regret *can* grow like $\Omega(\log n)$, and this section proves that it *must* grow like $\Omega(\log n)$ (see the discussion at the end of Section 4.1).

We will establish a universal $\Omega(\log n)$ regret rate by retooling the methodology developed in Section 4.3 for a lower bound. For example, the lower-bounding decomposition will depend on $\Delta_t(y_{t-1}^{\psi_t^{b_t}(0)})^-$ and $\Delta_t(y_{t-1}^{\psi_t^{b_t}(a_t)})^+$ (as opposed to $\Delta_t(y_{t-1}^{\psi_t^{b_t}(a_t)})^-$ and $\Delta_t(y_{t-1}^{\psi_t^{b_t}(0)})^+$); the lower-bounding version of Lemma 3 will ensure the proximity of b_t and β under the optimal algorithm (as opposed to Algorithm 2); and the lower-bounding version of Lemma 5 will establish that the expected myopic regret is $\Omega(1/t)$ (as opposed to $O(1/t)$).

In this section, $\{b_t\}_{t=n}^1$ will characterize the inventory levels that correspond to the optimal actions specified in Line (14):

$$b_n = \beta$$
$$\text{and} \quad b_{t-1} = \psi_t^{b_t}(\pi_t^{b_t} a_t). \tag{32}$$

Unfortunately, we now have little control over $\{b_t\}_{t=n}^1$, because the optimal policy is unknown. Nevertheless, we can still situate b_t near β for a substantial time interval.

Lemma 9. *The inventory vector tends to lie near β under the optimal policy for most of the second half of the horizon: For all $\delta > 0$, if n is sufficiently large then $n^{3/4} \leq t \leq n/2$ implies $\Pr(b_t \notin B_{\delta/2}(\beta)) \leq n^{-1/2}$.*

This lemma was the hardest result in this article to prove because the optimal policy is opaque. Generalizing the technique developed in Section 3.3, I argue that the regret incurred when b_t strays from β is at least as large as the value sacrificed when we chop the linear program into two separate problems: one with horizon t and endowment tb_t and the other with horizon $n-t$ and endowment $n\beta - tb_t$. The concavity of $\overline{V}_t^b$ in b ensures that this division is costly when b_t meaningfully differs from β.

As before, we will benchmark against the offline linear program, Line (18), rather than the offline integer program, Line (16). The following result will enable us to do so:

$$\mathrm{E}(\hat{R}_n) = \mathrm{E}(R_n) + O(1),$$
$$\text{where} \quad \hat{R}_t \equiv \overline{V}_t^{b_t} - v_t^{b_t}. \tag{33}$$

The first line holds because the linear program has a solution that partially satisfies at most m customers, and thus the integer program must derive at least as much value from resource endowment β as the linear program does from resource endowment $\beta - m\alpha/n$: $V_n^\beta \geq \overline{V}_n^{\beta - m\alpha/n}$. Because the shadow price decreases in the inventory level, this implies that $V_n^\beta \geq \overline{V}_n^\beta - m\alpha' y_n^{\beta - m\alpha/n}$, and hence that $R_n \geq \overline{V}_n^\beta - v_n^\beta - m\alpha' y_n^{\beta - m\alpha/n}$. Finally, Proposition 4 indicates that $\mathrm{E}(y_n^{\beta - m\alpha/n}) = O(1)$ as $n \longrightarrow \infty$, which establishes the result.

As before, I will now bound the benchmark regret with a sum of myopic regrets.

Lemma 10. *The benchmark regret under the optimal algorithm can be lower bounded by a sum of approximate myopic regrets: there exists sufficiently small $\delta > 0$ such that*

$$\hat{R}_n \geq \sum_{t=\lceil 2\|\alpha\|/\delta\rceil}^{n} r_t,$$

$$\textit{where} \quad r_t \equiv \mathbb{1}\{b_t \in B_{\delta/2}(\beta)\}(\pi_t^{b_t}\Delta_t(y_{t-1}^{\psi_t^{b_t}(0)})^- + (1-\pi_t^{b_t})\Delta_t(y_{t-1}^{\psi_t^{b_t}(a_t)})^+).$$

Combining the preceding lemma with the following lemma yields Theorem 3.

