A Single-Sample Polylogarithmic Regret Bound for Nonstationary Online Linear Programming

Haoran Xu*
Department of Management Science and Engineering
Stanford University

Owen Shen
Operations Research Center
Massachusetts Institute of Technology

Peter Glynn
Department of Management Science and Engineering
Stanford University

Yinyu Ye
Department of Management Science and Engineering
Stanford University

Patrick Jaillet
Operations Research Center
Massachusetts Institute of Technology

Abstract

We study nonstationary online linear programming (OLP), where orders arrive sequentially and follow independent but not-identical reward-resource distributions. The decision-maker seeks to maximize the expected total reward by making immediate and irrevocable acceptance or rejection decisions for each order, subject to a resource endowment available at the beginning of the planning horizon. Such problems arise in resource-constrained ML and AI systems, including inference admission control, online advertising and recommendation, and data-acquisition pipelines, where the mix and value of requests may change over time. We focus on a minimal-information regime in which the decision-maker observes only one independent sample from each future distribution before the horizon begins. We propose a novel re-solving algorithm that integrates a dynamic programming perspective with the dual-based frameworks traditionally employed in stationary environments. In the large-resource regime, where the resource endowment scales linearly with the number of orders, we prove that our algorithm achieves $O((\log n)^2)$ regret across a broad class of nonstationary distribution sequences. Our results demonstrate that polylogarithmic regret is attainable even under significant environmental shifts and minimal data availability, bridging the gap between stationary OLP and more volatile real-world resource allocation problems.

1 Introduction

Many modern ML and AI systems make resource-allocation decisions online. An inference platform must decide which requests to serve under GPU, latency, or energy budgets; an advertising or recommendation system allocates limited impressions across a changing stream of users; and a data-acquisition pipeline chooses which examples to label or process under a fixed budget. In each case, requests arrive sequentially, each accepted request consumes scarce resources, and the decision must be made before future requests are observed.

*Corresponding author: haoran14@stanford.edu

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Online Linear Programming (OLP) (Agrawal et al. [2014]) provides a canonical model for such problems. Each arriving request has a reward and a resource-consumption vector, and the decision-maker irrevocably accepts or rejects the request subject to a finite resource endowment. OLP and closely related models have been widely used to study online advertising, revenue management, and other resource-allocation problems (Li and Ye [2022]). A central question is whether simple price-based policies can achieve low regret relative to an offline benchmark that observes the full sequence in advance.

Classical OLP typically assumes that requests are drawn from a fixed distribution. This assumption is often unrealistic in deployed systems: user populations shift over the day, model releases change demand patterns, advertising campaigns evolve, and compute workloads fluctuate with product usage. These changes induce nonstationarity in both rewards and resource consumptions. Ignoring these changes can lead to poor performance, such as running out of resources too early or missing high-value opportunities. Thus, developing OLP algorithms that can adapt to these nonstationary environments is critical. These algorithms must maintain strong performance over different kinds of nonstationary environment and can be implemented when data is scarce and limited historical information is available.

In light of these challenges, this paper seeks to address nonstationary OLP problems. Specifically, we investigate the case that the reward-resource consumption pairs follow a sequence of independent but not necessarily identical distributions. We focus on a minimal data availability situation in which only one sample from each of these distributions are available at the start of the planning horizon. The main research question of this paper is: can we design a near-optimal algorithm for nonstationary OLP under this single-sample setting?

1.1 Main Contribution

In this work, we consider a nonstationary OLP framework involving n sequentially arriving orders and a fixed resource endowment provided at the start of the planning horizon. The decision-maker must make immediate and irrevocable acceptance or rejection decisions for each order to maximize the total expected reward, subject to the available resource endowment. We allow the marginal distributions of both rewards and resource consumptions to be independent but non-identical. We assume the marginal distributions of resource consumption share a common support—which may be discrete or continuous—while the conditional distribution of rewards, given resource consumption, is required to be continuous. Notably, we study the single-sample setting: at the start of the horizon, the decision-maker has access to only one sample from each of the n reward-resource consumption distributions. Furthermore, we assume a large-resource regime where the resource endowment scales linearly with the number of orders.

Adopting a dynamic programming perspective, we propose a dual-based pricing algorithm for nonstationary OLP. We define the system state as the average remaining resource. Upon the arrival of each order, the algorithm re-solves a linear program formulated using the current system state and the provided samples of both current and future orders. The optimal dual solution of this linear program yields a price that serves as a threshold, and an order is accepted if its reward exceeds this threshold. We measure the algorithm's performance using regret, defined as the difference between the expected optimal total reward of the offline problem and the expected reward earned by our algorithm. Under a set of boundedness and smoothness assumptions on the nonstationarity, we prove that our algorithm achieves $O(\log^2 n)$ regret for any distribution sequence satisfying these assumptions.

