An Improved Analysis of LP-based Control for Revenue Management

Guanting Chen†  Xiaocheng Li◊  Yinyu Ye‡

† Institute for Computational and Mathematical Engineering, Stanford University
◊ Imperial College Business School, Imperial College London
‡ Department of Management Science and Engineering, Stanford University

{guanting, chengli1, yinyu-ye}@stanford.edu

Abstract

In this paper, we study a class of revenue management problems where the decision maker aims to maximize the total revenue subject to budget constraints on multiple type of resources over a finite horizon. At each time, a new order/customer/bid is revealed with a request of some resource(s) and a reward, and the decision maker needs to either accept or reject the order. Upon the acceptance of the order, the resource request must be satisfied and the associated revenue (reward) can be collected. We consider a stochastic setting where all the orders are i.i.d. sampled, i.e., the reward-request pair at each time is drawn from an unknown distribution with finite support. The formulation contains many classic applications such as the quantity-based network revenue management problem and the Adwords problem. We focus on the classic LP-based adaptive algorithm and consider regret as the performance measure defined by the gap between the optimal objective value of the certainty-equivalent linear program (LP) and the expected revenue obtained by the online algorithm. Our contribution is two-fold: (i) when the underlying LP is nondegenerate, the algorithm achieves a problem-dependent regret upper bound that is independent of the horizon/number of time periods $T$; (ii) when the underlying LP is degenerate, the algorithm achieves a regret upper bound that scales on the order of $\sqrt{T \log T}$. To our knowledge, both results are new and improve the best existing bounds for the LP-based adaptive algorithm in the corresponding setting. We conclude with numerical experiments to further demonstrate our findings.

1 Introduction

Consider a revenue management problem over a horizon of $T$ time periods. The objective is to maximize the cumulative reward over the horizon subject to the budget constraints on $m$ types of resource. At each time period, a customer order arrives and it requests certain amount of each resource. We need to decide whether to accept or reject the customer order. Upon the acceptance of the order, we need to satisfy the resource request, and we will collect the revenue/reward associated with the order. From the resource viewpoint, we act as a market maker who allocates the resources among all the customer orders. The allocation decisions are made in an online manner and all the past decisions are irrevocable. In this paper, we study a stochastic model where the request-reward pair made by each customer order is assumed to be i.i.d. and follows an unknown distribution with finite support.
Specifically, the considered problem takes the following LP as its underlying form:

$$\max \sum_{t=1}^{T} r_t^\top x_t$$

s.t. $$\sum_{t=1}^{T} A_t x_t \leq B$$

$$1^\top x_t \leq 1, \ x_t \geq 0, \ t = 1, ..., T$$

where $$r_t = (r_{1t}, ..., r_{kt})^\top \in \mathbb{R}^k$$, $$A_t = (a_{1t}, ..., a_{kt}) \in \mathbb{R}^{m \times k}$$, and $$a_{st} = (a_{1st}, ..., a_{mst})^\top \in \mathbb{R}^m$$, for $$t = 1, ..., T$$ and $$s = 1, ..., k$$. The right-hand-side vector $$B = (B_1, ..., B_m)^\top$$ encapsulates the capacity for each resource. The decision variables are $$x = (x_1, ..., x_T)$$ where $$x_t = (x_{1t}, ..., x_{kt})^\top$$ for $$t = 1, ..., T$$. In an online setting, the parameters of the optimization problem (1) are revealed in an online fashion and one needs to determine the value of decision variables sequentially. At each time $$t$$, the coefficients $$(r_t, A_t)$$ are revealed, and we need to decide the value of $$x_t$$ instantly. Different from the offline setting, at time $$t$$, we do not have the information of the subsequent coefficients to be revealed, i.e., $$\{(r_t', A_t')\}_{t'=t+1}^T$$. The problem (1) in an online setting is often referred to as online linear programming (Agrawal et al., 2014; Kesselheim et al., 2014). With different specification of the input of the LP, the problem encompasses a wide range of applications, including secretary problem (Ferguson et al., 1989), knapsack problem (Kellerer et al., 2003), network routing problem (Buchbinder and Naor, 2009), matching and Adwords problem (Mehta et al., 2005), service reservation and scheduling problem (Conforti et al., 2014; Stein et al., 2020), network revenue management problem (Talluri and Van Ryzin, 2006), order fulfillment problem (Acimovic and Graves, 2015; Jasin and Sinha, 2015) and more.

We generally refer to the problem as the online revenue management problem. In this paper, we analyze the performance of the classic LP-based adaptive algorithm under both the cases when the underlying LP is nondegenerate or degenerate.

Our contribution is two-fold:

- When the underlying LP is nondegenerate, we derive a regret bound independent of the horizon $$T$$ which improves the existing logarithmic bound in Jasin (2015). The result highlights that the online revenue management is different in its information-theoretic nature from other constrained online learning problem such as bandits with knapsacks problem (BwK, Badanidiyuru et al. (2013)) and online convex optimization with constraints (OCOwC, Yu et al. (2017)). Specifically, the online revenue management problem concerns a full-information setting where the decision maker first observes the information (customer order) and then makes the decision of acceptance and rejection. In contrast, the problems of BwK and OCOwC are in a partial-information setting where the decision is made prior to the observation, and thus bounded regret is not achievable in such setting. Moreover, the finite support condition on the distribution is critical for the achievability of bounded regret. Bray (2019) shows an $$\Omega(\log T)$$ lower bound when the underlying distribution is continuously supported.

- When the underlying LP is degenerate, we derive a regret bound on the order of $$\sqrt{T \log T}$$. The LP-based adaptive algorithm that re-optimizes the decision rule at every time period (also known as frequent re-solving) is criticized for performance deterioration when the underlying LP becomes nearly degenerate. The existing bounds for the algorithm (Jasin and Kumar, 2012; Jasin, 2015; Wu et al., 2015) can be arbitrarily large when the underlying LP is still nondegenerate but close to degenerate. Our analysis provides a positive result that the algorithm achieves $$O(\sqrt{T \log T})$$ regret upper bound regardless of whether the underlying LP is nondegenerate or not. The analysis is nearly tight in that an $$\Omega(\sqrt{T})$$ lower bound is established by Bumpensanti and Wang (2020).
In the following, we review the related literature and highlight the technical novelty of our result and analysis.

1.1 Literature Review

There have been a proliferate literature on algorithm design and analysis for the online revenue management problem. We provide a short summary of our result and the related literature in Table 1. In the following, we will elaborate on our contribution upon the existing literature.

| Benchmark       | Regret Bound      | Key Assumption(s)                           |
|-----------------|-------------------|---------------------------------------------|
| Jasin and Kumar (2012) | Fluid | Bounded | Nondegeneracy, distribution knowledge |
| Jasin (2015)       | Fluid | \( \tilde{O} (\log T) \) | Nondegeneracy |
| Vera and Banerjee (2019) | Hindsight | Bounded | Distribution knowledge |
| Bumpensanti and Wang (2020) | Hindsight | Bounded | Distribution knowledge |
| Asadpour et al. (2019) | Full flex. | Bounded | Long-chain design, \( \xi \)-Hall condition |
| Vera et al. (2020)  | Hindsight | Bounded | Replenishment, binary entries |
| Ours              | Fluid | \( \tilde{O} (\sqrt{T}) \) | Nondegeneracy |

Table 1: Result comparison against literature: The precise definitions of the fluid benchmark and the hindsight benchmark are given in (3) and (4). We note that all the regret bounds here are problem-dependent bound that mainly focuses on the dependence on horizon length \( T \) but will inevitably involve certain parameters related to the underlying distribution/optimization problem.

Known distribution

One stream of literature investigates the existence of an online algorithm that achieves bounded regret under a known distribution setting. Specifically, a line of works (Jasin and Kumar, 2012; Wu et al., 2015; Bumpensanti and Wang, 2020) study the canonical quantity-based network revenue management problem and design algorithms that achieve bounded regret under the knowledge of the order arrival distribution. Jasin and Kumar (2012) show that the LP-based re-solving algorithm achieves bounded regret where the algorithm computes a control policy by periodically solving a linear program specified by the known arrival distribution. A critical condition in Jasin and Kumar (2012) is that the underlying LP should be nondegenerate. A subsequent work (Bumpensanti and Wang, 2020) fully solves the problem through an infrequent re-solving scheme which updates the control policy only at a few selected points. Analysis-wise, one key difference of our analysis compared to Jasin and Kumar (2012) and Bumpensanti and Wang (2020) is a new regret decomposition (Proposition 1). The regret decomposition helps us to circumvent the complication of dealing with the partially-accepted order types and transforms the analysis of partially-accepted order types to that of the constraint process. In this light, the analyses in Jasin and Kumar (2012) and Bumpensanti and Wang (2020) are purely primal-based while our analysis also utilizes the dual problem.

Another line of works (Vera et al., 2019; Vera and Banerjee, 2019; Banerjee and Freund, 2020) devise algorithms that achieve bounded regret for various problems including dynamic pricing, knapsack problem, and bin-packing problem. The authors develop a novel and intuitive approach called “compensated coupling” to derive regret upper bounds. The idea is to bound the cumulative expected loss induced by the different decisions made by the online algorithm against the hindsight optimal which has all the future information (formally defined in (4)). As in the aforementioned works on network revenue management, this line of works are also built upon the knowledge of the underlying distribution and the algorithms are closely related to the idea of approximate dynamic programming.
Unknown distribution

Efforts have also been made to relax the assumption on knowing the underlying distribution. Banerjee et al. (2020) consider one single historical trace of observations in substitute of the knowledge of the true distribution and derive an algorithm that achieves bounded regret for certain constrained online optimization problem. In certain sense, the single historical trace can contribute to $\Omega(n)$ observations (with $n$ being the length of horizon). In a similar spirit, Shivaswamy and Joachims (2012) characterize the number of historical observations needed to achieve a bounded regret for the stochastic multi-armed bandits problem. The historical observations required in these two works can be viewed as a warm start for the online procedure, and the empirical distribution constructed from the historical observations provides a moderately good estimation for the true underlying distribution at the very beginning of the online procedure.

As a subsequent work of Jasin and Kumar (2012), Jasin (2015) studies the same problem as ours and proves an $O(\log T)$ regret for the LP-based adaptive algorithm. Apart from the new regret decomposition mentioned above, we now provide the second explanation for why we manage to improve the regret bound. First, we point out that the key of the analysis under known distribution in Jasin and Kumar (2012) is a martingale analysis of the constraint process. A recent work (Balseiro et al., 2021) extends the result in Jasin and Kumar (2012) to more general constrained revenue management problems but still requires the distribution knowledge. The common point of Jasin and Kumar (2012) and Balseiro et al. (2021) is that with the knowledge of the underlying distribution, the constraint process is a martingale under the adaptive (re-solving) scheme. When the distribution is unknown, the constraint process is no longer a martingale. Jasin (2015) leaves the martingale approach and defines a sequence of high probability events to analyze the constraint process. In contrary, we still stick to the martingale approach: different from the known distribution case, we identify conditions under which the constraint process is not a martingale, but it still behaves “stably” to suffice for the bounded regret analysis. The conditions need to be carefully tuned so that they can be met with high probability. We remark that the idea applies to both cases of nondegenerate and degenerate. Under both cases, we first start to identify the desirable condition on the constraint process to achieve the corresponding regret bound and then show that such condition can be met with high probability.

A stream of works (Agrawal et al., 2014; Bray, 2019; Li and Ye, 2021) also consider the problem under unknown distribution but impose no finite support condition. Similar to our work, the primal-dual approach is often used in such setting for both algorithm design and analysis. Both lower and upper bounds of order $\log T$ are established for the case when the distribution support is infinite. In this light, the finite support condition is critical in the achievability of bounded regret. Intuitively, when the LP is nondegenerate, the finite support condition creates a well-separatedness between different order types, and the separation can be learned with high probability using $O(1)$ number of samples. This contrast between the finite and infinite support is analogous to the findings in the (unconstrained) newsvendor problem (Besbes and Muharremoglu, 2013).

Other related works

Another work related to our result is Asadpour et al. (2019), where the authors derive bounded regret for a resource allocation problem to study the effectiveness of the long-chain design without knowing the true distribution. Technically, the formulation in Asadpour et al. (2019) can be cast in (1) by imposing a binary structure on the constraint matrices $A_t$’s, along with certain other conditions. Another recent work (Vera et al., 2020) on a revenue management/resource allocation problem considers a setting where there is resource replenishment. Similar to our work, the paper focuses on analyzing the resource consumption process and takes a geometric viewpoint. In addition to the resource replenishment, the paper assumes the entries in the constraint matrix is binary and all the resources are binding, while
both assumptions are critical in the regret derivation. Two subsequent works (Kerimov et al., 2021a,b) consider a dynamic matching problem and develop algorithms achieving bounded regret utilizing the structure of the underlying LP.

2 Model, Assumption, and Algorithm

In this section, we formulate the problem and present the LP-based adaptive algorithm. We first focus on the one-dimensional case for notation simplicity where \( k = 1 \) in LP (1) and will discuss the general problem in the appendix. For \( k = 1 \), the online formulation of LP (1) reduces to a one-dimensional online LP problem,

\[
\max \sum_{t=1}^{T} r_t x_t \tag{2}
\]

s.t. \( \sum_{t=1}^{T} a_t x_t \leq B \)

\[0 \leq x_t \leq 1, \ t = 1, \ldots, T\]

where \( a_t = (a_{1,t}, \ldots, a_{m,t})^\top \in \mathbb{R}^m \) and the decision variables are \( x = (x_1, \ldots, x_T)^\top \in \mathbb{R}^T \). There are \( m \) constraints and \( T \) decision variables. Throughout the paper, we use \( i \) to index constraints and \( t \) to index decision variables.

Now, we introduce our first group of assumptions on the distribution that governs the generation of \((r_t, a_t)\)'s. In the next section, we will introduce an additional assumption on the nondegeneracy of the underlying LP.

Assumption 1 (Distribution). We assume

(a) Stochastic: The column-coefficient pair \((r_t, a_t)\)'s are i.i.d. sampled from a distribution \( P \). The distribution \( P \) takes a finite and known support \( \{(\mu_j, c_j)\}_{j=1}^{n} \) where \( \mu_j \in \mathbb{R} \) and \( c_j \in \mathbb{R}^m \). Specifically,

\[P((r_t, a_t) = (\mu_j, c_j)) = p_j\]

for \( j = 1, \ldots, n \). The probability vector \( p = (p_1, \ldots, p_n)^\top \) is unknown.

(b) Positiveness and Boundedness: \( 0 \leq \mu_j \leq 1, \ c_j \geq 0 \) and \( \|c_j\|_\infty \leq 1 \) for \( j = 1, \ldots, n \).

(c) Linear growth: The right-hand-side \( B = Tb \) for some \( b = (b_1, \ldots, b_m)^\top > 0 \).

Assumption 1 (a) imposes a stochastic assumption for the customer orders. In addition, it states that the support of the distribution is finite and known, but the parameters are unknown. In other words, it means that there are \( n \) known order types, and the order type at each time \( t \) follows a multinomial distribution with unknown parameters. Assumption 1 (b) requires all the entries of \((\mu_j, c_j)\) between 0 and 1. We remark that all the results in this paper still hold (up to a constant) when this part is violated, and the positiveness and boundedness are introduced only for notation simplicity. Lastly, the linear growth condition in Assumption 1 (c) is commonly assumed in problem setup and regret analysis online resource allocation problems. In our context, the condition is mild in that if \( B = o(T) \), we can always adjust the time horizon with \( T' \ll T \) such that \( B = T'b \), and consequently the linear growth condition holds for \( T' \).
2.1 Performance measure

In the literature of revenue management and the more general constrained online learning problems, a commonly considered performance benchmark is the certainty equivalent version of the “offline” LP (2),

\[
\text{OPT}_D := \max \sum_{j=1}^{n} p_j \mu_j y_j,
\]

subject to

\[
\sum_{j=1}^{n} p_j c_j \cdot y_j \leq b,
\]

\[
0 \leq y_j \leq 1, \quad j = 1, \ldots, n.
\]

Recall from Assumption 1 that \(\mu_j\) and \(c_j\) represent the revenue and requested resource consumption of the \(j\)-th order type, respectively. The right-hand-side \(b = B/T\) represents the average resource capacity per time period, and \(p_j\) is the probability of the \(j\)-th order type. The decision variables \(y_j\)’s prescribe a “probabilistic” decision rule for the orders, and \(y_j\) can be interpreted as the proportion of accepted orders (or the probability of accepting orders) for the \(j\)-th order type. The connection between the LPs (2) and (3) can be seen from Assumption 1 that the probability of the \(j\)-th order type is \(p_j\). In this light, LP (3) can be viewed as a deterministic version obtained by taking expectation of the objective and the left-hand-side of LP (2). For such reason, we refer to (3) as the deterministic LP (DLP). It is easy to verify that the optimal objective value of (3), with a proper scaling of factor \(T\), upper bounds the expected offline (hindsight) optimal objective value of (2). Specifically, the expected offline (hindsight) optimal objective value is denoted by \(E[\text{OPT}_{\text{Hind}}]\), where \(\text{OPT}_{\text{Hind}}\) is a random variable and is defined by

\[
\text{OPT}_{\text{Hind}} := \max \sum_{t=1}^{T} r_t x_t
\]

subject to

\[
\sum_{t=1}^{T} a_t x_t \leq B
\]

\[
0 \leq x_t \leq 1, \quad t = 1, \ldots, T,
\]

where \((r_t, a_t)\) follows the distribution in Assumption 1. It can be shown that \(E[\text{OPT}_{\text{Hind}}] \leq T \cdot \text{OPT}_D\). In this light, the DLP (fluid or certainty-equivalent) benchmark is a stronger one than the expected offline (hindsight) one.

For the online problem, at each time \(t\), we decide the value of \(x_t\): \(x_t = 1\) means that we accept the order and allocate \(a_t\) amount of resources to this order accordingly; \(x_t = 0\) means that we reject the order. In this paper, we focus on the case of integer-valued solution, i.e., \(x_t = 0\) or 1, but the analysis can be easily extended to the case where partial acceptance is allowed. Like the offline problem, we need to conform to the constraints throughout the procedure, i.e., no shorting of the resources is allowed. In this paper, we consider regret as the performance measure, formally defined as follows:

\[
\text{Reg}_T^\pi := E \left[ T \cdot \text{OPT}_D - \sum_{t=1}^{\tau} r_t x_t \right]
\]

where the quantity \(\text{OPT}_D\) represents the optimal objective value of the DLP problem (3) and \(x_t\)’s represent the online solution. Here \(\tau\) is the stopping time for an algorithm when one or more types of the resource is depleted. The superscript \(\pi\) denotes the online algorithm/policy according to which the online decisions are made. The expectation is taken with respect to \((r_t, a_t)\)’s and the (possible)
randomness introduced by the algorithm.

2.2 LP-based Adaptive Algorithm

Before formally describing the algorithm, we first introduce a few additional notations to characterize the constraint consumption process. Define $B_1 = B$ and $B_t = (B_{1,t}, ..., B_{m,t})^\top$ as the remaining resource capacity at the beginning of time $t$, i.e.,

$$B_t = B_{t-1} - a_{t-1}x_{t-1}.$$ 

Accordingly, we define $b_t = B_t/(T - t + 1)$ as the average resource capacity for the remaining time periods. In addition, we use $B_{T+1}$ to denote the remaining constraint at the end of horizon, and the initial $b_1 = (b_{1,1}, ..., b_{1,m})^\top = B_1/T = B/T = b$. We formally define the stopping time $\tau$ based on $B_t$

$$\tau := \min\{t : B_{i,t} \leq 1 \text{ for some } i = 1, ..., m\} - 1.$$ 

The rationale behind the definition is that when $B_{i,t} \leq 1$, there may arrive an order that is profitable but cannot be fulfilled due to the resource constraint. On the opposite, when $B_{i,t} > 1$ for all $i$, we can fulfill any possible arriving order (given $\|c_j\|_\infty \leq 1$). Let $n_j(t)$ denote the counting process of the $j$-th order type, i.e., the number of observations $(\mu_j, c_j)$ up to time $t$ ( inclusively) for $j = 1, ..., n$. Since no shorting is allowed, i.e., the remaining constraint vector $B_t$ must be element-wise non-negative for all $t = 1, ..., T$. Notice that the true probability distribution $p = (p_1, ..., p_n)$ is unknown. The counts $n_j(t)$’s will be used by the online algorithm to construct empirical estimates for the corresponding probabilities.

Now we formally present the LP-based adaptive algorithm as Algorithm 1. At each time $t$, the algorithm solves a sampled linear program (5) to compute the probability of acceptance for each order type $(\mu_j, c_j)$. The LP (5) takes a similar form as LP (3) but differs in two aspects: (i) the probabilities $p_j$’s in (3) are replaced with their empirical estimates since the underlying distribution is assumed unknown; (ii) the right-hand-side $b$ in (3) is replaced with its adaptive counterpart $b_t$. Algorithm 1 then uses the LP’s optimal solution $y_{j,t}^*$ to determine the online solution $x_t$ at time $t$. Recall that $y_{j,t}^*$ denotes the optimal proportion of acceptance rate for the $j$-th order type for the deterministic LP associated with time $t$. Thus the probabilistic decision rule in Algorithm 1 aims to follow the prescription of the optimal solution by accepting the $j$-th order type with probability $y_{j,t}^*$.