Lemma 11. *The approximate period t myopic regret under the optimal algorithm is $\Omega(1/t)$ in expectation for most of the second half of the horizon: There exists $C > 0$ such that $\mathrm{E}(r_t) \geq C/t$, for all sufficiently large t that satisfies $n^{3/4} \leq t \leq n/2$.*

To establish this last result, I show that if $b_t \in B_{\delta/2}(\beta)$—which happens with high probability, by Lemma 9—then $\sqrt{t}(y_t^{b_t} - y_\infty^{b_t})$ could be near *any* $\gamma \in \mathbb{R}^m$. Accordingly, both type I errors—rejecting customers that you should have satisfied—and type II errors—satisfying customers that you should have rejected—are unavoidable because the shadow price can always be larger or smaller than anticipated. Specifically, I show that there is at least a $\Omega(1/\sqrt{t})$ chance that both the expected type I and type II errors are $\Omega(1/\sqrt{t})$.

5. Conclusion

I have already said everything I want to, so I will use this space to recapitulate this work's primary contributions.

- I develop a new methodology for regulating the dual variables of an online linear program: use the convexity of the dual problem to bound the shadow prices in terms of the subgradient of the dual value function and then bound this subgradient by casting it as an empirical process (for which there are many established results). The technique is versatile —I used it five different times:
  1. In the proof of Proposition 3, I implicitly use the technique as I leverage an M-estimator result that establishes the asymptotic normality of the dual solution by casting the random dual objective function as an empirical process.
  2. In the proof of Proposition 4, I bound $\mathbb{1}\{y_t^b \notin B_\epsilon(y_\infty^b)\}\|y_t^b - y_\infty^b\|^2$ with a fixed multiple of $\|\dot{\Lambda}_t^b((y_t^b + y_\infty^b)/2) - \dot{\Lambda}_\infty^b((y_t^b + y_\infty^b)/2)\|^2$, which I bound with $\sup_{y\in B_\epsilon(y_\infty^b)}\|\dot{\Lambda}_t^b(y) - \dot{\Lambda}_\infty^b(y)\|^2$, which in turn I bound with an empirical processes result.
  3. In the proof of Proposition 6, I bound the probability that $\sup_{b\in B_\delta(\beta)}\|y_t^b - y_\infty^b\|$ is large with the probability that $\sup_{b\in B_\delta(\beta)}\|\dot{\Lambda}_t^b(y_\infty^b + \eta k\omega_j^b) - \dot{\Lambda}_\infty^b(y_\infty^b + \eta k\omega_j^b)\|$ is large, which I bound with the probability that $\sup_{y\in B_\nu(y_\infty^\beta)}\|\dot{\Lambda}_t^\beta(y) - \dot{\Lambda}_\infty^\beta(y)\|$ is large, which in turn I bound with an empirical processes result.
  4. In the lemma that proves Proposition 5, I lower bound the probability that $\sup_{b\in B_\delta(\beta)}\|\sqrt{t}(y_t^b - y_\infty^b) - \gamma\|$ is small with the probability that $\sup_{b\in B_\delta(\beta)}\|\sqrt{t}(\dot{\Lambda}_t^\beta(y_\infty^b + (\gamma + \eta k\omega_j^b)/\sqrt{t}) - \dot{\Lambda}_\infty^\beta(y_\infty^b + (\gamma + \eta k\omega_j^b)/\sqrt{t})) + \ddot{\Lambda}_\infty(y_\infty^\beta)\gamma\|$ is small, which I bound with the expected value of $\sup_{y\in B_\nu(y_\infty^\beta)}\|\dot{\Lambda}_t^\beta(y) - \dot{\Lambda}_\infty^\beta(y) - (\dot{\Lambda}_t^\beta(y_\infty^\beta) - \dot{\Lambda}_\infty^\beta(y_\infty^\beta))\|^2$, which in turn I bound with an empirical processes result.
  5. In the proof of Lemma 9, I lower bound the cost of an additional restriction that mandates $b_t = \beta + \xi$, with $|\Lambda_t^\beta(y_n^\beta) - \Lambda_\infty^\beta(y_n^\beta) - \Lambda_t^{\beta+\zeta}(y_t^{\beta+\zeta}) + \Lambda_\infty^{\beta+\zeta}(y_t^{\beta+\zeta})|$, which I bound with $\sup_{y,\overline{y}\in\Omega}|\Lambda_t^b(y) - \Lambda_\infty^b(y) - \Lambda_t^{\overline{b}}(\overline{y}) + \Lambda_\infty^{\overline{b}}(\overline{y})|$, which in turn I bound with an empirical processes result.
- I use my new empirical processes methodology to precisely characterize the convergence of dual variable y_t^b to its deterministic limit, y_∞^b. Specifically, I show that under weak conditions
  1. $\sqrt{t}(y_t^b - y_\infty^b)$ converges to a multivariate normal for all $b \in B_\delta(\beta)$,
  2. $\mathrm{E}(\sup_{b\in B_\delta(\beta)}\mathbb{1}\{y_t^b \notin B_\epsilon(y_\infty^b)\}\|y_t^b - y_\infty^b\|^p)$ falls exponentially fast in t for all $p \geq 0$,
  3. $\mathrm{E}(\inf_{b\in B_\delta(\beta)}\|y_t^b - y_\infty^b\|^2) = \Omega(1/t)$, and
  4. $\mathrm{E}(\sup_{b\in B_\delta(\beta)}\|y_t^b - y_\infty^b\|^2) = O(1/t)$.