Our regret analysis centers on two pillars: establishing a uniform dual convergence result for the algorithm's prices and characterizing the evolution of the system state. We first demonstrate the existence of a near-optimal offline policy that mirrors our algorithm's structure but utilizes a constant price throughout the planning horizon. We then derive our regret bounds by proving that the prices generated by our algorithm closely track this fixed "offline price". A key theoretical finding of this analysis is that, despite the nonstationary arrivals and rewards, the prices utilized by the algorithm remain relatively stationary over the planning horizon.

Crucially, while the prices are stable, the system states themselves evolve nonstationarily. To analyze the relationship between the prices and the system states, we introduce the concept of δ-path, defined as a sequence of points in the state space over the planning horizon. We first construct a neighborhood around each point along the δ-path. Our uniform dual convergence result then establishes that if the

system state lies within the neighborhood of the j-th point of the δ-path when order j arrives, the mean squared error between the computed price and the offline price is $O(\frac{\log(n-j)}{n-j})$ uniformly over the neighborhood. Furthermore, we show that the system state remains within these neighborhoods until the final stages of the planning horizon. Together, these results yield the $O(\log^2 n)$ regret upper bound.

1.2 Related Works

The literature on Online Linear Programming (OLP) generally focuses on two primary models: the random permutation model (Agrawal et al. [2014]) and the stochastic input model (Li and Ye [2022]). This paper adopts the stochastic input model. Most research under this model assumes that rewards and resource consumptions follow independent and identically distributed (i.i.d.) processes. Within the i.i.d. setting, two major distributional assumptions prevail. The first assumes the joint distribution of reward and resource consumption has finite support, which typically leads to constant regret (Chen et al. [2022], Vera and Banerjee [2021]). The second assumes the conditional distribution of reward, given resource consumption, is continuous, which generally yields logarithmic regret (Li and Ye [2022], Bray [2024]). Our work focuses on the setting of continuous conditional distributions. Under this setting, when the available resources grow proportionally with the number of orders, Li and Ye [2022] proposed the Action-History-Dependent Learning Algorithm (AhdLA) and established an $O(\log n \log\log n)$ regret bound. Bray [2024] studied the same algorithm under slightly broader assumptions, establishing an $\Omega(\log n)$ regret lower bound and refining the dual convergence analysis to prove a tighter $O(\log n)$ upper bound. In this paper, we further generalize the problem setting by relaxing the i.i.d. requirement to consider distributions that are independent but non-identical.

AhdLA belongs to the class of dual-based "re-solving" algorithms, which compute prices by solving a dual linear program at each time period. These prices determine the thresholds used to accept or reject incoming orders. While such algorithms can achieve optimal regret (Bray [2024]), they are often computationally demanding. As an alternative, first-order methods compute prices via iterative gradient descent. While Li. et al. [2023] and Balseiro et al. [2023] establish $O(\sqrt{n})$ regret for such methods, Balseiro et al. [2023] further demonstrates that this performance holds for nonstationary inputs that are ergodic or seasonal. Jiang et al. [2025] explored nonstationary inputs without requiring ergodic or seasonal properties, assuming instead that rewards and consumptions are independent but non-identical. By introducing a "deviation budget" (W_n) to measure the accuracy of prior estimates of the distributions, they showed that their Informative Gradient Descent Algorithm achieves $O(\max\{\sqrt{n}, W_n\})$ regret. Cheung and Lyu [2025] proposes an algorithmic design framework for more general nonstationary online planning problems. When this framework is applied to the setting in Jiang et al. [2025], fewer samples are required to construct good prior estimates of the distributions to achieve $O(\sqrt{n})$ regret.

Moving to an even more constrained data environment, Ghuge et al. [2025] investigated the "single-sample" setting, where only one sample from each distribution is available. They proved that a $1-\epsilon$ approximation ratio is achievable, provided the resource endowment is $\tilde{\Omega}(1/\epsilon^6)$. Our paper also operates in the single-sample setting but leverages additional structural assumptions to derive superior bounds. We demonstrate that $O(\log^2 n)$ regret—equivalent to a $1 - O(\frac{\log^2 n}{n})$ approximation ratio—can be achieved with standard $\Theta(n)$ resources. This work further distinguishes itself from Feng et al. [2026], who also assume a single-sample setting but restrict resource consumption to finite support and require conditional reward distributions to remain identical over time. In contrast, our model allows both the conditional reward distributions to be non-identical and the resource consumption support to be infinite, while achieving a slightly improved regret upper bound of $O(\log^2 n)$.