The algorithm is not new, and the adaptive design (using a dynamic right-hand-side) is commonly known as the re-solving technique in the network revenue management literature. Essentially, the algorithm has the same structure as the re-solving algorithms in Jasin and Kumar (2012), Jasin (2015), Bumpensanti and Wang (2020), and Li and Ye (2021). The algorithm here re-solves the problem in each time period, and it uses the sample counts as estimates for the true probabilities. In the following two sections, we analyze the performance of the algorithm for the cases when the underlying DLP (3) is (i) nondegenerate and (ii) degenerate.
Algorithm 1  Adaptive Allocation Algorithm

1: Input: $B, T, \{(\mu_j, c_j)\}_{j=1}^n$
2: Initialize $B_1 = b, b_1 = B_1/T$
3: Set $x_1 = 1$
4: for $t = 2, \ldots, T$ do
5: Compute $B_t = B_{t-1} - a_{t-1}x_{t-1}$
6: Compute $b_t = B_t/(T - t + 1)$
7: Solve the following linear program where the decision variables are $(y_1, \ldots, y_n)$:
\[
\begin{align*}
\max & \quad \sum_{j=1}^n \frac{n_{t-1}(j)}{t-1} \mu_j y_j \\
\text{s.t.} & \quad \sum_{j=1}^n \frac{n_{t-1}(j)}{t-1} c_j y_j \le b_t \\
& \quad 0 \le y_j \le 1, \quad j = 1, \ldots, n
\end{align*}
\]
8: Denote the optimal solution as $y^*_t = (y^*_1, \ldots, y^*_n, t)$
9: Observe $(r_t, a_t)$ and identify $(r_t, a_t) = (\mu_j, c_j)$ for some $j$
10: Set
\[
\begin{align*}
x_t = \begin{cases} 1, & \text{with probability } y^*_j, t \\
0, & \text{with probability } 1 - y^*_j, t 
\end{cases}
\end{align*}
\]
when the constraint permits; otherwise set $x_t = 0$.
11: Update the counts
\[
n_j(t) = \begin{cases} n_j(t - 1) + 1, & \text{if } (r_t, a_t) = (\mu_j, c_j) \\
    n_j(t - 1), & \text{otherwise}
\end{cases}
\]
12: end for
13: Output: $x = (x_1, \ldots, x_T)$

3  Regret Analysis for Nondegenerate Case

The standard form of the DLP (3) is as below.
\[
\begin{align*}
\max & \quad \mu^\top y \\
\text{s.t.} & \quad Cy + s = b \quad \text{(dual variable: } \lambda) \\
& \quad y + z = 1 \\
& \quad y, s, z \ge 0
\end{align*}
\]  

where the decision variable vector $y$ is the same as the one in (3). With a slight abuse of notation (omitting the effect of the probability vector $p$), we use $\mu$ to denote the vector $(p_1\mu_1, \ldots, p_n\mu_n)^\top$ and $C$ to denote the matrix $(p_1c_1, \ldots, p_nc_n)$. The additional decision variables $s \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ represent the slack variables for the corresponding constraints. The dual program for both the DLP (3) and its standard form (6) is
\[
\begin{align*}
\min & \quad b^\top \lambda + \sum_{j=1}^n \gamma_j \\
\text{s.t.} & \quad p_j c_j^\top \lambda + \gamma_j \ge p_j \mu_j, \quad j = 1, \ldots, n \\
& \quad \lambda \ge 0, \gamma_j \ge 0, \quad j = 1, \ldots, n
\end{align*}
\]

where the decision variables are $\lambda$ and $\gamma_j$'s.
Denote the optimal solution to the LP (6) as \((y^*, s^*, z^*)\) and the dual optimal solution (also known as dual price) of (7) for the resource constraints in (6) as \(\lambda^*\). Accordingly, we define the sets of basic and non-basic variables/order types as
\[
J^* := \{j : \mu_j \geq c_j^\top \lambda^*, j = 1, \ldots, n\}, \quad J' := \{j : \mu_j < c_j^\top \lambda^*, j = 1, \ldots, n\},
\]
and the sets of binding and non-binding constraints as
\[
I^* := \{i : b_i = C_i^\top y^*, i = 1, \ldots, m\}, \quad I' := \{i : b_i > C_i^\top y^*, i = 1, \ldots, m\},
\]
where \(C_{ij}\) denotes the \(i\)-th row of the constraint coefficient matrix \(C\) in (6). Here, the dual optimal solution \(\lambda^*\) provides a pricing rule for the resource consumption of a certain order.

Throughout this section, we assume a nondegeneracy structure for the standard form LP as below.

**Assumption 2** (Nondegeneracy). The optimal solution to (6) is unique and nondegenerate, i.e.,
\[
|\{i : y^*_i \neq 0, i = 1, \ldots, m\}| + |\{j : s^*_j \neq 0, j = 1, \ldots, n\}| + |\{z^*_j \neq 0, j = 1, \ldots, n\}| = m + n.
\]

The assumption is a standard one in the literature of linear programming, and with an arbitrarily small perturbation any LP can satisfy the assumption (Megiddo and Chandrasekaran, 1989). An implication of Assumption 2 is a stability structure for the underlying LP as follows.

**Lemma 1** (Stability under nondegeneracy). Under Assumption 1 and 2, there exists a positive constant \(L\) which depends on \(\mu, C\) and \(b\) such that if
\[
\max \{\|\hat{C} - C\|_\infty, \|\hat{\mu} - \mu\|_\infty, \max_{i \in I^*} |b_i - \hat{b}_i|, \max_{i \in I'} |b_i - \hat{b}_i|\} \leq L,
\]
then the following LP shares the same optimal basis and set of binding constraints with LP (6),
\[
\begin{align*}
\max & \quad \hat{\mu}^\top y \\
\text{s.t.} & \quad \hat{C} y + s = \hat{b} \\
& \quad y + z = 1 \\
& \quad y, s, z \geq 0.
\end{align*}
\]

From a geometric viewpoint, this lemma ensures that the optimal solutions of (6) and (8) coincide at the same corner point of the corresponding feasible simplex. We remark that for the non-binding constraints \(i \in I'\), it only needs a lower bound for \(\hat{b}_i\). For the online problem, the parameters \((\mu, C)\) are estimated through observations, and the resource level may deviate from the initial \(b\). Thus the perturbed LP (8) is analogous to the adaptive LP (5) used in the algorithm. In the previous analysis of the network revenue management problem (Jasin and Kumar, 2012; Jasin, 2015), the nondegeneracy assumption is used in a similar way to establish a stability for the underlying LP. We defer the proof of Lemma 1 to Appendix E where we further relate the parameter \(L\) explicitly with several key parameters of the underlying LP. The relationship refines the analysis in Mangasarian and Shiau (1987) and may be of independent interest. In the rest of the paper, we will express our regret bounds in terms of the parameter \(L\) in Lemma 1.

### 3.1 Regret Decomposition

The starting point of our analysis is to decompose the regret into three parts: (i) the first two parts concern the “incorrect” number of order acceptance; (ii) the third part concerns the remaining resources
weighted by the dual price. We summarize the result in the following proposition, and defer the proof to Appendix A.1.

**Proposition 1.** Under Assumption 1 and 2, the following equality holds

\[
\text{Reg}^{\pi}_T = \sum_{j \in J^*} (\mu_j - c_j^T \lambda^*) \cdot \mathbb{E} [n_j(T) - n_j^*(\tau)] + \sum_{j \in J'} (c_j^T \lambda^* - \mu_j) \cdot \mathbb{E} [n_j^*(\tau)] + \lambda^* \cdot \mathbb{E} [B_T]
\]

where \(n_j^*(t)\) denotes the number of accepted orders of the \(j\)-th type up to time \(t\) (inclusively) under policy \(\pi\). Here \(n_j(t)\) denotes the total number of orders of the \(j\)-th type up to time \(t\) (inclusively) as defined earlier.

The equality has an intuitive interpretation. Recall that \(\tau\) is the stopping time of the algorithm and the first time that some resource is (almost) depleted, and \(B_T\) denotes the remaining resource vector when the algorithm terminates. Thus the last part on the right-hand-side of (9) penalizes the wasted resources when the process terminates. In particular, only residuals of the binding resources will be penalized. As to the first two parts on the right-hand-side of (9), we categorize the order types and elaborate as below:

- **All-accepted orders:** \(\mu_j - c_j^T \lambda^* > 0\). For these orders, the optimal decision should be to accept all of them. We will observe \(n_j(T)\) such orders throughout the horizon and aim to have the number of acceptance \(n_j^*(\tau)\) close to that.

- **All-rejected orders:** \(\mu_j - c_j^T \lambda^* < 0\). On the opposite of the previous case, the optimal decision should be to reject all of these orders. Each acceptance of such order will induce a cost of \(c_j^T \lambda^* - \mu_j\); resources of value \(c_j^T \lambda^*\) are spent, but only reward of value \(\mu_j\) is received.

- **Partially-accepted orders:** \(\mu_j - c_j^T \lambda^* = 0\). The condition may lead to a proportional acceptance of the orders, i.e., \(0 \leq y_j^* \leq 1\). Analysis-wise, there is no need to worry about these orders because they do not contribute to the first two terms of the right-hand-side of (9).

We make the following remarks for the regret decomposition. First, when the underlying probability distribution is unknown, the above categorization is also unknown a priori and should be learned. Second, in general, to identify the right proportion of acceptance for the partially-accepted order types is more challenging than the other two categories of orders. However, with the regret decomposition, we only need to focus on analyzing the constraint consumption \(B_T\) and avoid the complication related to the analysis of the partially-accepted orders. This is in contrast with the existing works (Jasin and Kumar, 2012; Jasin, 2015; Bumpensanti and Wang, 2020) where it needs to carefully chase after (i) the number of the acceptance and (ii) the number of total arrivals of the partially-accepted order types. This paradigm shift from decision-variable-centric to constraint-centric is crucial in tightening the regret bound. Third, the nondegeneracy assumption makes the all-accepted orders and the all-rejected orders well separated from the remaining order types. Without this assumption, some order type’s categorization (e.g. as either partially-accepted or all-rejected) can only be revealed in hindsight or at the very end of the online procedure (Bumpensanti and Wang, 2020).

The following corollary extends Proposition 1 to the case of a more general stopping time, and its proof can be found in Appendix A.2.

**Corollary 1.** The following inequality holds

\[
\text{Reg}^{\pi}_T \leq \sum_{j \in J^*} (\mu_j - c_j^T \lambda^*) \cdot \mathbb{E} [n_j(\tau') - n_j^*(\tau')] + \sum_{j \in J'} (c_j^T \lambda^* - \mu_j) \cdot \mathbb{E} [n_j^*(\tau')] + (T - \mathbb{E}[\tau']) \cdot \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| + \lambda^* \cdot \mathbb{E} [B_{\tau'}]
\]

\[\tag{10}\]

10
where \( \tau' \) is a stopping time adapted to the process \( B_t \)'s and \( \tau' \leq \tau \) almost surely.

The corollary replaces the algorithm termination time \( \tau \) in Proposition 1 with a general stopping time \( \tau' \) and it includes an additional term \( T - \mathbb{E}[\tau'] \) measuring the closeness of \( \tau' \) to the end of the horizon. This generalization gives much more flexibility in choosing a proper stopping time when analyzing the regret. In the following two subsections, we will analyze the terms in Corollary 1 part by part.

### 3.2 Order Acceptance – First Two Terms in Regret Decomposition

Next, we utilize Lemma 1 to analyze the first and second term in (10). Recall that in the setting of the online resource allocation problem, one passively collects the observations of orders. The sampled LP (5) solved in Algorithm 1 will gradually converge to the DLP (3) if given the same right-hand-side. So, a concentration argument leads to a bound on the number of time periods it takes until the condition in Lemma 1 is met. Since then, the algorithm will make no further mistake on the acceptance (or rejection) of all-accepted orders (or all-rejected orders) defined by Proposition 1. One caveat is that part of the condition in Lemma 1 concerns the right-hand-side of the LP, so we also need to impose some restrictions on \( b_t \) – the right-hand-side of the sampled LP (5).

Specifically, we define a stopping time based on the constant \( L \) in Lemma 1,

\[
\tau_S := \min \{ t \leq T : |b_{i,t} - b_i| > L \text{ for some } i \in I^* \} \cup \{ t \leq T : b_{i,t} - b_i > -L \text{ for some } i \in I' \} \cup \{ T + 1 \}.
\]

Hence for any time \( t < \tau_S \), the right-hand-side \( b_t \) meets the condition in Lemma 1. Thus the underlying adaptive LP shares the same structure as the DLP (3) when \( t \leq \tau_S \). This property makes the stopping time \( \tau_S \) easier to analyze than the original stopping time \( \tau \). Because when \( t \) approaches to \( \tau \), the underlying LP’s optimality and bindingness structure may already change; but this will not happen for \( t < \tau_S \).

By the definition of \( \tau_S \), if we restrict our attention to time periods before \( \tau_S \), the numbers of mistakes made on the all-accepted and all-rejected orders are purely caused by the inaccurate estimation on the left-hand-side. To obtain an upper bound, the concentration argument can be applied and it leads to the following proposition.

**Proposition 2.** Under Assumptions 1 and 2, the output of Algorithm 1 satisfies

\[
\sum_{j \in J^*} (\mu_j - c_j^\top \lambda^*) \cdot \mathbb{E} \left[ n(\tau_S) - n_j^a(\tau_S) \right] + \sum_{j \in J'} (c_j^\top \lambda^* - \mu_j) \cdot \mathbb{E} \left[ n_j^a(\tau_S) \right] \leq \frac{2n \max_j |\mu_j - c_j^\top \lambda^*|}{1 - \exp(-2L^2)}.
\]

We defer the proof to Appendix A.3. Note that the right-hand-side of the above inequality is not dependent on the time horizon \( T \). The proposition’s proof directly follows an application of Lemma 1. First, as mentioned earlier, the definition of \( \tau_S \) precludes the possibility that the changing right-hand-size \( b_t \) in Algorithm 1 affects the LP’s stability. Then, with a sufficient number of observations, the condition in Lemma 1 can be satisfied with high probability. From that time on, the sampled LP (5) and the deterministic LP (3) will share the same optimal basis, and consequently the algorithm will not make any further mistake on the all-accepted and all-rejected orders.

If we put together Proposition 2 with Corollary 1, the remaining task is to bound the remaining time periods \( \mathbb{E}[T - \tau_S] \) and the remaining resources \( \mathbb{E}[B_{\tau_S}] \). In fact, these two aspects are closely related with each other and they will be the focus of the next subsection.
3.3 Constraint Consumption Process – Last Two Terms in Regret Decomposition

In this section, we analyze the constraint consumption process $B_t$ (equivalently, $b_t$) and handle the last two terms in the generic upper bound (10). To better illustrate the proof idea, we focus in this section on the case that all the constraints are binding. We will show in Appendix C how the analysis can be adapted to the case where there are both binding and non-binding constraints.

**Assumption 3.** All the resource constraints are binding, i.e., $s^* = 0$ in (6).

From Lemma 1, we know that under Assumptions 1 and 2, the LP’s optimality and bindingness structure remains to hold when $b$ is perturbed. Let

$$B := \bigotimes_{i=1}^{m} [b_i - L, b_i + L].$$

where $L$ is the constant in Lemma 1. The following lemma states that for any $\tilde{b} \in B$, all the constraints of the corresponding LP are binding.

**Lemma 2.** Under Assumption 1, 2, and 3, for each $\tilde{b} \in B$, there exists an optimal solution $\tilde{y}^* = (\tilde{y}^*_1, ..., \tilde{y}^*_n)^\top$ of the DLP (3) with the right-hand-side being $\tilde{b}$ that satisfies

$$\sum_{j=1}^{n} p_j c_j \cdot \tilde{y}^*_j = \tilde{b}.$$

Our goal of analyzing $B_t$ is to bound (i) the remaining time periods when the algorithm terminates, $T - E[\tau_S]$, and (ii) the resource left-over, $\lambda^\top E[\tau_S]$. Throughout the analysis, we will reserve the notation $b = (b_1, ..., b_m)^\top$ for the initial average resource capacity and use $b'$ to denote an arbitrary value in $\mathbb{R}^m$. Ideally, the process $b_t$ should stay near $b$ throughout the horizon, as this would imply that the resource is exhausted only at the very end of the horizon. Let $x_t(b')$ denote the online solution output by Algorithm 1 at the $t$-th time period as a function of the input $b_t = b'$. Consider the following event defined in the space of the history observations up to time $t - 1$,

$$E_t := \left\{ H_{t-1} \mid \sup_{b' \in B} \|E[a_t x_t(b')|H_{t-1}, b_t = b'] - b'\|_\infty \leq \epsilon_{t-1} \right\}$$

where the history $H_{t-1} = (r_1, a_1, ..., r_{t-1}, a_{t-1})$. Here we choose

$$\epsilon_t := \begin{cases} 1 & t \leq \kappa T, \\ t^{-\frac{1}{2}} & t > \kappa T, \end{cases}$$

(11)

where the constant $\kappa \in (0, 1)$ is to be specified later and will be roughly on the same order of $L$. Without loss of generality, we assume $\kappa T$ takes an integer value.

Now, we provide some intuitions for the definition of $E_t$. First, recall that

$$b_{t+1} = \frac{B_{t+1}}{T - t} = \frac{B_t - a_t x_t}{T - t} = b_t - \frac{1}{T - t} (a_t x_t - b_t)$$

for $t = 1, ..., T - 1$. The definition of $E_t$ is aligned with the hope that the expected resource consumption at each time $t$ stays close to $b_t$. Specifically, the event $E_t$ controls the expectation of $a_t x_t - b_t$, and in its definition, the supremum taken over $b' \in B$ is necessary because $b_t$ is random. As to the choice of $\epsilon_t$:

- $\epsilon_t$ should not be too small so that the events $E_t$’s will happen with high probability.
the probability of obtaining the good estimate for the model parameters. Piecing the two components concerns the probability of the fluctuation of the process \( \tau < \tau_S \), i.e., the expected constraint consumption in a single time period has a large deviation from zero. The stopping time \( \tilde{\tau} \) chops off such “bad” event and through the lens of \( \tau_S \), and the constraint process becomes easier to analyze.

Specifically, with \( \tilde{\tau} \), we define an auxiliary process \( \tilde{b}_t \) as follows

\[
\tilde{b}_t = \begin{cases} 
   b_t, & t < \tilde{\tau}, \\
   b_{\tilde{\tau}}, & t \geq \tilde{\tau}.
\end{cases}
\]

By its definition, the process \( \tilde{b}_t \) freezes its value once \( b_t \) exits the region \( \mathcal{B} \) or the bad event \( \tilde{\mathcal{E}}_t \) happens. With \( \tilde{b}_t \), we have

\[
P(\{b_s \notin \mathcal{B} \text{ for some } s \leq t\}) = P(\{b_s \notin \mathcal{B} \text{ for some } s \leq t, \cap_{s=1}^t E_s\}) + P(\{b_s \notin \mathcal{B} \text{ for some } s \leq t, \cup_{s=1}^t E_s\})
\]

\[
\leq P(\tilde{b}_s \notin \mathcal{B} \text{ for some } s \leq t) + \sum_{s=1}^t P(E_s)
\]

(12)

where \( E \) denotes the complement of an event \( \mathcal{E} \). For the first part of the second line, it is because given \( \cap_{s=1}^t E_s \), the event that \( b_s \notin \mathcal{B} \) for some \( s \leq t \) is equivalent to the event \( \tilde{\tau} \leq t \) and thus it entails \( b_s \notin \mathcal{B} \).

For the second part, it is obtained by ignoring the condition on \( b_s \) and then taking a union bound with respect to \( s = 1, ..., t \).

We justify the decomposition of the left-hand-side of (12) by relating it with the last two terms in Corollary 1 as the following lemma.

**Lemma 3.** We have the following relation between the stopping time and the left-hand-side of (12)

\[
E[T - \tau_S] \leq \sum_{t=1}^T P(\{b_s \notin \mathcal{B} \text{ for some } s \leq t\}),
\]

and for the last two terms in Corollary 1, we have

\[
(T - E[\tau_S]) \cdot \max_{j \in [n]} |\mu_j - \epsilon_j^T \lambda^*| + \lambda^T \cdot E[B_{\tau_S}] \leq ||\lambda^*||_1 \cdot (3E[T - \tau_S] + 4).
\]

Now we discuss the motivation for defining \( \tilde{b}_t \) and why the inequality (12) is useful. The inequality (12) separates our goal – the probability on the left-hand-side – into two components. The first component concerns the process \( \tilde{b}_t \) which is a relatively “well-behaved” process in that when \( t < \tilde{\tau} \), the process \( \tilde{b}_t \)'s fluctuation is subject to the event \( \tilde{\mathcal{E}}_t \); when \( t \geq \tilde{\tau} \), the process freezes. The event \( \mathcal{E}_t \) further controls the fluctuation of the process \( b_t \), and thus the process behaves roughly like a martingale. The second component concerns the probability of \( \tilde{\mathcal{E}}_t \)'s, which can be analyzed individually for each \( t \). Overall, the inequality (12) disentangles the stability of the process \( b_t \) from the estimation error. Its first component concerns the process stability given a good estimate (the event \( \mathcal{E}_t \)), while its second component concerns the probability of obtaining the good estimate for the model parameters. Piecing the two components
together, we obtain a bound on the probability that $b_t$ exists the region $B$.