I cannot find any direct antecedents for the first three results in the literature, but the fourth one is a strengthened version of the finding of Li and Ye (2022) that $\mathrm{E}(\|y_t^b - y_\infty^b\|^2) = O((\log\log t)/t)$. In addition to removing the $\log\log t$ wiggle room, my bound also holds uniformly across $b \in B_\delta(\beta)$. Crucially, I position the $\sup_{b\in B_\delta(\beta)}$ inside the expectation, which makes the result especially strong. For example, I would not have been able to accommodate online learning had the E and $\sup_{b\in B_\delta(\beta)}$ operators been commuted. Precisely, if the supremum preceded the expectation, then I could have controlled the variance of look-back shadow price $\underleftarrow{y}_t^b$ for any fixed resource vector b, but I could not have done so for the realized b_t, as this random variable correlates with the random mapping $b \longmapsto \underleftarrow{y}_t^b$.

- I broaden the applicability of the compensated coupling scheme of Vera and Banerjee (2019) by devising new bounds on the trajectory of the inventory random walk, $\{b_t\}_{t=n}^1$. My delicate assumptions hold only in the δ-ball around β, so I cannot bound the regret without first proving that b_t resides in $B_\delta(\beta)$, with high probability. For the upper bound with known demand distribution, I show that following the certainty-equivalent principle—that is, estimating the unobserved shadow price, $y_t^{b_t}$, with its fluid approximation, $y_\infty^{b_t}$—constrains b_t to $B_\delta(\beta)$ for all but last $O(1)$ periods. For the upper bound with unknown demand distribution, I inductively prove that the error introduced by replacing deterministic shadow price estimate $y_\infty^{b_t}$ with stochastic shadow price estimate $\underleftarrow{y}_t^{b_t}$ falls fast enough to ensure an orderly $\{b_t\}_{t=n}^1$ walk. Finally, for the lower bound, I argue that the regret conditional on $b_t \notin B_\delta(\beta)$ must be at least as high as the cost of a $b_t \notin B_\delta(\beta)$ constraint imposed on the offline problem. I then lower bound the cost of this constraint to upper bound the probability that $b_t \notin B_\delta(\beta)$ under the optimal policy.