In addition, the concept of variation budget introduced in Besbes et al. [2015] for general stochastic optimization is used in our work to formulate conditions under which polylogarithmic regret is achievable. Our work is also related to papers studying other classes of stochastic optimization problems in nonstationary environments, such as Bandits with Knapsacks (Besbes et al. [2014],Liu et al. [2022],Zhang and Cheung [2024]), Reinforcement Learning (Lykouris et al. [2025]), and online allocation of reusable resources (Zhang and Cheung [2025]).

Another related but distinct online optimization problem in which nonstationarity has been studied is online convex optimization (OCO). In OCO, over a planning horizon of length T, the learner chooses a decision x_t from a fixed convex set $\mathcal{K} \subset \mathbb{R}^d$ at each time step t, after which a convex loss

function f_t is revealed. The static regret is measured by the difference between $\sum_{t=1}^{T} f_t(x_t)$ and $\min_{x\in\mathcal{K}} \sum_{t=1}^{T} f_t(x)$, either in expectation or in the worst case. For a discussion of nonstationarity and dynamic regret in OCO, we refer readers to Zhao et al. [2024]. There are two major differences between this framework and OLP. First, OLP decisions are subject to a global resource capacity constraint. In contrast, OCO selects decisions from the same uncoupled feasible region at each time step. Consequently, at each time period, an OLP algorithm must balance earning immediate rewards against preserving resources for future orders, which is a trade-off that is absent in OCO. Second, in OLP, the decision is made after the current period's information is revealed, whereas in OCO, the decision is made before the loss function is revealed. These structural differences fundamentally alter the problem formulations and regret definitions, necessitating different informational requirements to guarantee sublinear regret in OLP and OCO. Despite these differences between OCO and OLP, the dual price learning step in our algorithm has a conceptual connection to OCO and to online learning more generally.

The paper is organized as follows. We introduce the problem formulation, assumptions, and algorithm in Section 2. A sketch of regret analysis is given in Section 3. Numerical experiments are provided in Section 4. We conclude and summarize our paper in Section 5. The proofs and complete regret analysis can be found in the appendix.

2 Problem Setup and Algorithm

In the problem of OLP, there are m types of resources available to the decision maker (DM) at the beginning of the planning horizon, and there are n orders arriving sequentially. Denote the total amount of available resources as $b_{0,n} \in \mathbb{R}_+^m$, and let $\{(u_{j,n}, a_{j,n})\}_{j=1}^n$ be a stochastic process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $(u_{j,n}, a_{j,n}) \in \mathbb{R}_+ \times \mathbb{R}_+^m$ is the pair of the reward and the resource consumption of the jth order of these n orders. When order j arrived, the DM needs to make a decision $x_{j,n}^\pi \in \{0, 1\}$, where π denotes the algorithm used by the DM. If $x_{j,n}^\pi = 1$, the DM accepts order j, consumes $a_{j,n}$ resources, and earns a reward of $u_{j,n}$. If $x_{j,n}^\pi = 0$, the DM rejects order j, and both consumes and earns nothing. Each decision made by the DM should be immediate and irrevocable without violating the budget constraint. More precisely, let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_{j,n}$ be the sigma algebra generated by $\{(u_{k,n}, a_{k,n})\}_{k=1}^j$, and we say π is a feasible policy if $x_{j,n}^\pi$ is measurable with respect to $\mathcal{F}_{j,n}$ for all j and $\sum_{j=1}^n a_{j,n} x_{j,n}^\pi \leq b_{0,n}$ almost surely. The total reward earned by algorithm π is $R_n(\pi) = \sum_{j=1}^n u_{j,n} x_{j,n}^\pi$, and the goal of the DM is to maximize the expected total reward $\mathbb{E}[R_n(\pi)]$ over all feasible policy π.

In this paper, we consider nonstationary OLP problem in the sense that, for each n, the stochastic process $\{(u_{j,n}, a_{j,n})\}_{j=1}^n$ can be nonstationary. In addition, we study our problem in the "large resource" scenario defined by the following assumption.