Next, we will analyze the two components in (12) separately and then combine the results to derive the regret bound.

**Analysis of the first component in (12).**

The following theorem states a concentration result for a general martingale difference sequence $X_t$’s. The approach that we analyze the first component can be viewed as a two-step procedure: we first construct a martingale as an approximation of the process $\tilde{b}_t$. Specifically, the constructed martingale and $\tilde{b}_t$ share the same initial value $b$, and the difference between the martingale and $\tilde{b}_t$ is controllably small. Second, we apply Theorem 1 for the constructed martingale and argue that both the martingale and $\tilde{b}_t$ will stay within $B$ with high probability.

**Theorem 1.** (Hoeffding’s inequality for dependent data (van de Geer, 2002)) Consider a sequence of random variables $\{X_t\}_{t=1}^T$ adapted to the filtration $\mathcal{F}_t$’s and $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = 0$ for $t = 1, \ldots, T$

where $\mathcal{F}_0 = \emptyset$. Suppose $L_t, U_t$ are $\mathcal{F}_{t-1}$-measurable random variables such that $L_t \leq X_t \leq U_t$ almost surely for $t = 1, \ldots, T$. Let $S_t = \sum_{s=1}^{t} X_t$ and $V_t = \sum_{s=1}^{t} (U_s - L_s)^2$. Then, the following inequality holds for any $b > 0, c > 0$ and $T \in \mathbb{N}^+$,

$$
\mathbb{P}(|S_t| \geq b, V_t \leq c^2 \text{ for some } t \in \{1, \ldots, T\}) \leq 2e^{-\frac{2b^2}{c^2}}.
$$

The following lemma utilizes the result in Theorem 1 and provides an upper bound on the first component in (12). Thus it completes the analysis of the first component in (12). The proof can be found in Appendix A.5.

**Lemma 4.** For $T \geq T_1$ and $t \leq T - 2$, the following inequality

$$
\mathbb{P} \left( \tilde{b}_s \notin \bigotimes_{i=1}^{m} [b_i - \Delta, b_i + \Delta] \text{ for some } s \leq t \right) \leq 2me^{-\frac{\Delta^2(t-1)}{8}}
$$

holds for any $\Delta > 0$. The constant $T_1$ is defined as the minimal integer such that $T_1 \geq \frac{1}{\exp\left(\frac{1}{2}\right)-1} + 2$ and $\frac{\log T_1}{T_1^{1/4}} \leq \frac{e^{1/4}}{4}$. For the parameter in the definition of $\epsilon_t$ in (11), we set $\kappa = 1 - \exp\left(-\frac{L}{8}\right)$.

From the lemma, if we set $\Delta = L$ and $\kappa = 1 - \exp\left(-\frac{L}{8}\right)$, we have

$$
\mathbb{P} \left( \tilde{b}_s \notin B \text{ for some } s \leq t \right) \leq 2me^{-\frac{L^2(t-1)}{8}}.
$$

Thus we obtain a bound for the first component in (12).

**Analysis of the second component in (12).**

Define the events for $t = 2, \ldots, T$

$$
A_t^{(j)} := \left\{ \left| \frac{n_j(t-1)}{t-1} - p_j \right| \leq L \right\}
$$

and

$$
B_t^{(j)} := \left\{ \left| \frac{n_j(t-1)}{t-1} - p_j \right| \leq \frac{1}{n(t-1)^{1/4}} \right\}.
$$

14
In addition, we define \( A_1^{(j)} = B_1^{(j)} = \Omega \).

Now we argue that
\[
\left( \bigcap_{j=1}^n A_1^{(j)} \right) \cap \left( \bigcap_{j=1}^n B_1^{(j)} \right) \subseteq \mathcal{E}_t.
\]

To see this, for \( t \geq 2 \), let \( y_t^* = (y_{1t}^*, \ldots, y_{nt}^*)^\top \) be the optimal solution of (5) with \( b_t = b' \) for some \( b' \in \mathcal{B} \). By the algorithm, we have the expected resource consumption at time \( t \)
\[
E[a_t x_t(b') | \mathcal{H}_{t-1}, b_t = b'] = \sum_{j=1}^n p_j c_j \cdot y_{jt}^*.
\]
Moreover, we know that given the event \( \bigcap_{j=1}^n A_1^{(j)} \), the perturbation of \( C \) and \( \mu \) (described in Lemma 1) will be within \( L \). From Lemma 1 and Assumption 3 we know that all the constraints of the LP with right-hand-side being \( b' \) are binding
\[
b' = \sum_{j=1}^n \frac{n_j(t-1)}{t-1} c_j \cdot y_{jt}^*.
\]
Then, taking the difference,
\[
E[a_t x_t(b') | \mathcal{H}_{t-1}, b_t = b'] - b' = \sum_{j=1}^n \left( p_j - \frac{n_j(t-1)}{t-1} \right) c_j \cdot y_{jt}^*.
\]
Next, given the event \( \bigcap_{j=1}^n B_1^{(j)} \), we have
\[
\|E[a_t x_t(b') | \mathcal{H}_{t-1}, b_t = b'] - b'\|_\infty = \left\| \sum_{j=1}^n \left( p_j - \frac{n_j(t-1)}{t-1} \right) c_j y_{jt}^* \right\|_\infty \leq \sum_{j=1}^n \left| p_j - \frac{n_j(t-1)}{t-1} \right| \leq \min \left\{ (t-1)^{-\frac{1}{2}}, 1 \right\},
\]
where we use the fact that \( \|c_j\|_\infty \leq 1 \) from Assumption 1. This meets the definition of the event \( \mathcal{E}_t \) and this result is summarized in Lemma 5. To analyze the events \( A_1^{(j)} \) and \( B_1^{(j)} \), we can simply apply the concentration argument. In this way, we complete our analysis of the second component in (12).

**Lemma 5.** We have \( \left( \bigcap_{j=1}^n A_1^{(j)} \right) \cap \left( \bigcap_{j=1}^n B_1^{(j)} \right) \subseteq \mathcal{E}_t \) and the following inequality holds for each \( t = 1, \ldots, T \),
\[
\mathbb{P} \left( \tilde{\mathcal{E}}_t \right) \leq 2n \exp \left( -2L^2(t-1) \right) + 2n \exp \left( -\frac{2(t-1)^{1/2}}{n^2} \right).
\]
The proof is left in Appendix A.6.

### 3.4 Final Regret Bound

We can derive the final regret bound by combining Corollary 1, Proposition 2, Lemma 3, 4 and 5. The following theorem states that the regret is uniformly bounded in terms of \( T \). We defer the detailed proof to Appendix A.7. The regret bound is related with parameters such as the number of customer/order types \( n \), the number of constraints \( m \) and the stability parameter \( L \) (in Lemma 1). In this sense, the regret should be interpreted as a problem-dependent bound rather than a worst-case bound. The implication is that when the underlying LP is well-posed such that the parameter \( L \) can be treated as a
constant, the algorithm’s regret does not scale up with the number of time horizon $T$. We also remark that the theorem requires Assumption 3 where all the constraints are binding. We make the assumption to better illuminate the analysis of the underlying constraint process. In Appendix C, we will remove the assumption and study the general setting where there exist both binding and non-binding constraints. We note that the proof idea and the regret bound for the general setting without Assumption 3 are the same as the case of all binding constraints in this section.

**Theorem 2.** Under Assumptions 1, 2 and 3, the regret of Algorithm 1 satisfies

$$\text{Reg}_T^\pi \leq \frac{(48m + 4n + 12)}{L^2} \cdot \|\mathbf{x}^*\|_1 + o(1).$$

where $\pi$ denotes the policy specified by Algorithm 1, $\mathbf{x}^*$ is the dual optimal solution of LP (7), and the last term $o(1) \to 0$ as $T \to \infty$.

4 Regret Analysis for Degenerate Case

Now we analyze the algorithm’s performance without the nondegeneracy assumption (Assumption 2). As noted by Bumpensanti and Wang (2020), the existing results on bounded regret (Jasin and Kumar, 2012; Wu et al., 2015) require the nondegeneracy assumption. The regret bounds therein, though bear no dependency in $T$, will become arbitrarily large when the LP is nondegenerate but nearly degenerate. In this section, we present a positive result on the performance of Algorithm 1 for the degenerate case. Specifically, we show that Algorithm 1 achieves a $O(\sqrt{T \log T})$ regret without the nondegeneracy assumption. The bound is nearly tight in that $\Omega(\sqrt{T \log T})$ lower bound for Algorithm 1 is established by Bumpensanti and Wang (2020) against both the fluid benchmark $\text{OPT}_D$ and hindsight benchmark.

We first provide some intuition of the analysis. Consider the following three LPs:

$$R_t := \max_{\mathbf{y}} \mathbf{\mu}_t^\top \mathbf{y} \quad \text{s.t.} \quad C_t \mathbf{y} \leq \mathbf{b}_t \quad 0 \leq \mathbf{y} \leq 1,$$

$$\bar{R}_t := \max_{\mathbf{y}} \mathbf{\mu}^\top \mathbf{y} \quad \text{s.t.} \quad C \mathbf{y} \leq \mathbf{b} \quad 0 \leq \mathbf{y} \leq 1,$$

$$\text{OPT}_D := \max_{\mathbf{y}} \mathbf{\mu}^\top \mathbf{y} \quad \text{s.t.} \quad C \mathbf{y} \leq \mathbf{b} \quad 0 \leq \mathbf{y} \leq 1,$$

where

$$\mathbf{\mu}_t = \left( \frac{n_1(t-1)}{t-1} \mu_1, \cdots, \frac{n_n(t-1)}{t-1} \mu_n \right)^\top,$$

$$C_t = \left( \frac{n_1(t-1)}{t-1} c_1, \cdots, \frac{n_n(t-1)}{t-1} c_n \right),$$

$$\mathbf{\mu} = (p_1 \mu_1, \cdots, p_n \mu_n)^\top, \quad C = (p_1 c_1, \cdots, p_n c_n).$$

We note that the left LP in above is the sampled LP (5) used in Algorithm 1, while the right LP is the deterministic LP (3) with optimal objective value $\text{OPT}_D$. It is easy to see that $R_t$ is the expected reward (conditional on the history) that Algorithm 1 collects at time $t$. Then the single-period regret at time $t$ follows

$$\text{OPT}_D - R_t = (\text{OPT}_D - \bar{R}_t) + (\bar{R}_t - R_t).$$

To analyze the right-hand-side, we define for $2 \leq t \leq T - 1$,

$$C_t^{(i)} = \left\{ b_{it} > b_i - \frac{\sqrt{4n \log 2T}}{\sqrt{t}} - \frac{\sqrt{4n \log 2T} + \sqrt{2 \log 2T}}{\sqrt{T-1}} \right\},$$

$$D_t^{(j)} = \left\{ \frac{n_j(t-1)}{(t-1)p_j} - 1 < \frac{\sqrt{\log 2T}}{\sqrt{2p^2(t-1)}} \right\},$$

where $p = \min\{p_1, \ldots, p_n\}$. 
To motivate the definition of these events, first, given the events $C^i_t$'s, $b_t$ is close to $b$, and consequently, $\text{OPT}_D - \bar{R}_t$ is small. Second, given the events $D^j_t$'s, the left LP and the middle LP in above are close to each other, and thus, $\bar{R}_t - R_t$ is small.

The following lemma establishes that the above events will happen with high probability. The analysis of the event $C^i_t$ essentially reduces to the analysis of the constraint process, which is in a similar spirit as that of the previous section. The analysis of the event $D^j_t$ is simply based on a concentration argument. The proof can be found in Appendix B.1.

**Lemma 6.** We have

$$
P\left(\bigcap_{i=1}^m C^i_t\right) \geq 1 - \frac{m}{T}
$$

and

$$
P\left(\bigcap_{j=1}^n D^j_t\right) \geq 1 - \frac{n}{T}.
$$

Denote $C_t = \bigcap_{i=1}^m C^i_t$ and $D_t = \bigcap_{j=1}^n D^j_t$. We formalize the intuition above into the following lemma, and leave the proof in Appendix B.2.

**Lemma 7.** Under event $C_t \cap D_t$, we have the single-period regret

$$
\text{OPT}_D - R_t \leq \max\{1, \bar{\lambda}\} \cdot \left(m \left(\frac{\sqrt{4n \log 2T}}{\sqrt{t}} + \frac{\sqrt{4n \log 2T} + \sqrt{21 \log 2T}}{\sqrt{T - t}}\right) + n \frac{\sqrt{\log 2T}}{\sqrt{T}}\right),
$$

where $\bar{\lambda} := \max\{||\lambda||_\infty : \lambda \in FD_0\}$ and $FD_0$ denotes the set of basic feasible solutions for the dual of $DLP$ (3).

Combining the above two lemmas, we yield the regret bound without the nondegeneracy as follows. The theorem provides a regret bound sublinear in $T$ for Algorithm 1 without the nondegeneracy assumption. We remark that the analysis also covers the case when the distribution is known: in that case, the algorithm will use the LP in the middle $\bar{R}_t$ to guide the online decision, and thus the event $D^j_t$ will happen with probability 1. In contrast, the previous bounds in Jasin and Kumar (2012) and Wu et al. (2015) will be arbitrarily large when the underlying LP approaches a nondegenerate one. The key difference between our analysis and the previous analyses is the focus on the constraint process. On one hand, the definition of the events $C^i_t$ imposes a condition under which a sublinear regret is achievable. On the other hand, martingale tools enable a careful analysis of the constraint process to meet such condition.

**Theorem 3.** Under Assumption 1, Algorithm 1 gives a regret upper bounded by

$$
\text{Reg}_T^\pi \leq \max\{1, \bar{\lambda}\} \cdot \left(m \left(\sqrt{2} + \sqrt{16n}\right) + \frac{n}{\sqrt{22^2}}\right) \sqrt{T \log 2T} + 1 + n + m
$$

$$
= O((m\sqrt{n} + n)\sqrt{T \log T}).
$$

The proof of the theorem is deferred to Appendix B.3. Apart from the $\Omega(\sqrt{T})$ lower bound result in Bumpensanti and Wang (2020), we can also understand from the analysis the reason why bounded regret cannot be achieved for the degenerate case. In the analysis for the degenerate case, the regret decomposition in Proposition 1 no longer holds and thus we have to rely on the single-period regret bound in Lemma 7. The single-period regret bound is looser in a sense that it treats all the time period separately. Specifically, in the nondegenerate case, if the algorithm performs poorly in one time period, it may recover with better reward later through the adaptive (re-solving) mechanism. But this recovering mechanism cannot be captured when we treat all the time periods separately.
5 Numerical Experiment and Discussions

We conclude with numerical experiments to illustrate our analysis. We perform our simulation experiments under both degenerate and nondegenerate cases, and the results are consistent with our theoretical findings. Specifically, we consider the following problem instance. There are two types of resources and three types of customer orders. The unknown probability vector of three order types is 
\[(p_1, p_2, p_3) = (0.3, 0.3, 0.4),\]
the reward vector is 
\[(\mu_1, \mu_2, \mu_3) = (1, 1.2, 0.8),\]
and the resource consumption is 
\[c_1 = (1, 2)^	op, c_2 = (2, 1)^	op, c_3 = (1, 1)^	op.\]
Thus the underlying LP is
\[
\begin{align*}
\text{max} & \quad 0.3y_1 + 0.36y_2 + 0.32y_3 \\
\text{s.t.} & \quad 0.3y_1 + 0.6y_2 + 0.4y_3 \leq b_1 \\
& \quad 0.6y_1 + 0.3y_2 + 0.4y_3 \leq b_2 \\
& \quad 0 \leq y_j \leq 1, \quad j = 1, 2, 3.
\end{align*}
\]
(14)

For a nondegenerate problem instance, we set \((b_1, b_2) = (1, 1)\); for a degenerate problem instance, we set \((b_1, b_2) = (1, 1.15)\).

Figure 1 describes the relationship between the regret and the horizon length \(T\) under both nondegenerate and degenerate cases. It reports both the sample mean and the 99%-confidence interval (in a light color). When the underlying LP is nondegenerate, the regret does not scale with \(T\), but when the LP is degenerate, the regret grows on the order of \(O(\sqrt{T})\).

Furthermore, we compare the performance of Algorithm 1 for the cases of with and without distributional knowledge under a nondegenerate problem instance. The result is presented in Figure 2. The first two plots report the mean and variance computed based on 200 independent trials. The last plot displays the histogram of the difference of the regrets under the two cases. The histogram is generated under 800 independent trials with horizon \(T = 1000\). For the case when we know the underlying distribution, Algorithm 1 is implemented based on replacing the estimate with the true distribution when solving the LP (5).

This result further substantiates the theme of our discussion: the learning of the distribution (under the finite support and nondegeneracy condition) will not affect the regret’s dependency on the horizon \(T\). The algorithm performances under the cases of known and unknown distribution match in both expectation and variance. In addition, the difference of their performances is also symmetrically distributed as seen from the histogram.

Lastly, we make the following remarks:

- Partial acceptance: In the algorithm and the analysis, we focus on the case of binary decision where
the decision maker either accepts or rejects the order, \( x_t = 0 \) or \( 1 \). If partial acceptance is allowed, an alternative decision rule is to let \( x_t = y^*_j t \) in Algorithm 1. Compared to the binary decision rule, this partial rule has one less layer of randomness but no other difference. In that case, all the analyses and regret bounds in this paper still hold.

- **Known distribution:** In this paper, we focus on the case when the underlying distribution is unknown. We remark that our analyses can also be applied to the case when the distribution is known. As noted earlier, the algorithm with known distribution replaces all the estimators in Algorithm 1 with their true parameters. For its analysis, all the probability events related to parameter estimation will happen with probability one and all the remaining analyses in our paper still go through.

- **Infrequent re-solving:** Several existing works consider the case of known distribution and discuss the infrequent re-solving scheme where the adaptive LP (5) in Algorithm 1 is solved on an infrequent basis (see, for example, Reiman and Wang (2008); Agrawal et al. (2014); Jasin (2015); Bumpensanti and Wang (2020)). While our paper discusses the case of unknown distribution, we believe the infrequent re-solving scheme is still compatible with our analysis at least for the nondegenerate case. Specifically, in the proof of Proposition 4, we identify the number of observations needed to accurately estimate the optimal basis of the underlying LP. After these number of observations, the underlying distribution is estimated accurately enough and we can then adopt the algorithm proposed by Bumpensanti and Wang (2020) using the estimated distribution. The re-solving time points can be chosen to meet the condition in \( E_t \), and then the analysis of the constraint process still holds. While the goal of our paper is mainly to provide a thorough analysis for the classic version of the algorithm with a frequent re-solving scheme, we believe such investigation is interesting and deserves future study.

- **More efficient algorithm:** Algorithm 1 can be computationally costly as it solves a linear program in every time period, so do the infrequent re-solving algorithms when the underlying LP is large-scale. A few recent works study the more efficient algorithms to solve the problem. Li et al. (2020) and Balseiro et al. (2020) both perform gradient descent in the dual space and use the dual solution to guide the primal decisions. Sun et al. (2020) propose a gradient descent version of the infrequent re-solving algorithm in Bumpensanti and Wang (2020). Gupta (2021) adapts the renowned Sum-of-Squares algorithm (originally for bin packing problem) and develops an efficient algorithm that achieves bounded regret with known distribution and nondegeneracy. Though originally motivated from a computational consideration, these works provide new perspectives for both algorithm design and analysis.
References

Acimovic, Jason, Stephen C Graves. 2015. Making better fulfillment decisions on the fly in an online retail environment. *Manufacturing & Service Operations Management* 17(1) 34–51.

Agrawal, Shipra, Zizhuo Wang, Yinyu Ye. 2014. A dynamic near-optimal algorithm for online linear programming. *Operations Research* 62(4) 876–890.

Arlotto, Alessandro, Itai Gurvich. 2019. Uniformly bounded regret in the multisecretary problem. *Stochastic Systems*.

Asadpour, Arash, Xuan Wang, Jiawei Zhang. 2019. Online resource allocation with limited flexibility. *Management Science*.

Badanidiyuru, Ashwinkumar, Robert Kleinberg, Aleksandrs Slivkins. 2013. Bandits with knapsacks. *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 207–216.

Balseiro, Santiago, Omar Besbes, Dana Pizarro. 2021. Survey of dynamic resource constrained reward collection problems: Unified model and analysis. *Available at SSRN 3963265*.

Balseiro, Santiago, Haihao Lu, Vahab Mirrokni. 2020. The best of many worlds: Dual mirror descent for online allocation problems. *arXiv preprint arXiv:2011.10124*.

Banerjee, Siddhartha, Daniel Freund. 2020. Uniform loss algorithms for online stochastic decision-making with applications to bin packing. *Abstracts of the 2020 SIGMETRICS/Performance Joint International Conference on Measurement and Modeling of Computer Systems*. 1–2.

Banerjee, Siddhartha, Itai Gurvich, Alberto Vera. 2020. Constant regret in online allocation: On the sufficiency of a single historical trace.