- I use my new control over shadow prices and inventory levels to extend $O(\log n)$ and $\Omega(\log n)$ regret bounds of Lueker (1998) to a multiresource setting. Rather than use Lueker's approach, I performed this generalization with compensated coupling, as bounding the value function across its entire domain would have been infeasible in higher dimensions (at least for me).
- I tighten the regret bound of Li and Ye (2022) for the online linear program (OLP) with online learning from $O(\log n \log \log n)$ to $O(\log n)$, and I provide a corresponding $\Omega(\log n)$ lower bound. Hence, I show that incorporating online learning—that is, the dynamic estimation of μ, the joint distribution of utility u_t, and resource consumption a_t—does not position a revenue management problem in a new difficulty class.

Acknowledgments

The author thanks Siddhartha Banerjee, Itai Gurvich, Ioannis Stamatopoulos, Neil Walton, and three anonymous reviewers for insightful feedback.

Appendix

Table A.1. List of Symbols

Symbol	Definition	Reference
$[x]$	The set $\{1,\ldots,x\}$	
$x \wedge y$	Vector with ith element $\min(x_i, y_i)$	
$x \vee y$	Vector with ith element is $\max(x_i, y_i)$	
x^+	$\max(0, x)$	
x^-	$\max(0, -x)$	
e_j	Unit vector indicating jth position	
ι	Vector of ones	
$\mathbb{1}\{\}$	Indicator function	
$B_\delta(b)$	Open ball with radius δ about b	
m	Number of resources to manage	
n	Number of time periods	
t	Generic time period	
b	Generic inventory vector, defined as total holdings divided by total time	
b_t	Period-t inventory vector associated with given algorithm	Algorithms 2 and 3 and line (32)
β	Initial inventory vector	
$\tau(\delta)$	First time inventory vector leaves $B_\delta(\beta)$	Line (26)
u_t	Utility received by satisfying period-t customer	
a_t	Resources consumed by satisfying period-t customer	
Δ_t	Surplus utility function	Line (21)
μ	Joint distribution of (u_t, a_t)	Assumption 1
α	Upper bound on a_t	Assumption 4
x_t	Period-t decision variable	
ψ_t^b	Function determining period-$(t-1)$ inventory vector	Line (13)
π_t^b	Optimal action	Line (14)
v_t^b	Online objective value	Line (15)
$\hat{v}_t$	Objective value associated with given algorithm	Lines (28) and (31)
V_t^b	Offline objective value	Line (16)
$\overline{V}_t^b$	Offline objective value with linear programming relaxation	Line (18)
R_n	Regret	Line (17)
$\hat{R}_t$	Benchmark regret associated with given algorithm	Lines (27), (30), and (33)
r_t	Approximate myopic regret associated with given algorithm	Lemmas 4, 7, and 10
Λ_t^b	Dual objective	Line (20)
$\underleftarrow{\Lambda}_t^b$	Look-back dual objective	Line (29)
$\dot{\Lambda}_t$	Dual objective subgradient	Line (25)
Λ_∞^b	Limiting dual objective	Line (23)
$\dot{\Lambda}_\infty$	Limiting dual gradient	Line (24)
$\ddot{\Lambda}_\infty$	Limiting dual Hessian	Lemma 1
ω_i^b	ith orthonormal eigenvector of limiting dual Hessian	After Lemma 2
σ_i^b	ith largest eigenvalue of limiting dual Hessian	After Lemma 2
y_t^b	Dual optimal solution	Line (22)
$\underleftarrow{y}_t^b$	Look-back dual optimal solution	Before line (29)
y_∞^b	Limiting dual optimal solution	Assumption 5 and Lemma 2
y	Generic dual solution	