Besbes, Omar, Alp Muharremoglu. 2013. On implications of demand censoring in the newsvendor problem. *Management Science* 59(6) 1407–1424.

Bray, Robert. 2019. Does the multisecretary problem always have bounded regret? *Available at SSRN 3497056*.

Buchbinder, Niv, Joseph Naor. 2009. Online primal-dual algorithms for covering and packing. *Mathematics of Operations Research* 34(2) 270–286.

Bumpensanti, Pornpawee, He Wang. 2020. A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Science*.

Conforti, Michele, Gérard Cornuéjols, Giacomo Zambelli, et al. 2014. *Integer programming*, vol. 271. Springer.

Ferguson, Thomas S, et al. 1989. Who solved the secretary problem? *Statistical science* 4(3) 282–289.

Gupta, Varun. 2021. Greedy algorithm for multiway matching with bounded regret. *arXiv preprint arXiv:2112.04622*.

Jaksch, Thomas, Ronald Ortner, Peter Auer. 2010. Near-optimal regret bounds for reinforcement learning. *Journal of Machine Learning Research* 11 1563–1600.

Jasin, Stefanus. 2015. Performance of an lp-based control for revenue management with unknown demand parameters. *Operations Research* 63(4) 909–915.
Jasin, Stefanus, Sunil Kumar. 2012. A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research* **37**(2) 313–345.

Jasin, Stefanus, Amitabh Sinha. 2015. An lp-based correlated rounding scheme for multi-item ecommerce order fulfillment. *Operations Research* **63**(6) 1336–1351.

Kellerer, Hans, Ulrich Pferschy, David Pisinger. 2003. Knapsack problems. 2004.

Kerimov, Süleyman, Itai Ashlagi, Itai Gurvich. 2021a. Dynamic matching: Characterizing and achieving constant regret. *Available at SSRN 3824407*.

Kerimov, Süleyman, Itai Ashlagi, Itai Gurvich. 2021b. On the optimality of greedy policies in dynamic matching. *Available at SSRN 3918497*.

Kesselheim, Thomas, Andreas Tönnis, Klaus Radke, Berthold Vöcking. 2014. Primal beats dual on online packing lps in the random-order model. *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. ACM, 303–312.

Li, Xiaocheng, Chunlin Sun, Yinyu Ye. 2020. Simple and fast algorithm for binary integer and online linear programming. *Advances in Neural Information Processing Systems* **33** 9412–9421.

Li, Xiaocheng, Yinyu Ye. 2021. Online linear programming: Dual convergence, new algorithms, and regret bounds. *Operations Research*.

Mangasarian, Olvi L, T-H Shiau. 1987. Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems. *SIAM Journal on Control and Optimization* **25**(3) 583–595.

Megiddo, Nimrod, Ramaswamy Chandrasekaran. 1989. On the ε-perturbation method for avoiding degeneracy. *Operations Research Letters* **8**(6) 305–308.

Mehta, Aranyak, Amin Saberi, Umesh Vazirani, Vijay Vazirani. 2005. Adwords and generalized on-line matching. *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)*. IEEE, 264–273.

Reiman, Martin I, Qiong Wang. 2008. An asymptotically optimal policy for a quantity-based network revenue management problem. *Mathematics of Operations Research* **33**(2) 257–282.

Shivaswamy, Pannagadatta, Thorsten Joachims. 2012. Multi-armed bandit problems with history. *Artificial Intelligence and Statistics*. 1046–1054.

Stein, Clifford, Van-Anh Truong, Xinshang Wang. 2020. Advance service reservations with heterogeneous customers. *Management Science* **66**(7) 2929–2950.

Sun, Rui, Xinshang Wang, Zijie Zhou. 2020. Near-optimal primal-dual algorithms for quantity-based network revenue management. *Available at SSRN 3728397*.

Talluri, Kalyan T, Garrett J Van Ryzin. 2006. *The theory and practice of revenue management*, vol. 68. Springer Science & Business Media.

van de Geer, Sara A. 2002. On hoeffding’s inequality for dependent random variables. *Empirical process techniques for dependent data*. Springer, 161–169.

Vera, Alberto, Alessandro Arlotto, Itai Gurvich, Eli Levin. 2020. Dynamic resource allocation: The geometry and robustness of constant regret.
We first state three results that will be used in the proof.

**Lemma 8.** Denote \( \Xi_d = \{ q \in \mathbb{R}^d | \sum_{j=1}^d q_j = 1, 0 < q_j < 1, j = 1, \ldots, d \} \). Let \( \text{Mul}(t, q) \) to be the multi-nomial distribution. For \( q \in \Xi_d \) and \( \hat{q} \sim \frac{1}{t} \cdot \text{Mul}(t, q) \), the inequality

\[
P(||\hat{q} - q||_1 \geq \epsilon) \leq 2^d \exp\left(-\frac{t \epsilon^2}{2}\right)
\]

holds for any \( \epsilon > 0 \).

**Proof:** See Appendix C1 in Jaksch et al. (2010).

**Lemma 9.** (Hoeffding) Let \( X_1, \ldots, X_T \) be independent random variables such that \( a_t \leq X_t \leq b_t \) almost surely, we have

\[
P\left(\sum_{t=1}^T X_t - \sum_{t=1}^T \mathbb{E}[X_t] \geq \epsilon\right) \leq 2 \exp\left(-\frac{2 \epsilon^2}{\sum_{t=1}^T (b_t - a_t)^2}\right).
\]

**Lemma 10.** (Azuma-Hoeffding) Let \( S_t \) be a martingale such that \( S_0 = 0 \) and the increment \( S_t - S_{t-1} \) is bounded by \( \sigma_t \) with probability 1, then for any \( \epsilon > 0 \) and \( T \geq 2 \), we have

\[
P(S_T \geq \epsilon) \leq \exp\left(-\frac{\epsilon^2}{2 \sum_{t=1}^T \sigma_t^2}\right).
\]
A.1 Proof of Proposition 1

We have

\[ \text{Reg}_T^\tau = \mathbb{E} \left[ T \cdot \text{OPT}_D - \sum_{t=1}^{\tau} r_t x_t \right] = \mathbb{E} \left[ \lambda^T B + \sum_{t=1}^{\tau} (r_t - a_t^T \lambda^*)_+ \right] - \sum_{t=1}^{\tau} \mathbb{E} [r_t x_t] \]

Substituting the duality of the DLP (3), we get

\[ = \mathbb{E} \left[ \lambda^T B + \sum_{t=1}^{\tau} (r_t - a_t^T \lambda^*)_+ \right] - \sum_{t=1}^{\tau} \mathbb{E} [(r_t - a_t^T \lambda^*) x_t + a_t^T \lambda^* x_t] \]

\[ = \mathbb{E} \left[ \lambda^T \left( B - \sum_{t=1}^{\tau} a_t x_t \right) \right] + \mathbb{E} \left[ \sum_{t=1}^{\tau} (r_t - a_t^T \lambda^*)_+ - \sum_{t=1}^{\tau} (r_t - a_t^T \lambda^*) x_t \right] \]

\[ = \lambda^T \mathbb{E} [B] + \mathbb{E} \left[ \sum_{t=1}^{\tau} (r_t - a_t^T \lambda^*)_+ - \sum_{t=1}^{\tau} (r_t - a_t^T \lambda^*) x_t \right] + \mathbb{E} \left[ \sum_{t=1}^{\tau} (a_t^T \lambda^* - r_t)_+ x_t \right] \]

\[ = \lambda^T \mathbb{E} [B] + \mathbb{E} \left[ \sum_{j \in \mathcal{J}^*} (\mu_j - c_j^T \lambda^*) \cdot (n_j(T) - n_j^\tau(\tau)) \right] + \mathbb{E} \left[ \sum_{j \in \mathcal{J}^*} (c_j^T \lambda^* - \mu_j) \cdot n_j^\tau(\tau) \right], \]

where the second line comes from duality of the DLP (3), and the last line comes from the definition of \( B^\tau \) and the definition of optimal basis \( \mathcal{J}^* \) and its compliment \( \mathcal{J}' \).

A.2 Proof of Corollary 1

Recall Proposition 1 that

\[ \text{Reg}_T^\tau = \sum_{j \in \mathcal{J}^*} (\mu_j - c_j^T \lambda^*) \cdot \mathbb{E} [n_j(T) - n_j^\tau(\tau)] + \sum_{j \in \mathcal{J}^*} (c_j^T \lambda^* - \mu_j) \cdot \mathbb{E} [n_j^\tau(\tau)] + \lambda^T \cdot \mathbb{E} [B^\tau]. \]

Since \( \tau' \leq \tau \), we know \( n_j^\tau(\tau') \leq n_j^\tau(\tau) \) for all \( j \). If we apply this to the above equality, we obtain

\[ \text{Reg}_T^\tau = \sum_{j \in \mathcal{J}^*} (\mu_j - c_j^T \lambda^*) \cdot \mathbb{E} [n_j(\tau') - n_j^\tau(\tau')] + \sum_{j \in \mathcal{J}^*} (c_j^T \lambda^* - \mu_j) \cdot \mathbb{E} [n_j^\tau(\tau')] + \lambda^T \cdot \mathbb{E} [B^\tau]. \]

\[ \leq \sum_{j \in \mathcal{J}^*} (\mu_j - c_j^T \lambda^*) \cdot \mathbb{E} [n_j(\tau') - n_j^\tau(\tau')] + \sum_{j \in \mathcal{J}^*} (c_j^T \lambda^* - \mu_j) \cdot \mathbb{E} [n_j^\tau(\tau')] + \lambda^T \cdot \mathbb{E} [B^\tau]. \]

\[ + \sum_{j \in \mathcal{J}^*} (\mu_j - c_j^T \lambda^*) \cdot \mathbb{E} [n_j(T) - n_j(\tau')] + \sum_{j \in \mathcal{J}^*} (c_j^T \lambda^* - \mu_j) \cdot \mathbb{E} [n_j^\tau(\tau) - n_j^\tau(\tau')] + \lambda^T \cdot \mathbb{E} [B^\tau]. \]
To bound the leftover terms compared to the statement of the corollary,

\[
\sum_{j \in J^*} (\mu_j - c_j^T \lambda^*) \cdot \mathbb{E} [n_j(T) - n_j(\tau')] + \sum_{j \in J'} (c_j^T \lambda^* - \mu_j) \cdot \mathbb{E} [n_j^0(\tau) - n_j^0(\tau')] + \lambda^* \cdot \mathbb{E} [B_\tau - B_{\tau'}]
\]

\[
\leq \max_{j \in \mathbb{N}} |\mu_j - c_j^T \lambda^*| \cdot \mathbb{E} \left[ \sum_{j \in J^*} (n_j(T) - n_j(\tau')) + \sum_{j \in J'} (n_j^0(\tau) - n_j^0(\tau')) \right]
\]

\[
\leq \max_{j \in \mathbb{N}} |\mu_j - c_j^T \lambda^*| \cdot \mathbb{E} \left[ \sum_{j \in J^*} (n_j(T) - n_j(\tau')) + \sum_{j \in J'} (n_j(T) - n_j(\tau')) \right]
\]

\[
= \max_{j \in \mathbb{N}} |\mu_j - c_j^T \lambda^*| \cdot \mathbb{E} [T - \tau],
\]

where the first inequality comes from the fact that since \( \tau' \leq \tau \), we have \( B_\tau \leq B_{\tau'} \) entry-wisely. The second inequality is due to \( n_j^0(\tau) - n_j^0(\tau') \leq n_j(\tau) - n_j(\tau') \leq n_j(T) - n_j(\tau') \), and the intuition is that for a certain order type, within a certain time window, the number of occurrence will always be greater than the number of acceptance. The last line is because of the definition of \( n_j \) such that \( \sum_{j=1}^n n_j(t) = t \) for \( t > 0 \). Combining the above inequalities finishes the proof.

### A.3 Proof of Proposition 2

It suffices to bound (i) the term \( \mathbb{E} [n_j^0(\tau_s)] \) for \( j \in J' \) and (ii) the term \( \mathbb{E} [n(\tau_s) - n_j^0(\tau_s)] \) for \( j \in J^* := \{ j : \mu_j > c_j^T \lambda^* \} \).

Consider the event (the same as its definition in the main paper)

\[
\mathcal{A}_t^{(j)} = \left\{ \left| \frac{n_j(t) - 1}{t - 1} - p_j \right| \leq L \right\},
\]

and with the convention \( \mathcal{A}_t^{(j)} = \Omega \).

We first bound \( \mathbb{E} [n_j(\tau_s)] \) where \( j \in J' \). Notice that

\[
\mathbb{E} [n_j(\tau_s)] \leq \mathbb{E} \left[ \sum_{t=1}^T I(\{\text{Accept order type } j \in J' \text{ at time } t\} \cap \{ t < \tau_s \}) \right]
\]

\[
\leq \sum_{t=1}^T \mathbb{P}(\{\text{Accept order type } j \in J' \text{ at time } t\} \cap \{ t < \tau_s \}).
\]

By the definition of \( \mathcal{A}_t^{(j)} \) and \( L \), for \( j \in J' \), under the event \( \{ t < \tau_s \} \cap \{ \cap_{j=1}^n \mathcal{A}_t^{(j)} \} \), we have that when solving the LP (5)

\[
\left| \frac{n_j(t - 1)}{t - 1} c_j - p_j c_j \right| \leq L \quad \text{and} \quad \left| \frac{n_j(t - 1)}{t - 1} \mu_j - p_j \mu_j \right| \leq L.
\]

The above inequalities combined with the fact \( \{ t < \tau_s \} \) ensures the condition in Lemma 1. Thus we know that the perturbed LP (5) share the same optimal basis with DLP (3). The probabilistic decision elicited by the algorithm based on (5) will then reject the order type \( j \in J' \). Therefore we have

\[
\left\{ \left\{ \text{Accept order type } j \in J' \text{ at time } t \right\} \cap \{ t < \tau_s \} \right\} \subseteq \left\{ \bigcup_{j=1}^n \mathcal{A}_t^{(j)} \right\} \cap \{ t < \tau_s \}.
\]
Thus, we have for $j \in J'$

$$
E \left[ n_j^a(\tau_S) \right] \leq \sum_{t=1}^{T} P(\{\text{Accept order type } j \in J' \text{ at time } t \} \cap \{t < \tau_S\})
$$

$$
\leq \sum_{t=1}^{T} P \left( \bigcup_{j=1}^{n} \bar{A}_t^{(j)} \right) \cap \{t < \tau_S\}
$$

$$
\leq \sum_{t=1}^{T} P \left( \bigcup_{j=1}^{n} \bar{A}_t^{(j)} \right) \leq \sum_{t=1}^{T} \sum_{j=1}^{n} P(\bar{A}_t^{(j)})
$$

$$
\leq n + \sum_{t=1}^{T-1} 2n \exp \left(-2tL^2\right) \leq \frac{2n}{1 - \exp(-2L^2)}.
$$

Here the first two lines come from the above argument, the third line comes from the union bound, and
the last line comes from concentration inequality (Lemma 9).

Next, we start to bound $E \left[ n(\tau_S) - n_j^a(\tau_S) \right]$ for $j \in J'_0 = \{j : \mu_j > c^T_j \lambda^*\}$. In a similar way as
before, under the event $\{t < \tau_S\} \cap \{\bigcap_{j=1}^{n} \bar{A}_t^{(j)}\}$, the LP (5) shares the same optimal basis with DLP
(3). Therefore,

$$
\{\{\text{Reject order type } j \in J'_0 \text{ at time } t\} \cap \{t < \tau_S\}\} \subseteq \left\{\bigcup_{j=1}^{n} \bar{A}_t^{(j)} \cap \{t < \tau_S\}\right\}.
$$

Then, we have for $j \in J'_0$,

$$
E \left[ n(\tau_S) - n_j^a(\tau_S) \right] \leq E \left[ \sum_{t=1}^{T} I(\{\text{Reject order type } j \in J'_0 \text{ at time } t \} \cap \{t < \tau_S\}) \right]
$$

$$
\leq \sum_{t=1}^{T} P(\{\text{Reject order type } j \in J'_0 \text{ at time } t \} \cap \{t < \tau_S\})
$$

$$
\leq \sum_{t=1}^{T} P \left( \bigcup_{j=1}^{n} \bar{A}_t^{(j)} \cap \{t < \tau_S\} \right)
$$

$$
\leq \frac{2n}{1 - \exp(-2L^2)},
$$

thereby finishing the proof.

### A.4 Proof of Lemma 3

From definition we know that

$$
P(\tau_S < t) = P(b_s \notin \mathcal{B} \text{ for some } s \leq t),
$$

and since $\tau_S \geq 0$, we know

$$
E[\tau_S] = \sum_{t=1}^{+\infty} P(\tau_S \geq t) = \sum_{t=1}^{+\infty} (1 - P(\tau_S < t)) \geq \sum_{t=1}^{+\infty} (1 - P(\tau_S \leq t)) = T - \sum_{t=1}^{+\infty} P(b_s \notin \mathcal{B} \text{ for some } s \leq t).
$$

Next, to finish the proof, notice that $\max_{j \in [n]} |\mu_j - c^T_j \lambda^*| \leq ||\lambda||_1$, and

$$
\lambda^*^T E[B_{\tau_S}] \leq \lambda^*^T E[(T - \tau_S + 2)b_{\tau_S - 1}] \leq \lambda^*^T(b + L) E[(T - \tau_S + 2)] \leq 2||\lambda^*||_1(T + 2 - E[\tau_S]),
$$
where the last inequality comes from the fact that for all the binding resources \( i \) we will have initial average inventory \( b_i \leq 1 \), for all the nonbinding resource \( i \) we know \( \lambda_i^* = 0 \), and lastly we can assume \( L \leq 1 \).

A.5 Proof of Lemma 4

We first analyze the process \( \tilde{b}_{i,t} \), the \( i \)-th component of \( \tilde{b}_t \), and then take union bound with respect to \( i \).

Define

\[
H_t = \{(r_s,a_s)\}_{s=1}^t \quad \text{for} \quad t = 1,\ldots,T.
\]

Let

\[
Y_t := \tilde{b}_{i,t+1} - \tilde{b}_{i,t}
\]

for \( t \geq 1 \) and

\[
X_t := Y_t - \mathbb{E}[Y_t|H_{t-1}].
\]

In this way, to analyze the process \( \tilde{b}_{i,t} \), we can equivalently analyze the summation \( \sum_{s=1}^{t-1} Y_t \). From the definition of the process \( \tilde{b}_t \), we know that when \( t \geq \tilde{\tau} - 1 \), we have

\[
\tilde{b}_{i,t+1} = \tilde{b}_{i,t},
\]

and when \( 1 \leq t < \tilde{\tau} \), we have

\[
\tilde{b}_{i,t+1} = \tilde{b}_{i,t} + \frac{1}{T-t}(\tilde{b}_{i,t} - a_{i,t}x_t).
\]

From the fact that \( \tilde{b}_{i,t} \) is \( H_{t-1} \)-measurable, we can bound the absolute value of \( |x_t| \) such that

\[
|X_t| = \left| \frac{1}{T-t}(\tilde{b}_{i,t} - a_{i,t}x_t) - \mathbb{E} \left[ \frac{1}{T-t}(\tilde{b}_{i,t} - a_{i,t}x_t) \bigg| H_{t-1} \right] \right| \\
= \frac{1}{T-t} |\mathbb{E}[a_{i,t}x_t|H_{t-1}] - a_{i,t}x_t| \leq \frac{1}{T-t}
\]

for each \( t \leq T - 1 \). So we can define \( L_t \) and \( U_t \) as

\[
L_t := -\frac{1}{T-t}, \\
U_t := \frac{1}{T-t},
\]

and the conditions of Theorem 1 are met for the process \( X_t, L_t \) and \( U_t \). Then as in Theorem 1,

\[
V_t = \sum_{s=1}^{t}(U_s - L_s)^2 = \sum_{s=1}^{t} \frac{4}{(T-s)^2} \leq \frac{4}{T-t-1}
\]

for \( t = 1,\ldots,T-2 \). From Theorem 1, we know that

\[
P \left( \sum_{j=1}^{s} X_j \geq \Delta \text{ for some } s \leq t \right) \leq 2e^{-\frac{\Delta^2}{4(T-t-1)}} 
\]  

holds for all \( \Delta > 0 \) and \( t \leq T - 2 \).

With this bound on the summation of \( X_t \), we return to analyze the summation of \( Y_t \) by bounding
the difference between these two sequences. By the definition, we have

\[
|X_t - Y_t| = |E[Y_t | \mathcal{H}_{t-1}]|
\]

\[
= |E[\hat{b}_{i,t+1} - \hat{b}_{i,t} | \mathcal{H}_{t-1}]|
\]

\[
= \frac{1}{T-t} E[(a_{i,t} x_t - \hat{b}_{i,t}) I(t < \tau) | \mathcal{H}_{t-1}] + \frac{1}{(T-t) t^{1/4}} I(t > \kappa T)
\]

\[\leq \epsilon_t \frac{T-t}{T-t} = \frac{1}{T-t} I(t \leq \kappa T) + \frac{1}{(T-t) t^{1/4}} I(t > \kappa T) \tag{17}\]

for \(1 \leq t \leq T-1\). The second line comes from the definition of \(Y_t\). The third line comes from splitting the second line with two indicators \(I(t < \tau)\) and \(I(t \geq \tau)\). The process \(\tilde{b}_t\) freezes and \(\tilde{b}_{i,t+1} = \tilde{b}_{i,t}\) for \(t \geq \tau\). The second last line comes from the definition of \(E_t\) and the definition of \(\tau\), and the last line comes from the definition of \(\epsilon_t\).

By taking summation of (17), we have for \(s \leq T-2\),

\[
\left| \sum_{j=1}^{s} X_j - \sum_{j=1}^{s} Y_j \right| \leq \sum_{j=1}^{\kappa T} \frac{1}{T-t} + \sum_{j=\kappa T+1}^{s} \frac{1}{(T-j) j^{1/4}}.
\]

The next is to find a proper value of \(\kappa\) such that the equation above is bounded by \(\frac{\Delta}{2}\). For the first part, we have

\[
\sum_{j=1}^{\kappa T} \frac{1}{T-t} \leq \int_{\kappa T+1}^{T-1} \frac{1}{x} dx = \log \left( \frac{T-1}{T-\kappa T-1} \right)
\]

\[
\leq \log \left( \frac{T-1}{T-\kappa T-2-2\kappa} \right) = \log \left( \frac{T-1}{T-2} \right) - \log (1-\kappa).
\]

For the second part,

\[
\sum_{j=\kappa T+1}^{T-1} \frac{1}{(T-j) j^{1/4}} \leq \frac{1}{(\kappa T)^{1/4}} \sum_{j=\kappa T+1}^{T-1} \frac{1}{T-j} \leq \frac{\log T}{(\kappa T)^{1/4}}.
\]

Henceforth, if we set \(\kappa = 1 - \exp \left( -\frac{\Delta}{2} \right)\) and define \(T_1\) as the minimal integer such that \(T_1 \geq \frac{1}{\exp \left( \frac{\Delta}{2} \right)-1} + 2\)

and \(\frac{\log T_1}{T_1^{1/4}} \leq \frac{\kappa^{1/4} \Delta}{4}\), then the following inequality holds for \(T \geq T_1\)

\[
\sum_{j=1}^{\kappa T} \frac{1}{T-t} + \sum_{j=\kappa T+1}^{T-1} \frac{1}{(T-j) j^{1/4}} \leq \Delta 4 + \Delta 4 = \Delta 2.
\]

With the choice of \(\kappa\) and \(T \geq T_1\), we have

\[
\left| \sum_{j=1}^{s} X_j - \sum_{j=1}^{s} Y_j \right| \leq \frac{\Delta}{2}
\]
holds almost surely. Consequently,
\[
\{ |\hat{b}_{i,s} - b_i| > \Delta \text{ for some } s \leq t \} = \left\{ \sum_{j=1}^{s} Y_j > \Delta \text{ for some } s \leq t \right\}
\]
\[
= \left\{ \sum_{j=1}^{s} Y_j > \Delta \text{ for some } s \leq t - 1 \right\}
\]
\[
\subseteq \left\{ \sum_{j=1}^{s} X_j > \Delta/2 \text{ for some } s \leq t - 1 \right\}.
\]
Therefore, if we apply union bound with respect to constraint index \(i = 1, \ldots, m\), we have
\[
P\left( \hat{b}_s \notin m_{i=1}^{m} [b_i - \Delta, b_i + \Delta] \text{ for some } s \leq t \right) \leq 2me^{-\frac{\Delta^2(t-\Omega)}{8}}
\]
for \(t \leq T - 2\) and \(T \geq T_1\).

A.6 Proof of Lemma 5

In the statement before Lemma 5, we have already shown \(\bigcap_{j=1}^{n} A_{t}^{(j)} \cap \bigcap_{j=1}^{m} B_{t}^{(j)} \subseteq \mathcal{E}_t\) for \(t \geq 2\), and we just have to discuss the special case that \(\mathcal{E}_1 = \Omega\). From Algorithm 1, we know that when \(t = 1\), we will always accept the order, and have \(b = b_1\) with probability 1. Therefore,
\[
\mathbb{E}[a_1x_1(b)|\mathcal{H}_0, b_1 = b] - b = \mathbb{E}[a_1] - b.
\]
Then notice that we have \(-b_i \leq \mathbb{E}[a_1] - b_i \leq 1 - b_i\) for all \(i\), because \(||c_j||_{\infty} \leq 1\) for all \(j\). Moreover, all resource being binding implies that \(0 \leq b_i \leq 1\). Combining these facts we can have
\[
||\mathbb{E}[a_1x_1(b)|\mathcal{H}_0, b_1 = b] - b||_{\infty} = ||\mathbb{E}[a_1] - b||_{\infty} \leq 1.
\]
Next, to show
\[
P(\mathcal{E}_t) \geq P\left( \left( \bigcap_{j=1}^{n} A_{t}^{(j)} \right) \cap \left( \bigcap_{j=1}^{m} B_{t}^{(j)} \right) \right) \geq 1 - \sum_{j=1}^{n} P\left( A_{t}^{(j)} \right) - \sum_{j=1}^{m} P\left( B_{t}^{(j)} \right),
\]
we analyze each component in the summation with Hoeffding’s inequality (Lemma 9). Specifically,
\[
n_j(t) = \sum_{s=1}^{t} I( (rs, a_s) = (\mu_j, c_j) )
\]
where \(I(\cdot)\) denotes the indicator function and \(I( (rs, a_s) = (\mu_j, c_j) )\)'s are i.i.d. random variables. In addition, \(\mathbb{E}[I( (rs, a_s) = (\mu_j, c_j) )] = p_j\) for each \(j = 1, \ldots, n\). Therefore, we have
\[
P\left( A_{t}^{(j)} \right) = P\left( \left| \frac{n_j(t-1)}{t-1} - p_j \right| > L \right) \leq 2 \exp \left( -2L^2(t-1) \right)
\]

28
and
\[ P( \bar{B}^{(j)} ) = P \left( \left| \frac{n_j(t-1)}{t-1} - p_j \right| > \frac{1}{n(t-1)^{1/4}} \right) \]
\[ \leq 2 \exp \left( \frac{2(t-1)^{1/2}}{n^2} \right). \]

Combining these two inequalities, we have
\[ P \left( \bigcap_{j=1}^n A^{(j)} t \cap \bigcap_{j=1}^n B^{(j)} t \right) \geq 1 - 2 \exp \left( -2L^2(t-1) \right) - 2n \exp \left( -\frac{2(t-1)^{1/2}}{n^2} \right). \]

### A.7 Proof of Theorem 2

We first state a lemma that takes a summation for both sides of the inequality in Lemma 5.

**Lemma 11.** The following inequality holds for \( T \in \mathbb{N}^+ \)
\[ \sum_{t=1}^T \sum_{s=\kappa T+1}^T P(\mathcal{E}_s) \leq \frac{nT}{L^2} \exp \left( -2L^2 \kappa T \right) + 4n^3T^{3/2} \exp \left( -\frac{1}{n^2}(\kappa T)^{1/2} \right). \]

**Proof:** Firstly, from Lemma 5,
\[ \sum_{t=1}^T \sum_{s=\kappa T+1}^T P(\mathcal{E}_s) \leq \sum_{t=1}^T \sum_{s=\kappa T+1}^T 2n \exp \left( -2L^2 s \right) + 2n \exp \left( -\frac{s^{1/2}}{n^2} \right) \]
\[ \leq \int_{\kappa T}^T \left( 2n \exp \left( -2L^2 s \right) + 2n \exp \left( -\frac{s^{1/2}}{n^2} \right) \right) ds. \]

For the second term observe that for any \( \alpha > 0, \beta, \kappa \in (0, 1) \) we can have the following bound
\[ \int_{\kappa T}^T e^{-\alpha x} dx \leq \frac{T^{1-\beta}}{\alpha \beta} \int_{\kappa T}^T x^{1-\beta} e^{-\alpha x} dx \leq \frac{T^{1-\beta}}{\alpha \beta} e^{-\alpha (\kappa T)^{\beta}}. \]

Combining (19) and (20) yields
\[ \sum_{t=1}^T \sum_{s=\kappa T+1}^T P(\mathcal{E}_s) \leq \frac{nT}{L^2} \exp \left( -2L^2 \kappa T \right) + 4n^3T^{3/2} \exp \left( -\frac{1}{n^2}(\kappa T)^{1/2} \right). \]

**Proof of Theorem 2.**

First we provide a slightly more careful analysis for (12). Recall the stopping time
\[ \tau_S := \min\{t \leq T : b_t \notin \mathcal{B} \} \cup \{T+1\}. \]

Then, for \( t = 1, ..., T \),
\[ P(\tau_S \leq t) = P( b_s \notin \mathcal{B} \text{ for some } s \leq t) \]
\[ \leq P( \bar{b}_s \notin \mathcal{B} \text{ for some } s \leq t) + \sum_{s=1}^t P((r_1, a_1, ..., r_{s-1}, a_{s-1}) \notin \mathcal{E}_s) \]
\[ = P( \bar{b}_s \notin \mathcal{B} \text{ for some } s \leq t) + \sum_{s=\kappa T+1}^t P((r_1, a_1, ..., r_{s-1}, a_{s-1}) \notin \mathcal{E}_s). \]

where the second line comes from (12) and the third line comes from the definition of \( \mathcal{E}_s \). By a "more
careful" analysis, it means that the second component takes into account the definition of $\epsilon_t$ and thus removes the first $\kappa T$ summands. Next, we are going to apply Corollary 1 to bound the regret.

With $T_1$ and $\kappa$ defined in Lemma 4 corresponding to $\Delta = L$, if $T \geq T_1$, we have

$$\mathbb{E}[\tau_S] \geq \sum_{t=1}^{T} (1 - \mathbb{P}(\tau_S \leq t))$$

$$\geq T - 2 - \sum_{t=1}^{T-2} \mathbb{P} \left( \tilde{b}_s \notin \mathcal{B} \text{ for some } s \leq t \right) - \sum_{t=1}^{T} \sum_{s=\kappa T+1}^{t} \mathbb{P}(\tilde{E}_s) \quad (21)$$

$$\geq T - 2 - 2m \frac{1 - e^{-(T-2)L^2/8}}{1 - e^{-L^2/8}} - \sum_{t=1}^{T} \sum_{s=\kappa T+1}^{t} \mathbb{P}(\tilde{E}_s)$$

where the last line applies Lemma 4 for the first summation. The following result applies Lemma 5 for the second summation.

By combining (21) with Lemma 11, and noting that

$$2m \frac{1 - e^{-(T-2)L^2/8}}{1 - e^{-L^2/8}} \leq \frac{16m}{L^2}, \quad \text{and} \quad \kappa = 1 - \exp \left( -\frac{L}{8} \right) \geq \frac{L}{32} \quad \text{when} \quad L \leq 1,$$

we have

$$\mathbb{E}[\tau_S] \geq T - 2 - \frac{16m}{L^2} \frac{T^3}{16} - 4n^3T^{3/2} \exp \left( -\frac{L^{1/2}}{6n^2} \cdot T^{1/2} \right).$$

Recall from Corollary 1 we have

$$\text{Reg}_T \leq \sum_{j \in J^*} (\mu_j - c_j^T \lambda^*) \cdot \mathbb{E} [n_j(\tau') - n_j^0(\tau')] + \sum_{j \in J^*} (c_j^T \lambda^* - \mu_j) \cdot \mathbb{E} [n_j^0(\tau')]$$

$$+ (T - \mathbb{E} [\tau^*]) \cdot \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| + \lambda^T \cdot \mathbb{E} [B_{\tau^*}].$$

If we substitute $\tau_S$ for this $\tau' \leq \tau$ and apply (21), Lemma 3 and 11, we have

$$(T - \mathbb{E}[\tau_S]) \cdot \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| = O \left( \frac{m + n}{L^2} + n^3T^{3/2} \exp \left( -\frac{T^{1/2}}{n^2} \right) \right),$$

$$\lambda^T \mathbb{E} [B_{\tau_S}] \leq 2||\lambda^*||_1 \left( 4 + \frac{16m}{L^2} \frac{T^3}{16} + nT \exp \left( -\frac{L^3T}{16} \right) + 4n^3T^{3/2} \exp \left( -\frac{L^{1/2}}{6n^2} \cdot T^{1/2} \right) \right)$$

$$= O \left( \frac{m + n}{L^2} + n^3T^{3/2} \exp \left( -\frac{T^{1/2}}{n^2} \right) \right).$$

Using the result for the bound of $(T - \mathbb{E}[\tau_S]) \cdot \max_{j \in [n]} |\mu_j - c_j^T \lambda^*|$ and $\lambda^T \mathbb{E} [B_{\tau_S}]$, we have

$$\text{Reg}_T \leq \sum_{j \in J^*} (\mu_j - c_j^T \lambda^*) \cdot \mathbb{E} [n_j(\tau_S) - n_j^0(\tau_S)] + \sum_{j \in J^*} (c_j^T \lambda^* - \mu_j) \cdot \mathbb{E} [n_j^0(\tau_S)]$$

$$+ \left( \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| + 2||\lambda^*||_1 \right) \left( 4 + \frac{16m}{L^2} \frac{T^3}{16} + nT \exp \left( -\frac{L^3T}{16} \right) + 4n^3T^{3/2} \exp \left( -\frac{L^{1/2}}{6n^2} \cdot T^{1/2} \right) \right)$$

Lastly, we finish the proof by applying Proposition 2 to the first two terms. The binding assumption do not prevent us from getting a problem dependent parameter $L$ for the bound in Proposition 2. Therefore
we have

{\text{Reg}_T} \leq \frac{2n \max_j |\mu_j - e_j^T \lambda^*|}{1 - \exp(-2L^2)}
+ \left( \max_{j \in [n]} |\mu_j - e_j^T \lambda^*| + 2||\lambda^*||_1 \right) \left( 4 + \frac{16m}{L^2} + \frac{nT}{L^2} \exp \left( -\frac{L^3T}{16} \right) + 4n^3T^{3/2} \exp \left( -\frac{L^{1/2}}{6n^2} \cdot T^{1/2} \right) \right)

by noticing that \(1 - \exp(-2L^2) \geq \frac{1}{2}L^2\) for \(L \leq 1\), we have

{\text{Reg}_T} \leq \frac{(48m + 4n + 12) \cdot ||\lambda^*||_1}{L^2} + 3||\lambda^*||_1 \cdot \left( \frac{nT}{L^2} \exp \left( -\frac{L^3T}{16} \right) + 4n^3T^{3/2} \exp \left( -\frac{L^{1/2}}{6n^2} \cdot T^{1/2} \right) \right)

= \frac{(48m + 4n + 12) \cdot ||\lambda^*||_1}{L^2} + o(1).

B Proof of Section 4

B.1 Proof of Lemma 6

For notational simplicity, assume that there is only one constraint, i.e. \(m = 1\) and both \(b_t\) and \(a_t\) are one-dimensional. To obtain the multi-dimensional result, we can simply take a union bound.

Recall the dynamic of the constraint process

\[ b_{t+1} = b_t + b_t - a_t x_t \]

Define

\[ Y_t = b_{t+1} - b_t, \quad X_t = Y_t - E[Y_t|F_{t-1}], \]

then we know \(S_t = \sum_{s=1}^t X_s\) is a martingale, and the difference is bounded by

\[ |S_t - S_{t-1}| = \left| \frac{b_t - a_t x_t - E[b_t - a_t x_t|F_{t-1}]}{T - t} \right| \leq \frac{1}{T - t}. \]

Next, denote the sample mean estimator as \(\hat{p}_s = \left( \frac{n_1(s)}{s}, \ldots, \frac{n_n(s)}{s} \right)\) and define the event

\[ \mathcal{G}_t = \cap_{s=1}^{t-1} \left\{ ||\hat{p}_s - p||_1 \leq \frac{\sqrt{4n \log 2T}}{\sqrt{s}} \right\}. \]

Under the event \(\mathcal{G}_t\) we have

\[
\sum_{s=1}^{t-1} E[Y_s|F_{s-1}] = \sum_{s=1}^{t-1} b_s - E[a_s x_s|\mathcal{H}_{s-1}] \leq \sum_{s=1}^{t-1} \sum_{j=1}^n c_j y_j,s \hat{p}_{j,s} - \sum_{j=1}^n c_j y_j,s p_j \leq \sum_{s=1}^{t-1} \frac{-1}{(T - s)\sqrt{s}} \sqrt{4n \log 2T},
\]

where the first inequality is because there are resources that might be non-binding, and the second
inequality is because \( \|c_j\|_\infty \leq 1 \) for any \( j \) and \( \|y_t\|_\infty \leq 1 \). Then,

\[
\mathbb{P}\left( b_t - b < -\frac{\sqrt{4n \log 2T}}{\sqrt{t}} - \frac{\sqrt{4n \log 2T + 2 \log 2T}}{\sqrt{T-t}} \right) \\
= \mathbb{P}\left( \sum_{s=1}^{t-1} X_s + \sum_{s=1}^{t-1} \mathbb{E}[Y_s|F_{s-1}] < -\frac{\sqrt{4n \log 2T}}{\sqrt{t}} - \frac{\sqrt{4n \log 2T + 2 \log 2T}}{\sqrt{T-t}} \right) \\
\leq \mathbb{P}\left( \left\{ \sum_{s=1}^{t-1} X_s + \sum_{s=1}^{t-1} \mathbb{E}[Y_s|F_{s-1}] < -\frac{\sqrt{4n \log 2T}}{\sqrt{t}} - \frac{\sqrt{4n \log 2T + 2 \log 2T}}{\sqrt{T-t}} \right\} \cap G_t \right) + \mathbb{P}(\bar{G}_t) \\
\leq \frac{1}{2T} + \sum_{s=1}^{t-1} \frac{1}{4T^2} \leq \frac{1}{T}.
\]

Here the first inequality comes from the introduction of the event \( G_t \)'s, the second inequality comes from (22) and the inequality

\[
\sum_{s=1}^{t} \frac{1}{(T-s)\sqrt{s}} \leq \frac{1}{\sqrt{t}} + \frac{1}{\sqrt{T-t}}
\]

the third inequality comes from applying Lemma 10 to get

\[
\mathbb{P}\left( \sum_{s=1}^{t-1} X_s < -\frac{\sqrt{4n \log 2T}}{\sqrt{T-t}} \right) \leq \exp\left( -\frac{2 \log 2T}{T-t} \right) \leq \frac{1}{2T},
\]

and the bound on \( \mathbb{P}(\bar{G}_t) \) comes from applying Lemma 8 and observe

\[
\mathbb{P}\left( \|\hat{p}_s - p\|_1 > \frac{\sqrt{4n \log 2T}}{\sqrt{s}} \right) \leq 2^n \exp\left( -2n \log 2T \right) \leq \exp\left( n \log 2 - 2n \log 2T \right)
\]

\[
\leq \exp\left( -n \log 2T \right) \leq \exp\left( -2 \log 2T \right) = \frac{1}{4T^2}.
\]

Notice that in the last line we assume the number of order types \( n \geq 2 \) because otherwise the problem will be trivial.

Having shown the property for \( G_t^{(1)} \), we then go ahead to show the result for \( D_t^{(j)} \). For \( j \leq n \), from Hoeffding’s inequality (Lemma 9) we know that for \( t > 1 \)

\[
\mathbb{P}\left( \frac{n_j(t-1)}{(t-1)p_j} - 1 > \frac{\sqrt{\log 2T}}{2p^2(t-1)} \right) = \mathbb{P}\left( \frac{n_j(t-1)}{(t-1)p_j} - p_j > \frac{\sqrt{\log 2T p_j}}{2p^2(t-1)} \right)
\]

\[
\leq 2 \exp\left( -\frac{1}{2} \frac{p_j^2 \log 2T}{t} \right) = \frac{1}{T},
\]

thereby finishing the proof.

### B.2 Proof of Lemma 7

For simplicity, we define the perturbed vector

\[
\xi_t := \left( \frac{n_1(t-1)}{(t-1)p_1}, \cdots, \frac{n_n(t-1)}{(t-1)p_n} \right) ^\top.
\]
Recall the DLP

\[ \text{OPT} := \max \mu^\top y \]
\[ \text{s.t. } Cy \leq b \\
0 \leq y \leq 1, \quad (25) \]

where \(0\) and \(1\) are \(n\)-dimensional vectors with entries being 0 and 1, respectively. Notice that at every time \(t\), the sample LP (5)

\[ R_t := \max \mu_t^\top y \]
\[ \text{s.t. } C_t y \leq b_t \\
0 \leq y \leq 1, \quad (26) \]

is equivalent to

\[ R_t := \max \mu^\top y' \]
\[ \text{s.t. } Cy' \leq b_t \\
0 \leq y' \leq \xi_t, \quad (27) \]

where \(y'_j = y_j \frac{n_j(t)}{(t-1)\|p_j\|} \). The reason for the formulation (27) is that we can transform the randomness in \(\mu, C\) to the randomness in the inventory process \(b_t\) and the constraint for \(y_j\), and this will facilitate the analysis of the reward accumulated at each time period \(t\). More specifically, we can view the objective value as a function of \(b_t\) and \(\xi_t\), and we can bound the difference of the objective function to the optimal value by bounding \(b_t - b\) and \(\xi_t - 1\). To analyze the dynamics of \(b_t\) and \(\xi_t\), we define events that will give us a “right” deviation to ensure a \(\tilde{O}(\sqrt{T})\) regret.