Endnotes

[1] Switching from finite to continuous secretary valuations completely changes the mechanics of the model. With finite valuations, the probability of making a period t hiring mistake decreases exponentially in t, whereas the expected cost of such a mistake remains constant. Hence the total regret grows with n like $\sum_{t=1}^{n} \exp(-t) = \Theta(1)$. With continuous valuations, however, the probability of making a period t hiring mistake and the expected cost of such a mistake both decrease like $1/\sqrt{t}$. Hence, the total cost grows with n like $\sum_{t=1}^{n}(1/\sqrt{t}) \cdot (1/\sqrt{t}) = \Theta(\log n)$.

[2] To see that the expected regret with $n\beta$ initial open slots equals that with $n(1-\beta)$ initial open slots, we can re-express the problem of maximizing the capability of each of the $n\beta$ applicants you hire to maximizing one minus the capability of the $n(1-\beta)$ applicants you reject. However, one minus a uniform is also a uniform, so this mirror-image problem must yield mirror-image regrets.

[3] I make three minor changes to the online linear programming model: I impose additional nonnegativity constraints, $u_1, a_1 \geq 0$, I do not include constraints that are slack in the limit, and I use a cleaner version of the continuous value assumption, which I inherited from Lueker (1998). The first two modifications are trivial: Accommodating negative u_1 and a_1 would be simple because all that matters is the difference, $\Delta_1(y) = u_1 - a_1'y$. A simple concentration of measure argument establishes that a constraint that does not bind in the limit has only a $O(1)$ effect on the expected regret because the probability of it binding decreases exponentially fast in n. (I incorporated constraints that are slack in the limit in a previous version of the manuscript.) However, the third change is noteworthy because Assumption 6 is more straightforward and flexible. For example, this assumption permits unbounded shadow prices and hence unbounded utilities, and it extends the model to cover the specification of Arlotto and Xie (2020).

[4] After I posted the result, Jiang et al. (2024) independently developed a slightly weaker version of Proposition 4.

References

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

Arlotto A, Xie X (2020) Logarithmic regret in the dynamic and stochastic knapsack problem with equal rewards. *Stochastic Systems* 10(2):170–191.

Balseiro SR, Besbes O, Pizarro D (2023) Survey of dynamic resource-constrained reward collection problems: Unified model and analysis. *Oper. Res.*, ePub ahead of print May 9, https://doi.org/10.1287/opre.2023.2441.

Besbes O, Kanoria Y, Kumar A (2023) Dynamic resource allocation: Algorithmic design principles and spectrum of achievable performances. Preprint, submitted October 6, https://arxiv.org/abs/2205.09078.

Caley A (1875) Mathematical questions with their solutions. *Edu. Times* 23:18–19.

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

Jiang J, Zhang J (2020) Online resource allocation with stochastic resource consumption. Preprint, submitted December 14, https://arxiv.org/abs/2012.07933.

Jiang J, Ma W, Zhang J (2024) Degeneracy is OK: Logarithmic regret for network revenue management with indiscrete distributions. Preprint, submitted February 8, https://arxiv.org/abs/2210.07996.

Li X, Ye Y (2022) Online linear programming: Dual convergence, new algorithms, and regret bounds. *Oper. Res.* 70(5):2948–2966.

Lueker GS (1998) Average-case analysis of off-line and on-line knapsack problems. *J. Algorithms* 29:277–305.

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

Vera A, Banerjee S, Gurvich I (2019) Online allocation and pricing: Constant regret via bellman inequalities. *Oper. Res.* 69(3):821–840.

Wang Y, Wang H (2022) Constant regret resolving heuristics for price-based revenue management. *Oper. Res.* 5463(3):1–20.

Robert L. Bray is an associate professor of operations management at Northwestern University's Kellogg School of Management. Most of his research falls under the broad area of empirical operations management, but he also likes working on a good math problem when one presents itself.