Define \(\text{OPT}(b_t, \xi_t)\) as the objective value of (27) with the right hand side constraint being \((b_t, \xi_t)\). Clearly, we have

\[ \text{OPT}(b_t, \xi_t) = R_t \text{ and } \text{OPT}(b, 1) = \text{OPT}_D. \]

The plan is to show the difference \(\text{OPT}(b, 1) - \text{OPT}(b + \Delta b, 1)\) and \(\text{OPT}(b + \Delta b, 1) - \text{OPT}(b, 1 + \Delta \xi)\) are bounded, and then derive a bound for \(\text{OPT}(b, 1) - \text{OPT}(b + \Delta b, 1 + \Delta \xi)\).

We start to show

\[ \text{OPT}(b, 1) - \text{OPT}(b + \Delta b, 1) \leq \tilde{\lambda} ||(-\Delta b) + ||_1. \]

We consider the dual program

\[ \text{OPT}_{\text{Dual}} := \min b^\top \lambda + \sum_{j=1}^n \gamma_j \]
\[ \text{s.t. } p_j c_j^\top \lambda + \gamma_j \geq p_j \mu_j, \quad j = 1, \ldots, n \\
\lambda_i \geq 0, i = 1, \cdots, m \\
\gamma_j \geq 0, j = 1, \cdots, n \quad (28) \]

which is equivalent to

\[ \text{OPT}_{\text{Dual}} := \min b^\top \lambda + \sum_{j=1}^n \mu_j - c_j^\top \lambda_+ \]
\[ \text{s.t. } \lambda_i \geq 0, i = 1, \cdots, m. \]

The above program could be understood as a LP with finite dimension, where there exists finite many simplex solutions. Recall that

\[ \tilde{\lambda} := \max \{||\lambda||_\infty : \lambda \in \mathcal{F}D_0\} \]
and $\mathcal{F}D_0$ denotes the set of basic solutions for the dual of DLP (3). We know for any $\Delta b \in \mathbb{R}^m$,

$$\text{OPT}(b, 1) - \text{OPT}(b + \Delta b, 1) \leq \bar{\lambda}||(-\Delta b)_+||_1.$$  

This is because there will always exists a basic solution $\lambda$ that is the optimal solution, and the rate of change of the optimal value with respect to $b$ will always be bounded by $\bar{\lambda}$.

Next, we begin to show

$$\text{OPT}(b + \Delta b, 1) - \text{OPT}(b + \Delta b, 1 + \Delta \xi) \leq ||\Delta \xi||_1.$$  

This is obvious by taking a look at (27). If we increase/decrease the constraint of $y_j^*$, the optimal value can at most increase proportionally to $\mu_j$. Combining the pieces together we have

$$\text{OPT}(b, 1) - \text{OPT}(b + \Delta b, 1 + \Delta \xi)$$

$$= \text{OPT}(b, 1) - \text{OPT}(b + \Delta b, 1) + \text{OPT}(b + \Delta b, 1) - \text{OPT}(b + \Delta b, 1 + \Delta \xi)$$

$$\leq \bar{\lambda}||(-\Delta b)_+||_1 + ||\Delta \xi||_1.$$  

Then, under the condition that $b_t = b + \Delta b$, $\xi_t = 1 + \Delta \xi$ and $C_t \cap D_t$, from equations above we have

$$\text{OPT}_D - R_t = \text{OPT}(b, 1) - \text{OPT}(b_t, \xi_t) \leq \bar{\lambda}||(-\Delta b)_+||_1 + ||\Delta \xi||_1$$

$$m \bar{\lambda} \left( \frac{\sqrt{4n \log 2T}}{\sqrt{t}} + \frac{\sqrt{4n \log 2T} + \sqrt{2 \log 2T}}{\sqrt{T - t}} \right) + n \frac{\sqrt{\log 2T}}{\sqrt{2p^2(t - 1)}}.$$  

### B.3 Proof of Theorem 3

With Lemma 7, we can have an upper bound for the single-period regret. By taking the summation, we can obtain an upper bound for the cumulative regret. More specifically,

$$\text{Reg}^*_t = \sum_{t=1}^{T} (\mu^\top y^* - \mathbb{E}[a_t x_t])$$

$$= \sum_{t=1}^{T} \mathbb{E}[\text{OPT}_D - R_t]$$

$$\leq \sum_{t=1}^{T} \left( \mathbb{E}[\text{OPT}_D - R_t | C_t \cap D_t] \mathbb{P}(C_t \cap D_t) + 1 \cdot \mathbb{P}(C_t \cap D_t) + n + m \frac{1}{T} \right)$$

$$\leq 1 + \sum_{t=2}^{T} \left( \max\{1, \bar{\lambda}\} \left( m \left( \frac{\sqrt{4n \log 2T}}{\sqrt{t}} + \frac{\sqrt{4n \log 2T} + \sqrt{2 \log 2T}}{\sqrt{T - t}} \right) + n \frac{\sqrt{\log 2T}}{\sqrt{2p^2(t - 1)}} \right) + n + m \frac{1}{T} \right)$$

$$\leq \left( m \left( \sqrt{2} + \sqrt{16n} \right) + \frac{n}{\sqrt{2p^2}} \right) \max\{1, \bar{\lambda}\} \sqrt{T} \log 2T + 1 + n + m,$$  

where in the first inequality we use the fact that $||\mu||_\infty \leq 1$, therefore $\text{OPT}_D \leq 1$, and in the third inequality we apply Lemma 7 to get the regret bound.
C Regret Analysis for Nondegenerate Case with Both Binding and Nonbinding Constraint

In Section 3, we analyze the nondegenerate case under the assumption that all the constraints are binding (Assumption 3). In this section we remove the assumption and discuss the more general case where for the underlying DLP (3), both binding and non-binding constraints exist. We present the final result in Theorem 4. We first note that Corollary 1 and Proposition 2 hold without dependency on the bindingness of the underlying LP. So the remaining task is to reproduce the constraint process analysis under the general case to deal with the last two terms in Corollary 1.

C.1 Constraint process under general case

Recall the stopping time that the LP’s structure changes

\[ \tau_S := \min \{ t \leq T : |b_{i,t} - b_i| > L \text{ for some } i \in \mathcal{I}^* \} \cup \{ t \leq T : b_{i,t} - b_i < -L \text{ for some } i \in \mathcal{I}' \} \cup \{ T + 1 \}. \]

As the arguments in Section 3.2, the parameter \( L \) and stopping time \( \tau_S \) are critical in bounding the order acceptance in Proposition 2.

With a slight overload of the notation, we adjust the previous definition of the region \( \mathfrak{B} \) as

\[ \mathfrak{B} := \left( \bigotimes_{i \in \mathcal{I}^*} [b_i - L, b_i + L] \right) \bigotimes \left( \bigotimes_{i \in \mathcal{I}'} [b_i - L, \infty) \right). \]

In this way, the stopping time \( \tau_S \) can be expressed by

\[ \tau_S = \min \{ t \leq T : b_t \notin \mathfrak{B} \} \cup \{ T + 1 \}. \]

Unlike the previous section, the definition of \( \mathfrak{B} \) here differentiates between binding and non-binding constraints. For binding constraints, the definition is the same as before, and it aims to capture the first time that the remaining average resource capacity \( b_{i,t} \) deviates from \( b_i \) by \( L \). For non-binding constraints, we only concern a downward deviation – a lower bound for the deviation because that is sufficient to guarantee that the non-binding constraints remain non-binding.

Accordingly, we need to slightly adjust the previous definition of event \( \mathcal{E}_t \) as follows. Denote \( \bar{b} := \min\{b_1, ..., b_m\} \) and let

\[ \epsilon_t^* := \begin{cases} 1 & t \leq \kappa T, \\ \frac{1}{\tau T} & t > \kappa T, \end{cases} \quad \epsilon'_t := \begin{cases} 1 + \bar{b} - \frac{L}{\tau T} & t \leq \kappa T, \\ \frac{1}{\tau T} & t > \kappa T, \end{cases} \]

with \( \kappa \) to be specified and

\[ \mathcal{E}_t := \left\{ \mathcal{H}_{t-1} \left| \sup_{b' \in \mathfrak{B}} \| E[a_{\mathcal{I}^*, t} x_t(b')] \mathcal{H}_{t-1} - b'_{\mathcal{I}^*} \|_\infty \leq \epsilon^*_{t-1} \right. \quad \sup_{b' \in \mathfrak{B}} E[a_{\mathcal{I}', t} x_t(b')] \mathcal{H}_{t-1} - b'_{\mathcal{I}'} \leq \epsilon'_{t-1} \right\}, \]

where the subscripts \( \mathcal{I}^* \) and \( \mathcal{I}' \) denote the corresponding dimensions of the vectors. Here we define different tolerance levels \( \epsilon^* \) and \( \epsilon' \) for binding and non-binding dimensions, respectively. The intuition is that for non-binding dimensions, we can tolerate larger deviation for the resource consumption as long as it does not sabotage the non-bindingness.
According to the new event $\mathcal{E}_t$, we can define the stopping time

$$
\tilde{\tau} := \min\{t \leq T : b_t \notin B \text{ or } H_{t-1} \notin \mathcal{E}_t\} \cup \{T+1\},
$$

and the auxiliary process

$$
\tilde{b}_t = \begin{cases} 
    b_t, & t < \tilde{\tau}, \\
    b_T, & t \geq \tilde{\tau}.
\end{cases}
$$

These two definitions lead to the same decomposition as before.

$$
P(\tau_S \leq t) = P(b_s \notin B \text{ for some } s \leq t) \leq P\left(b_s \notin B \text{ for some } s \leq t\right) + \sum_{s=1}^{t} P((r_1, a_1, ..., r_{s-1}, a_{s-1}) \notin \mathcal{E}_s).
$$

The following lemma generalizes Lemma 4 to the general case (without Assumption 3). The proof follows the exact same arguments as Lemma 4. As to the statement, for definition of constants there is an additional term compared to Lemma 4 and it arises from the non-binding dimensions. The proof is left in later subsections.

**Lemma 12.** The following inequality holds for all $T \geq T_1$ and $t \leq T - 2$,

$$
P\left(b_s \notin B \text{ for some } s \leq t\right) \leq 2me^{\frac{L^2(\tau_{t-1})}{s}},
$$

where the constant $T_1$ is defined as the minimal integer such that

$$
T_1 \geq \left(\frac{1}{\exp\left(\frac{L}{s}\right) - 2}\right) \vee \left(\frac{1}{\exp\left(\frac{L}{8(1+L-B)}\right) - 2}\right)
$$

and \(\log T_1 \leq \frac{\kappa^{1/4}L}{4}\), where $\kappa$ is set by $\kappa = \left(1 - \exp\left(-\frac{L}{s}\right)\right) \land \left(1 - \exp\left(-\frac{L}{8(1+L-B)}\right)\right)$.

Next, for the second term in (30), we keep the same definition of events $\mathcal{A}^{(j)}_t$ and $\mathcal{B}^{(j)}_t$ in the previous section such that

$$
\mathcal{A}^{(j)}_t := \left\{\left\lfloor \frac{n_j(t-1)}{t-1} - p_j\right\rfloor \leq L\right\},
$$

$$
\mathcal{B}^{(j)}_t := \left\{\left\lfloor \frac{n_j(t-1)}{t-1} - p_j\right\rfloor \leq \frac{1}{n(t-1)^{1/4}}\right\}.
$$

Their definitions convey the same intuition as before. The event $\cap_{j=1}^n \mathcal{A}^{(j)}_t$ ensures that the sampled LP's bindingness structure aligns with that of the DLP, while the event $\cap_{j=1}^n \mathcal{B}^{(j)}_t$ ensures that given $\cap_{j=1}^n \mathcal{A}^{(j)}_t$, the expected resource consumption at time $t$ of the algorithm stays close to $b_t$. The following lemma is analogous to Lemma 5 and provides an lower bound for the “good” event $\mathcal{E}_t$. In addition, we can obtain a probability bound on the stopping time $\tau_S$.

**Lemma 13.** We have \(\cap_{j=1}^n \mathcal{A}^{(j)}_t\) \cap \(\cap_{j=1}^n \mathcal{B}^{(j)}_t\) \(\subset \mathcal{E}_t\) for $t = 1, ..., T$. Under Assumption 1 and 2, we have for $t \leq T - 2$,

$$
P(\mathcal{E}_t) = \begin{cases} 
    1 & \text{for } t \leq \kappa T \\
    1 - 2n \exp\left(-2L^2(t-1)\right) - 2n \exp\left(-\frac{2(t-1)^{1/2}}{n^2}\right) & \text{for } t > \kappa T
\end{cases}
$$
Consequently,
\[ \mathbb{E}[\tau_S] \geq T - 2 - \frac{16m}{L^2} - \frac{nT}{L^2} \exp \left( -2L^2\kappa T \right) - 4n^3T^{3/2} \exp \left( -\frac{\kappa^{1/2}}{n^2} \cdot T^{1/2} \right). \]

By putting the analysis of the stopping time \( \tau_S \) together with the analysis in Corollary 1 and Proposition 2, we can obtain the final regret bound as in the theorem below. As the bound for the all-binding case in the previous section, the regret bound bears no dependency in terms of the time horizon \( T \). Our analysis mainly focuses on removing the dependency on time \( T \), and the result thus indicates that the adaptive design of the algorithm can significantly mitigate the effect of the parameter learning/estimation error on the regret.

**Theorem 4.** Under Assumption 1 and 2, Algorithm 1 give a regret in the order of
\[ \text{Reg}_T^\pi \leq \left( \frac{48m + 4n + 12}{L^2} \right) \cdot \left\| \lambda^* \right\|_1 + o(1). \]

In the following subsections, we elaborate the proofs for the results in above.

### C.2 Proof of Lemma 12

We treat the case for binding index and non-binding index separately. For binding index \( i \in \mathcal{I}^* \) and \( t = 1,...,T \), let
\[ Y_t := \tilde{b}_{i,t+1} - \tilde{b}_{i,t}, \quad X_t := Y_t - \mathbb{E}[Y_t | \mathcal{H}_{t-1}]. \]

For binding resources, the setup is completely the same as the proof in Lemma 4. Therefore from the same definition that \( \kappa \leq 1 - \exp(-\frac{L^2}{8}) \), and that \( T_1 \) is the minimal integer such that \( T_1 \geq \frac{1}{\exp\left(\frac{L^2}{8}\right)} + 2 \) and \( \frac{\log T_1}{T_1} \leq \frac{\kappa^{1/4}}{4} \), we have
\[ \mathbb{P} \left( \tilde{b}_{i,s} \notin \mathfrak{B} \text{ for some } s \leq t \right) \leq 2e^{-\frac{L^2(\tau-1)}{8}} \]
for \( i \in \mathcal{I}^*, t \leq T - 2, \) and \( T \geq T_1 \).

Next, for non-binding index \( i \in \mathcal{I}' \), we use the same definition on \( Y_t, X_t \) such that
\[ Y_t := \tilde{b}_{i,t+1} - \tilde{b}_{i,t}, \quad X_t := Y_t - \mathbb{E}[Y_t | \mathcal{H}_{t-1}]; \]

and want to show
\[ \sum_{j=1}^{s} X_j - \sum_{j=1}^{s} Y_j \leq \frac{L}{2}. \]

From the fact that
\[ b_{i,t} - a_{i,t}x_1(b_t) \geq \inf_{b' \in \mathfrak{B}} b_{i,t} - \mathbb{E}[a_{i,t}x_1(b') | \mathcal{H}_{t-1}] \geq \begin{cases} \frac{b - L - 1}{1/T} & t \leq \kappa T, \\ \frac{1}{1/T} & t > \kappa T, \end{cases} \]
We know
\[X_t - Y_t = -\mathbb{E}[Y_t | \mathcal{H}_{t-1}]\]
\[= -\mathbb{E}\left[\frac{1}{T-t}(b_t - a_t x_t(b_t))I(\bar{t} > t) \middle| \mathcal{H}_{t-1}\right]\]
\[\leq \frac{c_{l+1}}{T-t}\]
\[= \frac{1 + L - \frac{b}{T}}{T - t}I(t \leq \kappa T) + \frac{1}{t^{1/4}(T-t)}I(t > \kappa T)\]

From the same approach in the proof of Lemma 4, by defining \(\kappa \leq 1 - \exp(-\frac{L}{\kappa (1 + L - \frac{L}{2})})\), and \(T_1 \geq \frac{1}{\exp(\frac{1}{\kappa (1 + L - \frac{L}{2})})} + 2\) and \(\log\frac{T_1}{T_1} \leq \frac{\kappa L^4}{4}\), we know that for \(T > T_1\),
\[\sum_{j=1}^{\kappa T} \frac{1 + L - \frac{b}{T}}{T - j} \leq (1 + L - \frac{b}{T})(\log \left(\frac{T - 1}{T - 2}\right) - \log (1 - \kappa)) \leq \frac{L}{4},\]
\[\sum_{j=\kappa T + 1}^{T-1} \frac{1}{(T-j)^{3/4}} \leq 1 \leq (\kappa T)^{3/4} \sum_{j=\kappa T + 1}^{T-1} \frac{1}{T - j} \leq \log T \leq (\kappa T)^{3/4} \leq \frac{L}{4}.
\]

Therefore, with the choice of \(\kappa\) and \(T \geq T_1\), we have
\[\sum_{j=1}^{s} X_j - \sum_{j=1}^{s} Y_j = \sum_{j=1}^{\kappa T} \frac{1 + L - \frac{b}{T}}{T - j} + \sum_{j=\kappa T + 1}^{s} \frac{1}{(T-j)^{3/4}} \leq \frac{L}{2}.
\]

Next, similar to Lemma 4, for \(t \leq T - 2\), we have that
\[\mathbb{P}\left(\sum_{j=1}^{s} X_j \leq -\frac{L}{2} \text{ for some } s \leq t\right) \leq e^{-\frac{L^2(t-1)}{8}}.
\]

Then, we have for \(i \in \mathcal{T}'\)
\[\{\bar{b}_{s-i} - b_i \leq -L \text{ for some } s \leq t\} = \left\{\sum_{j=1}^{s-1} Y_j \leq -L \text{ for some } s \leq t\right\}
\[\subseteq \left\{\sum_{j=1}^{s} X_j \leq -\frac{L}{2} \text{ for some } s \leq t - 1\right\}.
\]

Therefore, for \(i \in \mathcal{T}'\) we have that
\[\mathbb{P}\left(\bar{b}_{s-i} \notin \mathcal{B} \text{ for some } s \leq t\right) \leq 2e^{-\frac{L^2(t-1)}{8}}.
\]

Summing up and taking the union bound, we know that for all \(T \geq T_1\) and \(t \leq T - 2\), we have
\[\mathbb{P}\left(\bar{b}_s \notin \mathcal{B} \text{ for some } s \leq t\right) \leq 2me^{-\frac{L^2(t-1)}{8}}.
\]

C.3 Proof of Lemma 13

To show that \((\cap_{j=1}^{n} A_i^{(j)}) \cap (\cap_{j=1}^{n} B_i^{(j)}) \subset \mathcal{E}_t\), we firstly denote that it suffices to show such property for \(t \geq \kappa T\), because the requirement for \(t \leq \kappa T\) holds with probability 1 for both the binding resource in \(\mathcal{T}^*\)
and the non-binding resource in $\mathcal{I}$. To see this, notice that for any $b' \in \mathfrak{B}$ and $i \in \mathcal{I}^*$, we have

$$-b'_i \leq \mathbb{E}[a_{i,t}] - b'_i \leq 1 - b'_i.$$  

Because $b' \in \mathfrak{B}$ and $i \in \mathcal{I}^*$, without loss of generality we can take $L \leq 1 - b'_i$ then assume that $b'_i < 1$ (otherwise if there is no slackness for $b_i$, the problem will be degenerate). Therefore, we know that

$$\|\mathbb{E}[a_{\mathcal{I}^*,t}x_t(b')|\mathcal{H}_{t-1}] - b'_\mathcal{I}\|_\infty \leq 1.$$  

For $b' \in \mathfrak{B}$ and $i \in \mathcal{I}$, from $b'_i > b - L$ we know that

$$\mathbb{E}[a_{\mathcal{I}^*,t}x_t(b')|\mathcal{H}_{t-1}] - b'_\mathcal{I} \leq 1 + L - b.$$  

From above we can see that $P(\mathcal{E}_t) = 1$ for $t \leq \kappa T$.

For $t > \kappa T$, under $\left(\cap_{j=1}^n A_t^{(j)}\right)$ and $b_t = b' \in \mathfrak{B}$, we know that the sampled LP (5) is stable such that the optimal basis and the resource bindiness is the same as the DLP (3). Then, Lemma 1 tell us that there exists $b$ such that $b_{\mathcal{I}^*} = b'_{\mathcal{I}^*}$, and the same binding structure also implies that we can define $b_{\mathcal{I}^*}$ in the non-binding dimension to be

$$b = \sum_{j=1}^n c_j \frac{n_k(t-1)}{t-1} y^*_j(b'),$$

where $y^*$ is the solution for (5). More specifically, from above we know that

$$b_{\mathcal{I}^*} = b'_{\mathcal{I}^*} = \sum_{j=1}^n c_{\mathcal{I}^*,j} \frac{n_j(t-1)}{t-1} y^*_j(b'),$$

$$b_{\mathcal{I}^*} = \sum_{j=1}^n c_{\mathcal{I}^*,j} \frac{n_j(t-1)}{t-1} y^*_j(b') \leq b'_{\mathcal{I}^*}.$$  

Therefore, from equation (31) and Algorithm 1 we know that

$$b'_{\mathcal{I}^*} - \mathbb{E}[a_{\mathcal{I}^*,t}x_t(b')|\mathcal{H}_{t-1}] = b'_{\mathcal{I}^*} - \sum_{j=1}^n c_j p_j y^*_j(b')$$

$$= b'_{\mathcal{I}^*} - b_{\mathcal{I}^*} + \sum_{j=1}^n c_j \left( \frac{n_j(t-1)}{t-1} - p_j \right) y^*_j(b')$$

$$\geq (b_{\mathcal{I}^*} - L) + (L - b_{\mathcal{I}^*}) - \frac{1}{(t-1)^{1/4}} = -\frac{1}{(t-1)^{1/4}},$$

where in the last line, $b_{\mathcal{I}^*} \geq b_{\mathcal{I}^*} - L$ is from the fact that $b' \in \mathfrak{B}$, $-b_{\mathcal{I}^*} \geq L - b_{\mathcal{I}^*}$ is from Lemma 1 (more specifically, the positivity of slack variable and the stability of the optimal basic index), and the last one is from the inequality ensured by $\left(\cap_{j=1}^n A_t^{(j)}\right)$ that $\sum_{j=1}^n \left| \frac{n_j(t-1)}{t-1} - p_j \right| \leq \frac{1}{(t-1)^{1/4}}$. Therefore, under $\left(\cap_{j=1}^n A_t^{(j)}\right) \cap \left(\cap_{j=1}^n B_t^{(j)}\right)$ we have shown that

$$\sup_{b' \in \mathfrak{B}} \mathbb{E}[a_{\mathcal{I}^*,t}x_t(b')|\mathcal{H}_{t-1}] - b_{\mathcal{I}^*} \leq \frac{1}{(t-1)^{1/4}}.$$
Then, in order to show \(\cap_{j=1}^{n} A_t^{(j)} \cap \cap_{j=1}^{n} B_t^{(j)} \subset \mathcal{E}_t\), it suffices to show that
\[
\sup_{b' \in \mathbb{W}} \|\mathbb{E} [a_{T^*, t \varepsilon_t} (b')] |_{\mathcal{H}_{t-1}} - b_T'\|_{\infty} \leq \frac{1}{(t-1)^{1/4}}.
\]

Under event \(\cap_{j=1}^{n} A_t^{(j)} \cap \cap_{j=1}^{n} B_t^{(j)}\), from equation (31) and Algorithm 1 we know that
\[
\|\mathbb{E} [a_{T^*, t \varepsilon_t} (b')] |_{\mathcal{H}_{t-1}} - b_T'\|_{\infty} \leq \frac{1}{(t-1)^{1/4}},
\]
where the last equation follows from the definition of \(\cap_{j=1}^{n} B_t^{(j)}\). Therefore, we are done with showing \(\cap_{j=1}^{n} A_t^{(j)} \cap \cap_{j=1}^{n} B_t^{(j)} \subset \mathcal{E}_t\).

Next, for \(t \leq T \leq 2\), we have to show
\[
\mathbb{P}(\mathcal{E}_t) \begin{cases} = 1 & \text{for } t \leq \kappa T \\
 \geq 1 - 2n \exp \left(2\sum_{j=1}^{n} (p_j - \frac{n_j(t-1)}{t-1}) \right) \sum_{j=1}^{n} \left(1 - \exp \left(-\frac{2(t-1)^{1/2}}{n^2} \right) \right) & \text{for } t > \kappa T 
\end{cases}
\]
This is trivial since by definition of \(\mathcal{E}_t\), when \(t \leq \kappa T\), \(\mathcal{E}_t\) is an event with probability 1. For \(t > \kappa T\), the proof follows from applying Lemma 8.

Finally, the bound
\[
\mathbb{E}[\tau_\mathcal{S}] \geq T - 2 - \frac{16m}{L^2} \frac{nT}{L^2} \exp \left(-2L^2 \kappa T\right) - 4n^3T^{3/2} \exp \left(-\frac{\kappa^{1/2}}{n^2} \cdot 4T^{1/2} \right)
\]
follows from the same approach in (21), Lemma 11 and 12, because all the terms in (21) have the same bound as their counterparts in Section 3.3.

### C.4 Proof of Theorem 4

Recall the Corollary 1 also holds without Assumption 3, and we have
\[
\text{Reg}_t^{\mathcal{S}} \leq \sum_{j \in J^*} \left(\mu_j - c_j^T \lambda^*\right) \cdot \mathbb{E} \left[n_j(t') - n_j^2(t')\right] + \sum_{j \in J^*} \left(c_j^T \lambda^* - \mu_j\right) \cdot \mathbb{E} \left[n_j^2(t')\right]
\]
\[
+ \left(T - \mathbb{E}[\tau]\right) \cdot \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| + \lambda^T \cdot \mathbb{E} [B_{\tau}].
\]

By substituting \(\tau_\mathcal{S}\) for \(\tau\), we can again bound the first two terms using Proposition 2, and thereby having
\[
\text{Reg}_t^{\mathcal{S}} \leq \frac{2n \max_j |\mu_j - c_j^T \lambda^*|}{1 - \exp(-2L^2)} + \left(T - \mathbb{E}[\tau_\mathcal{S}]\right) \cdot \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| + \lambda^T \cdot \mathbb{E} [B_{\tau_\mathcal{S}}].
\]

As for the last two terms, from Lemma 13 we know that
\[
\mathbb{E}[\tau_\mathcal{S}] \geq T - 2 - \frac{16m}{L^2} \frac{nT}{L^2} \exp \left(-2L^2 \kappa T\right) - 4n^3T^{3/2} \exp \left(-\frac{\kappa^{1/2}}{n^2} \cdot 4T^{1/2} \right),
\]
and without loss of generality let us assume \(1 + L - \frac{b}{2} \leq 1\) and \(\frac{L}{8} \leq 1\), thereby having \(\kappa = 1 - \exp \left(-\frac{L}{32}\right) \geq \frac{L}{32}\), and
\[
\mathbb{E}[\tau_\mathcal{S}] \geq T - 2 - \frac{16m}{L^2} \frac{nT}{L^2} \exp \left(-\frac{L^3T}{16}\right) - 4n^3T^{3/2} \exp \left(-\frac{\kappa^{1/2}}{6n^2} \cdot 4T^{1/2} \right).
\]
Then, we have

\[
(T - E[rs]) \cdot \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| \\
\leq \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| \left(4 + \frac{16m}{L^2} + \frac{nT}{L^2} \exp \left(-\frac{L^3T}{16}\right) + 4n^3T^{3/2} \exp \left(-\frac{L^{1/2}}{6n^2} \cdot T^{1/2}\right) \right),
\]

and from Lemma 3, we have

\[
\lambda^{*T} E[B_{rs}] \leq 2||\lambda^*||_1 \left(4 + \frac{16m}{L^2} + \frac{nT}{L^2} \exp \left(-\frac{L^3T}{16}\right) + 4n^3T^{3/2} \exp \left(-\frac{L^{1/2}}{6n^2} \cdot T^{1/2}\right) \right).
\]

Lastly noticing \(\exp(-2L^2) \geq \frac{1}{2}L^2\) for \(L \leq 1\) and combining the results above we have

\[
\text{Reg}^\pi \leq \frac{2n \max_j |\mu_j - c_j^T \lambda^*|}{1 - \exp(-2L^2)} + (T - E[rs]) \cdot \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| + \lambda^{*T} \cdot E[B_{rs}] \\
\leq \frac{4n \max_j |\mu_j - c_j^T \lambda^*|}{L^2} + \left( \max_{j \in [n]} |\mu_j - c_j^T \lambda^*| + 2||\lambda^*||_1 \right) \cdot \left(4 + \frac{16m}{L^2} + \frac{nT}{L^2} \exp \left(-\frac{L^3T}{16}\right) + 4n^3T^{3/2} \exp \left(-\frac{L^{1/2}}{6n^2} \cdot T^{1/2}\right) \right) \\
\leq \frac{(48m + 4n + 12) \cdot ||\lambda^*||_1}{L^2} + 3||\lambda^*||_1 \cdot \left(\frac{nT}{L^2} \exp \left(-\frac{L^3T}{16}\right) + 4n^3T^{3/2} \exp \left(-\frac{L^{1/2}}{6n^2} \cdot T^{1/2}\right) \right) \\
= \frac{(48m + 4n + 12) \cdot ||\lambda^*||_1}{L^2} + o(1).
\]

## D Multi-dimensional Case

Now, we return to the general multi-dimensional formulation in LP (1). We note that the analysis of the multi-dimensional problem is essentially the same as the one-dimensional case discussed previously and the presentation here is mainly for completeness. So this section will mainly emphasize the different parts between the multi-dimensional case and the one-dimensional case, and the rest will follow.

Firstly, for the multi-dimensional case, the distribution assumption becomes the following. The assumption and its interpretation here are parallel to Assumption 1.

### Assumption 4. We assume

(a) The column-coefficient pair \((r_t, A_t)\)'s are i.i.d. sampled from a distribution \(\mathcal{P}\). The distribution \(\mathcal{P}\) takes a finite and known support \(\{(\mu_j, c_j)\}_{j=1}^n\) where \(\mu_j \in \mathbb{R}^k\) and \(c_j \in \mathbb{R}^{m \times k}\). Specifically, \(\mathcal{P}(r_t, A_t) = (\mu_j, c_j)\) is \(p_j\) for \(j = 1, ..., n\) and the parameters \(p = (p_1, ..., p_n)^\top\) are unknown.

(b) Positiveness and Boundedness: \(\mu_j, c_j \geq 0\) and \(||\mu_j||_\infty, ||c_j||_\infty \leq 1\) for \(j = 1, ..., n\).

(c) The right-hand-side \(B = Tb \geq 0\) where \(b = (b_1, ..., b_m)^\top\).

The multi-dimensional DLP will become the following

\[
\max_{\mu} \sum_{j=1}^n \mu_j^T y_j p_j \\
\text{s.t.} \sum_{j=1}^n c_j y_j p_j \leq b \\
\quad \quad 1^T y_j \leq 1, \quad y_j \geq 0, \quad j = 1, ..., n,
\]

41
Proposition 3. Under Assumption 4 and assume the DLP (32) is nondegenerate, the following equality holds

\[
\text{Reg}_T^\tau = \mathbf{X}^\top \cdot \mathbb{E} [\mathbf{B}_\tau] + \sum_{j \in \mathcal{J}^*} \mathbb{E} \left[ \max_{l \in \mathcal{K}_j} (\mu_l - c_l^\top \mathbf{X}^*) \cdot n_j(T) - \sum_{l \in \mathcal{K}_j} (\mu_l - c_l^\top \mathbf{X}^*) n_j^0(\tau) \right] \\
+ \sum_{j \in \mathcal{J}' \subset \mathcal{J}} \sum_{l \in \mathcal{K}_j} (c_l^\top \mathbf{X}^* - \mu_l) \cdot \mathbb{E} [n_j^0(\tau)].
\]

To bound the second and third term, an important observation is that if we are able to always choose the right \( l \in \mathcal{K}_j \) such that \( \mu_l - c_l^\top \mathbf{X}^* = \max_{l \in \mathcal{K}_j} (\mu_l - c_l^\top \mathbf{X}^*) \), the second term will be 0. Moreover, if we are able to reject all the orders in the order group \( j \in \mathcal{J}' \), the third term will be 0.

A similar corollary as Corollary 1 transforms the above result to the case that involves a general stopping time \( \tau' \).

Corollary 2. The following inequality holds

\[
\text{Reg}_{\tau'}^\tau \leq \mathbf{X}^\top \cdot \mathbb{E} [\mathbf{B}_{\tau'}] + \sum_{j \in \mathcal{J}^*} \mathbb{E} \left[ \max_{l \in \mathcal{K}_j} (\mu_l - c_l^\top \mathbf{X}^*) \cdot n_j(\tau') - \sum_{l \in \mathcal{K}_j} (\mu_l - c_l^\top \mathbf{X}^*) n_j^0(\tau') \right] \\
+ \sum_{j \in \mathcal{J}' \subset \mathcal{J}} \sum_{l \in \mathcal{K}_j} (c_l^\top \mathbf{X}^* - \mu_l) \cdot \mathbb{E} [n_j^0(\tau')] + (T - \mathbb{E}[\tau']) \cdot \max_{j \in \mathcal{J}^* \cup \mathcal{J}'} |\mu_l - c_l^\top \mathbf{X}^*|
\]

where \( \tau' \) is a stopping time adapted to the process \( \mathbf{B} \), and \( \tau' \leq \tau \) almost surely.
The following proposition mimics Proposition 2 to provide a bound for the second and third term for the above corollary. Its proof idea is similar to that of Proposition 2, both of which utilize the stability result in Lemma 1.

**Proposition 4.** Under Assumption 4 and assume the DLP (32) is nondegenerate, the output of Algorithm 1 satisfies

\[
\sum_{j \in J^*} E \left[ \max_{l \in K_j} (\mu_l - c_l^T \lambda^*) \cdot n_j(\tau') - \sum_{l \in K_j} (\mu_l - c_l^T \lambda^*) n_l^a(\tau') \right] + \sum_{j \in J^*} \sum_{l \in K_j} (c_l^T \lambda^* - \mu_l) \cdot E [n_l^a(\tau')] \\
\leq \frac{2n \max_j |\mu_j - c_j^T \lambda^*|}{1 - \exp(-2L^2)}
\]

In the following subsection, we lay out the details of the proof.

### D.1 Analysis of the constraint process for the multi-dimensional case

The dimension of the constraint process for the multi-dimensional case remains the same as before, and its dynamics can also be analyzed in a similar way. To complete the analysis, a few minor changes need to be made and are elaborated as follows. We can carry out the same analysis as before. The only caveat is that we need to adjust the definitions for vectors \(r_t\)'s and matrices \(A_t\)'s as below. For simplicity, we assume the case that all the resource is binding again and one will see that the non-binding case follows by the same analysis. Define

\[
\mathcal{B} := \bigotimes_{i=1}^m [b_i - L, b_i + L],
\]

\[
\mathcal{E}_t := \left\{ \mathcal{H}_{t-1} \bigg| \sup_{b' \in \mathcal{B}} \|\mathcal{E}_t(x_t(b'))|\mathcal{H}_{t-1} - b'\|_\infty \leq \epsilon_{t-1} \right\},
\]

\[
\epsilon_t := \begin{cases} 
1 & t \leq \kappa T, \\
1/t & t > \kappa T,
\end{cases}
\]

Again by defining

\[
\tau_S := \min\{t \leq T : b_t \notin \mathcal{B} \} \cup \{T + 1\},
\]

\[
\tilde{\tau} := \min\{t \leq T : b_t \notin \mathcal{B} \text{ or } \mathcal{H}_{t-1} \notin \mathcal{E}_t \} \cup \{T + 1\},
\]

and

\[
\tilde{b}_t = \begin{cases} 
b_t & t < \tilde{\tau}, \\
b_{\tilde{\tau}} & t \geq \tilde{\tau},
\end{cases}
\]

we can have the same decomposition

\[
P(\tau_S \leq t) = P(b_s \notin \mathcal{B} \text{ for some } s \leq t) \\
\leq P(b_s \notin \mathcal{B} \text{ for some } s \leq t) + \sum_{s=1}^{t} P((r_1, A_1, ..., r_{s-1}, A_{s-1}) \notin \mathcal{E}_s)
\]

(34)

For the first term in (34), we find that Lemma 4 follows easily since the analysis therein does not involve the extra dimension \(k\). For the second term, we can just change the decision variables \(y_j\)'s from scalars to vectors and everything will follow.
From the definition of the events

\[ \mathcal{A}_t^{(j)} = \left\{ \left\lfloor \frac{n_j(t-1)}{t-1} - p_j \right\rfloor \leq L \right\} \quad \text{and} \quad \mathcal{B}_t^{(j)} = \left\{ \left\lfloor \frac{n_j(t-1)}{t-1} - p_j \right\rfloor \leq \frac{1}{n(t-1)^{1/4}} \right\}, \]

we remark that the derivation of

\[ \left( \bigcap_{j=1}^n \mathcal{A}_t^{(j)} \right) \cap \left( \bigcap_{j=1}^n \mathcal{B}_t^{(j)} \right) \subset \mathcal{E}_t \]

is independent of extra dimension \( k \) because only the norms on \( p, A, B \) are required. Therefore, the result in Theorem 2 also holds for the multi-dimensional case under Assumption 4 and the assumption of nondegeneracy. We have laid out the same pathways as in the one-dimensional case, and from now on, the analysis in Section C in terms of the handling of the non-binding constraints will be the same.

### D.2 Proof of Proposition 3

Denote \( \mathcal{T}_t \) to be the arriving order group at time \( t \), also notice that here \( (r_t, a_t) \) is the reward/consumption pair accepted by our algorithm at time \( t \).

\[
\text{Reg}_T^\tau = \mathbb{E} \left[ T \cdot \text{OPT}_D - \sum_{t=1}^\tau r_t x_t \right]
\]

\[
= \mathbb{E} \left[ \lambda^\top B + \sum_{t=1}^\tau \max_{l \in \mathcal{T}_t} (\mu_l - c_l^\top \lambda^\top)_+ \right] - \sum_{t=1}^\tau \mathbb{E} [r_t x_t]
\]

\[
= \mathbb{E} \left[ \sum_{t=1}^\tau \lambda^\top \left( B - \sum_{t=1}^\tau a_t x_t \right) \right] + \mathbb{E} \left[ \sum_{t=1}^T \max_{l \in \mathcal{T}_t} (\mu_l - c_l^\top \lambda^\top)_+ - \sum_{t=1}^\tau (r_t - a_t^\top \lambda^\top)x_t \right]
\]

\[
= \lambda^\top \mathbb{E} [B_T] + \sum_{j \in J^*} \mathbb{E} \left[ \max_{l \in K_j} (\mu_l - c_l^\top \lambda^\top) \cdot n_j(T) - \sum_{l \in K_j} (\mu_l - c_l^\top \lambda^\top) n_l^\tau(\tau) \right]
\]

\[
+ \sum_{j \in J^*} \sum_{l \in K_j} (c_l^\top \lambda^\top - \mu_j) \cdot \mathbb{E} [n_l^\tau(\tau)],
\]

where the second inequality comes from the fact that we are dealing with the DLP (32) and the corresponding dual value is different, and the rest just follows.

### D.3 Proof of Proposition 4

We note that the proof will be very similar to the proof of Proposition 2, therefore we will only point out the differences. Firstly, we define events such that

\[ \mathcal{A}_t^{(j)} = \left\{ \left\lfloor \frac{n_j(t-1)}{t-1} - p_j \right\rfloor \leq L \right\}. \]
and start to bound \( \sum_{l \in K_j} \mathbb{E} [n_l^j (\tau_S)] \) for \( j \in J' \). With the same decomposition we have

\[
\sum_{l \in K_j} \mathbb{E} [n_l^j (\tau_S)] \leq \mathbb{E} \left[ \sum_{t=1}^{T} I \{ \text{Accept any order in order group } j \in J' \text{ at time } t \} \cap \{ t < \tau_S \} \right] \\
\leq \sum_{t=1}^{T} \mathbb{P} \{ \text{Accept any order in order group } j \in J' \text{ at time } t \} \cap \{ t < \tau_S \} .
\]

Notice that under the event \( \{ t < \tau_S \} \cap \left( \cap_{j'=1}^{n} A_{j'}^{(2)} \right) \), the perturbed LP (33) has the same optimal basis as (32). Hence for \( j \in J' \), from complementary slackness condition we know \( x_1 \cdot (c_i^\top \lambda - \mu_l) = 0 \) for all \( l \in K_j \). Therefore we will reject all the order in order group \( j \) with probability 1, and

\[
\{ \text{Accept any order in order group } j \in J' \text{ at time } t \} \cap \{ t < \tau_S \} \subseteq \left( \bigcup_{j'=1}^{n} A_{j'}^{(1)} \right) \cap \{ t < \tau_S \} .
\]

The rest of bounding \( \sum_{l \in K_j} \mathbb{E} [n_l^j (\tau_S)] \) follows the same approach in Proposition 2.

Next, to bound the term \( \mathbb{E} \left[ \max_{l \in K_j} (\mu_l - c_l^\top \lambda^*) \cdot n_j (\tau') - \sum_{l \in K_j} (\mu_l - c_l^\top \lambda^*) n_l^j (\tau') \right] \) for \( j \in J^* \). We define the suboptimal order \( l \) in order group \( j \) to be the order such that \( \mu_l - c_l^\top \lambda^* < \max_{l \in K_j} (\mu_l - c_l^\top \lambda^*) \), and define the optimal order \( l \) in order group \( j \) to be such that \( \mu_l - c_l^\top \lambda^* = \max_{l \in K_j} (\mu_l - c_l^\top \lambda^*) \). Notice that

\[
\mathbb{E} \left[ \max_{l \in K_j} (\mu_l - c_l^\top \lambda^*) \cdot n_j (\tau') - \sum_{l \in K_j} (\mu_l - c_l^\top \lambda^*) n_l^j (\tau') \right] \\
\leq 2 \max_{l \in K_j, j \in [n]} |\mu_l - c_l^\top \lambda^*| \\
\cdot \mathbb{E} \left[ \sum_{t=1}^{T} I \{ \text{Not accepting anything, or accepting the suboptimal } l \in K_j \text{ and } j \in J^* \text{ at time } t \} \right].
\]

Next, we analyze the case that happens with probability 1 under the event \( \{ t < \tau_S \} \cap \left( \cap_{j'=1}^{n} A_{j'}^{(2)} \right) \), where the optimal basis for the perturbed LP (33) and DLP (32) are the same. Observe that for \( j \in J^* \), no matter if there are one or multiple optimal order \( l \) in group \( j \), we will always have \( 1^\top y_j = 1 \). This is again because of the complementary slackness condition: firstly, we have \( (1-1^\top y_j) (\max_{l \in K_j} (\mu_l - c_l^\top \lambda^*)_+) = 0 \), and this ensures \( 1^\top y_j = 1 \); secondly, we have

\[
y_l \cdot (c_l^\top \lambda^* - \mu_l + \max_{l \in K_j} (c_l^\top \lambda^* - \mu_l)_+) = 0,
\]

and this ensures that for \( j \in J^* \), \( y_l > 0 \) if and only if it is the optimal order. Henceforth, we have

\[
\{ \text{Not accepting anything, or accepting the suboptimal } l \in K_j \text{ and } j \in J^* \text{ at time } t \} \cap \{ t < \tau_S \} \\
\subseteq \left( \bigcup_{j'=1}^{n} A_{j'}^{(1)} \right) \cap \{ t < \tau_S \}.
\]

Lastly, the rest of the proof follows the proof of Proposition 2.

#### E Proof of Lemma 1 and LP’s Stability

Here we explore a stability property for LPs by identifying conditions under which a small perturbation of LP’s input will not change the optimal basis and the bindingness of the constraints. As a side product, we prove Lemma 1 and relate the constant \( L \) therein with a number of parameters of the underlying LP.

In the context of our paper, the perturbation of LP can be viewed as the estimation error of using (5)
as a proxy for (3). We emphasize that the property is not pertaining to the resource allocation LP, so we present the result under the general standard form,

\[
\begin{align*}
\min & \quad c^\top x, \\
\text{s.t.} & \quad Ax = b, \\
& \quad x \geq 0.
\end{align*}
\]

(35)

where \(c \in \mathbb{R}^n\), \(A \in \mathbb{R}^{m \times n}\), and \(b \in \mathbb{R}^n\). We overload the notations a bit: within this subsection, the LP’s input \(c\), \(A\) and \(b\) (and their dimensions \(m, n\)) all refer to a general vector or matrix, different from their contextual meanings in the previous sections. As the convention, we define the basic/non-basic variable set

\[
B^* = \{ j : x_j^* > 0, j = 1, ..., n \},
\]

\[
B^' = \{ j : x_j^* = 0, j = 1, ..., n \}
\]

where \(B^*\) is also known as the optimal basis. Consider a second LP of the same size as (35),

\[
\begin{align*}
\min & \quad \hat{c}^\top x, \\
\text{s.t.} & \quad \hat{A}x = \hat{b}, \\
& \quad x \geq 0,
\end{align*}
\]

(36)

where \(\hat{c} \in \mathbb{R}^n\), \(\hat{A} \in \mathbb{R}^{m \times n}\), and \(\hat{b} \in \mathbb{R}^n\). We can interpret \(\hat{c}, \hat{A}, \hat{b}\) as sample-based estimates or perturbations of their counterparts in (35). The following proposition states the conditions under which the two LPs (35) and (36) share the same index set of basic variables.

**Proposition 5.** Suppose the optimal solution of (35) is unique and nondegenerate. Define

\[
\chi := \min \{ x_j^* : x_j^* > 0 \},
\]

\[
\sigma := \sigma_{\min}(A_{B^*}),
\]

\[
\delta := \min \{ A_j^\top \lambda^* - c_j : A_j^\top \lambda^* - c_j > 0 \},
\]

where \(A_{B^*}\) is the sub-matrix of \(A\) containing the columns in \(B^*\), \(A_j\) is the \(j\)-th column of the matrix \(A\), and \(\lambda^*\) is the dual optimal solution of (35).

If the following conditions hold

\[
\begin{align*}
\| \hat{A}_j - A_j \|_{\infty} & \leq \begin{cases} \frac{\min\{1, \sigma, \sigma^2\} \cdot \min\{\chi, \delta\}}{12m^2 \sqrt{m}}, & \text{for } j \in B^*, \\ \frac{\sigma \delta}{12m}, & \text{for } j \in B', \end{cases} \\
|\hat{c}_j - c_j| & \leq \begin{cases} \frac{\sigma \delta}{12m}, & \text{for } j \in B^*, \\ \frac{\delta}{6}, & \text{for } j \in B', \end{cases} \\
|\hat{b}_i - b_i| & \leq \frac{\sigma \chi}{8\sqrt{m}}, \quad \text{for } i = 1, ..., m,
\end{align*}
\]

(37)

then the LP (36) has the same index sets of basic and non-basic variables as the LP (35). In addition, the optimal solution of (35) is also unique and nondegenerate.

The condition in the proposition is expressed by the deviation of the perturbed LP’s inputs from the original LP’s input. It relates the stability of \(B^*\) with a few quantities of the LP: \(\chi\) captures the stability of the original primal optimal solution; \(\sigma\) describes the singularity of the constraint matrix restricted to the columns in \(B^*\); \(\delta\) is a sub-optimality measure for the non-basic variables and it is computed based
on the reduced costs. The proof of the proposition is based on standard linear algebra analysis and it formalizes the intuition that the optimal basis \( B^* \) should exhibit some continuity with respect to the LP’s input.

The nondegeneracy assumption (Assumption 2) can be further illustrated from Proposition 5. The role of the assumption is to ensure \( \chi, \sigma \) and \( \delta \) defined in the proposition to be positive, and consequently, the stability property holds for the DLP (3). To achieve a bounded regret for the online resource allocation problem, the nondegeneracy assumption is indeed necessary so that the parameters \( \chi, \sigma \) and \( \delta \) can be treated as constant and not dependent on \( T \). On one hand, when the nondegeneracy assumption is violated or when the parameters such as \( \chi, \sigma \) and \( \delta \) may scale with \( \frac{\sqrt{T}}{T} \) or \( \frac{1}{T} \), there can be examples for which constant regret is not achievable (Arlotto and Gurvich, 2019; Bumpensanti and Wang, 2020). On the other hand, the constant condition is arguably reasonable in the nondegeneracy context in that these parameters are computed based on the DLP (3) and its standard form both of which bear no dependence of \( T \).

E.1 Proof of Proposition 5
Recall that \( B^* \) and \( B' \) denote the optimal and non-optimal basis for (35), respectively. The idea here is to show that the perturbed LP (36) has the same optimal and non-optimal basis under condition (37). Consider a basic solution \( \hat{x} \) of the perturbed LP (36) defined as follows. If \( \hat{A} B^* \) is invertible, we can define \( \hat{x} \) as

\[
\begin{align*}
(\hat{x})_{B^*} &= (\hat{A} B^*)^{-1} \hat{b}, \\
(\hat{x})_{B'} &= 0
\end{align*}
\]

Next, we prove the following results:

(a) The matrix \( \hat{A} B^* \) is non-singular and thus \( (\hat{x})_{B^*} \) is a well-defined basic solution.

(b) \( (\hat{x})_{B^*} > 0 \), and thus \( \hat{x} \) is a basic feasible solution of (36).

(c) The reduced costs associated with non-basic variables in \( B' \) are all negative and hence \( \hat{x} \) is the unique optimal solution of the perturbed LP (36).

Throughout the proof, we use \( A^e \) to denote \( \hat{A} - A \), and similarly for \( c^e \) and \( b^e \). To show part (a), we prove that the smallest singular value of the matrix is positive. We use \( \sigma_{\min}(M) \) and \( \sigma_{\max}(M) \) to denote the smallest and the largest singular value of a matrix \( M \). Then we have

\[
\sigma_{\min}(\hat{A} B^*) \geq \sigma_{\min}(A_{B^*}) - \sigma_{\max}(-A_{B^*}^e)
\]

\[
= \sigma - \sigma_{\max}(-A_{B^*}^e)
\]

\[
\geq \sigma - \sqrt{m} \|A_{B^*}^e\|_{\infty},
\]

where the first line comes from Weyl’s inequality on matrix eigenvalues/singular values, the second line comes from the definition of \( \sigma \), and the third line is obtained from the relation between the spectral norm and the infinity norm of a matrix. From condition (37) and \( \|A_{B^*}^e\|_{\infty} \leq m \max_{i \in B^*} \|A_i\|_{\infty} \), we have

\[
\sigma_{\min}(\hat{A} B^*) \geq \frac{\sigma}{2},
\]
and consequently,
\[ \sigma_{\text{max}} \left( \left( \hat{A}_{B'}^{-1} \right) \right) \leq \frac{2}{\sigma}. \tag{38} \]

For part (b), we show \( \hat{x} \) is a feasible basic solution. From Assumption 2 we know that
\[ (x^*)_{B'} \geq \chi > 0, \]
where the inequality holds element-wise. To ensure that \( (\hat{x})_{B'} \) is strict positive, it suffices to show
\[ \| (\hat{x})_{B'} - (x^*)_{B'} \|_\infty = \| (A + A^*)_{B'}^{-1} (b + b') - (A_{B'})^{-1} b \|_\infty \leq \frac{\chi}{2}. \]

From condition (37), if \( \max_{i \in B'} \| A_i \|_\infty \leq \frac{\sigma^2 \chi}{8m\sqrt{m}} \) and \( \| b' \|_\infty \leq \frac{\sigma \chi}{\sqrt{m}} \), then
\[
\begin{align*}
\| (A + A^*)_{B'}^{-1} (b + b') - (A_{B'})^{-1} b \|_\infty & \leq \| (A + A^*)_{B'}^{-1} b - (A_{B'})^{-1} b \|_\infty + \| (A + A^*)_{B'}^{-1} b' \|_\infty \\
& \leq \| (A + A^*)_{B'}^{-1} b - (A_{B'})^{-1} b \|_\infty + \| (A + A^*)_{B'}^{-1} b' \|_\infty \\
& \leq \| (A_{B'})^{-1} (A + A^*)_{B'}^{-1} (b + b') - (A_{B'})^{-1} b \|_\infty + \| (A + A^*)_{B'}^{-1} b' \|_\infty \\
& \leq \| (A_{B'})^{-1} (A + A^*)_{B'}^{-1} \|_{\infty} \| A_{B'} \|_{\infty} + \| (A + A^*)_{B'}^{-1} b' \|_\infty \\
& \leq \frac{2\sqrt{m} \sigma_{\text{max}}}{\sigma^2} \| A_{B'} \|_{\infty} + \frac{2\sqrt{m} \sigma}{\sigma} \| b' \|_\infty \\
& \leq \frac{2\sqrt{m} \sigma}{\sigma^2} \max_{i \in B'} \| A_i \|_{\infty} + \frac{2\sqrt{m} \sigma}{\sigma} \| b' \|_\infty \leq \frac{\chi}{2}.
\end{align*}
\]

The third and seventh line come from the sub-multiplicativity of matrix \( L_\infty \) norm. The eighth line come from the definition of \( \sigma \) following Assumption 2 and the relation between the spectral norm \( \sigma_{\text{max}} \) and \( L_\infty \) norm. The last line is from the inequality \( \| A_{B'} \|_{\infty} \leq \max_{i \in B'} \| A_i \|_{\infty} \) and condition (37). Thus we finish the part on the feasibility.

For part (c), we prove that reduced costs of non-basic variables in \( B' \) are all strictly negative. For a non-basic variable \( x_i \), the reduced cost of the perturbed LP (36) (denoted by \( \hat{\psi}_i \)) can be expressed as follows,
\[
\hat{\psi}_i := c_i + c_i^* - (c + c^*)_{B'}^\top (A + A^*)_{B'}^{-1} (A_i + A_i^*)
\[
= c_i - c_{B'}^\top (A_{B'})^{-1} A_i + c_{B'}^\top (A_{B'})^{-1} A_i + c_i^* - (c + c^*)_{B'}^\top (A + A^*)_{B'}^{-1} (A_i + A_i^*)
\[
= \hat{\psi}_i + c_i^* + c_{B'}^\top ((A_{B'})^{-1} - (A + A^*)_{B'}^{-1}) A_i - c_{B'}^\top (A + A^*)_{B'}^{-1} A_i
\[
- c_{B'}^\top (A + A^*)_{B'}^{-1} A_i^* + c_{B'}^\top (A + A^*)_{B'}^{-1} A_i^*.
\tag{39}
\]

Since the reduced cost of LP (35)
\[ \psi_i := c_i - A_i^\top \lambda^* > 0 \]
for \( i \in B' \), a sufficient condition for \( \hat{\psi}_i > 0 \) is the absolution values of all the rest five terms are no greater than \( \| \hat{\psi}_i \| \). Next, we are going to bound each component in (39).

For \( c_i^* \), the inequality \( |c_i^*| \leq \frac{\| \hat{\psi}_i \|}{\sigma} \) is directly implied from the condition (37). For the second term in
we have

\[ |\hat{e}_{G^*}^\top ((A_{G^*})^{-1} - (A + A')_{G^*}^{-1}) A_i| \leq \|e_{G^*}\|_1 \cdot \|((A_{G^*})^{-1} - (A + A')_{G^*}^{-1}) A_i\|_\infty \]

\[ \leq m \|(A_{G^*})^{-1} (I - A_{G^*}(A + A')_{G^*}^{-1}) A_i\|_\infty \]

\[ = m \|(A_{G^*})^{-1} A_{G^*} (A + A')_{G^*}^{-1} A_i\|_\infty \]

\[ \leq m \|(A_{G^*})^{-1}(A + A')_{G^*}^{-1}\|_\infty \|A_{G^*}\|_\infty \|A_i\|_\infty \]

\[ \leq m \sqrt{m} \sigma_{\max} ((A_{G^*})^{-1}(A + A')_{G^*}^{-1}) \|A_{G^*}\|_\infty \]

\[ \leq \frac{2m \sqrt{m}}{\sigma^2} \|A_{G^*}\|_\infty \]

\[ \leq \frac{2m}{\sigma} \|e_{G^*}\|_\infty \]

where the first line is obtained by Holder’s inequality, the fourth line is obtained by the sub-multiplicativity and (38), and the sixth line comes from the relation between the spectral norm and \(L_\infty\) norm, and the last line is again from \(\|A_{G^*}\|_\infty \leq m \max_{i \in B^*} \|A_i\|_\infty\). Thus, from (37) we have

\[ |\hat{e}_{G^*}^\top ((A_{G^*})^{-1} - (A + A')_{G^*}^{-1}) A_i| \leq \frac{\delta}{6}. \]

For the third term in (39), we have

\[ |\hat{e}_{G^*}^\top (A + A')_{G^*}^{-1} A_i| \leq \|e_{G^*}\|_2 \|(A + A')_{G^*}^{-1} A_i\|_2 \]

\[ \leq \sqrt{m} \|e_{G^*}\|_\infty \sigma_{\max} ((A + A')_{G^*}^{-1}) \|A_i\|_2 \]

\[ \leq m \|e_{G^*}\|_\infty \sigma_{\max} ((A + A')_{G^*}^{-1}) \]

\[ \leq \frac{2m}{\sigma} \|e_{G^*}\|_\infty . \]

Thus, from (37) again we have

\[ |\hat{e}_{G^*}^\top (A + A')_{G^*}^{-1} A_i| \leq \frac{\delta}{6}. \]

In a similar manner, for the last two terms we have

\[ |\hat{e}_{G^*}^\top (A + A')_{G^*}^{-1} A_i'| \leq \frac{2m}{\sigma} \max_{i \in B'} \|A_i'|_\infty , \]

\[ |\hat{e}_{G^*}^\top (A + A')_{G^*}^{-1} A_i'| \leq \frac{2m}{\sigma} \max_{i \in B'} \|A_i'|_\infty \|e_{G^*}\|_\infty . \]

Both of them are no longer than \(\frac{\delta}{6}\) condition because of (37).

Therefore, for any non-basic variable \(x_i\) with \(i \in B'\), we conclude that its reduced cost in the perturbed LP \(\hat{\Psi}; \hat{\xi} \geq \frac{\delta}{6} > 0\). Thus we establish the optimality of the solution \(\hat{x}\). Lastly, given the stability of the sign for the reduced-cost, we know the optimal solution is unique.

### E.2 Proof of Lemma 1

Firstly, we apply the result of Proposition 5 to the DLP (3) (also the equivalent form (6)). Notice that in Proposition 5 the matrix dimension is \(m \times n\), while in the canonical form of the DLP (6), the dimension becomes \((n + m) \times (2n + m)\). However, instead of plugging in the dimension to the result of Proposition 5 directly, one can find that in (6), only \(C\), \(\mu\) and \(b\) component will have random perturbation, and other entries will always be 1. Having this observation, we know that under Assumption 1, 2, and the
condition that

\[
\|\hat{C}_j - C_j\|_\infty \leq \begin{cases} 
\frac{\min\{1, \sigma^2\} \cdot \min\{\chi, \delta\}}{12n^2 \sqrt{n+m}} \cdot \frac{\sigma}{\delta}, & \text{for } j \in J^*, \\
\frac{\sigma \delta}{12 \sqrt{n(n+m)}} \cdot \frac{\delta}{8 \sqrt{n+m}}, & \text{for } j \in J', 
\end{cases}
\]

\[
|\hat{\mu}_j - \mu_j| \leq \begin{cases} 
\frac{\sigma \delta}{12 \sqrt{n(n+m)} \cdot \frac{\delta}{8 \sqrt{n+m}}}, & \text{for } j \in J^*, \\
\frac{\sigma}{\delta}, & \text{for } j \in J', 
\end{cases}
\]

\[
|\hat{b}_i - b_i| \leq \frac{\sigma \chi}{8 \sqrt{n+m}}, \text{ for } i \in I^*
\]

\[
\hat{b}_i - b_i \geq -\frac{\sigma \chi}{8 \sqrt{n+m}}, \text{ for } i \in I'.
\]

the optimal solution to LP (8) is unique and it shares the same optimal basis and binding/non-binding structure with the solution of LP (6) (which equivalent to (3)). The last step is to show that the condition above could be extended to

\[
\|\hat{C}_j - C_j\|_\infty \leq \begin{cases} 
\frac{\min\{1, \sigma^2\} \cdot \min\{\chi, \delta\}}{12n^2 \sqrt{n+m}} \cdot \frac{\sigma}{\delta}, & \text{for } j \in J^*, \\
\frac{\sigma \delta}{12 \sqrt{n(n+m)} \cdot \frac{\delta}{8 \sqrt{n+m}}}, & \text{for } j \in J', 
\end{cases}
\]

\[
|\hat{\mu}_j - \mu_j| \leq \begin{cases} 
\frac{\sigma \delta}{12 \sqrt{n(n+m)} \cdot \frac{\delta}{8 \sqrt{n+m}}}, & \text{for } j \in J^*, \\
\frac{\sigma}{\delta}, & \text{for } j \in J', 
\end{cases}
\]

\[
|\hat{b}_i - b_i| \leq \frac{\sigma \chi}{8 \sqrt{n+m}}, \text{ for } i \in I^*
\]

\[
\hat{b}_i - b_i \geq -\frac{\sigma \chi}{8 \sqrt{n+m}}, \text{ for } i \in I'.
\]

To show this statement, we only have to look at the dual solution of the LP (8). Since the primal is nondegenerate, its dual solution must be unique. It is easy to construct a primal solution for a larger \( \hat{b}_i \) where \( i \in I' \): we just add the corresponding increased value to the primal slack variable, and it remains to show that the constructed primal solution is unique. The increase for \( \hat{b}_i \) where \( i \in I' \) will not improve objective value because it corresponds to the dual variable that \( \lambda^*_i = 0 \), and moreover, it will not bring any new optimal dual solutions (otherwise the original optimal primal solution set would be different). Therefore, the optimal primal solution remains unique for \( \hat{b}_i > b_i - \frac{\sigma \chi}{8 \sqrt{n+m}} \). Therefore if we take

\[
L = \min \left\{ \frac{\min\{1, \sigma^2\} \cdot \min\{\chi, \delta\}}{12n^2 \sqrt{n+m}}, \frac{\sigma \delta}{12 \sqrt{n(n+m)}}, \frac{\sigma \delta}{12 \sqrt{n(n+m)}} \cdot \frac{\delta}{8 \sqrt{n+m}} \cdot \frac{\sigma \chi}{8 \sqrt{n+m}} \right\},
\]

we finish proving the statement of the lemma. Here the parameters \( \sigma, \chi, \delta \), follow their definitions in Proposition 5, but are based on the standard form DLP (6).