Assumption 1. *There exists $d_0 \in \mathbb{R}_+^m$ such that $b_{0,n} = n \cdot d_0$.*

We call d_0 the initial average resource, and we want to study the performance of an algorithm when n goes to infinity with d_0 fixed. The performance of an algorithm is measured by its regret. We call the following linear program (LP) as Offline LP, and denote its optimal objective value as R_n^{off}.

$$\begin{array}{lll} \max & \sum_{j=1}^n u_{j,n} x_{j,n} & \\ S.T & \sum_{j=1}^n a_{j,n} x_{j,n} \leq b_{0,n} & \\ & 0 \leq x_{j,n} \leq 1 & \forall 1 \leq j \leq n \end{array} \tag{1}$$

Because R_n^{off} is given by the linear relaxation of the resource allocation problem, and computing R_n^{off} requires knowing information of all the n orders in advance, $R_n^{off} \geq R_n(\pi)$ almost surely. Let P_n be the joint distribution of $\{(u_{j,n}, a_{j,n})\}_{j=1}^n$ and $P = \{P_n\}_{n=1}^{+\infty}$ be a sequence of distributions. When there are n orders in total, the regret of an algorithm π under P is defined to be

$$Regret_n(\pi, P) = \mathbb{E}_{P_n}\left[R_n^{off} - R_n(\pi)\right]$$

When nonstationarity is considered in OLP problems, an algorithm that performs well over a reasonably large class of distributions of the nonstationary process $\{(u_{j,n}, a_{j,n})\}_{j=1}^n$ is usually desired. Let $\mathcal{P}$ be the family of sequences of distributions that we are interested in. The goal of this

paper is to find appropriate regularity conditions on $\mathcal{P}$ and construct an algorithm π under which there exists a constant C such that for all $P \in \mathcal{P}$, when n is large enough,

$$Regret_n(\pi, P) \leq C(\log n)^2 \tag{2}$$

2.1 Single-sample Setting

We certainly prefer less restrictive regularity conditions so that the proposed algorithm works on larger $\mathcal{P}$; however, how strong the regularity conditions are depends on how much we know about the nonstationarity. When the distribution of $\{(u_{j,n}, a_{j,n})\}_{j=1}^n$ is completely unknown for each n, we provide the following example shows that, even if there are only two sequences of distributions in $\mathcal{P}$ and each of the sequence of distributions is uniformly bounded, there is no feasible online algorithm that has sublinear regret for all $P \in \mathcal{P}$

Example 1. *Let $\mathcal{P} = \{P_1, P_2\}$ and $P_{i,n}$ be the nth joint distribution in P_i. Under both $P_{1,n}$ and $P_{2,n}$, $a_{j,n} = 1$ almost surely for all j. Under $P_{1,n}$, $\{u_{j,n}\}_{j=1}^n$ a sequence of i.i.d. $Uniform(0,1)$ random variables. Under $P_{2,n}$, $\{u_{j,n}\}_{j=1}^n$ is a sequence of independent random variables such that $u_{j,n} \sim Uniform(0,1)$ if $j \leq \lceil \frac{n}{2} \rceil$, and $u_{j,n} \sim Uniform(2,3)$ otherwise. The initial average resource is $d_0 = \frac{1}{4}$. Then, $\max_{i\in\{1,2\}}\{Regret_n(\pi, P_i)\}$ is $\Omega(n)$ for any feasible algorithm π.*

The proof of Example 1 is in Appendix A. Thus, to achieve polylogarithmic regret for each sequences of distributions in a reasonably large distribution classe $\mathcal{P}$, some information regarding the nonstationarity is required. In this paper, we focus on the following *single-sample* setting.

Assumption 2. *$\{\tilde{u}_{j,n}, \tilde{a}_{j,n}\}_{j=1}^n$ is available at the beginning of the planning horizon such that (i) $\{\tilde{u}_{j,n}, \tilde{a}_{j,n}\}_{j=1}^n$ and $\{u_{j,n}, a_{j,n}\}_{j=1}^n$ are independent. (ii) $\{\tilde{u}_{j,n}, \tilde{a}_{j,n}\}_{j=1}^n$ and $\{u_{j,n}, a_{j,n}\}_{j=1}^n$ are equal in distribution.*

The single-sample setting can be justified by natural periodicity of operations in real-world supply chain management and AI systems. A common pattern in practical resource allocation is a sequence of operational cycles where, at the beginning of each cycle, resources are provided for future allocation within the cycle. While the joint distributions of reward and resource consumption are often nonstationary within a single cycle, the cycles themselves are typically structured such that these distributions remain consistent across consecutive cycles. Consequently, historical data from a previous cycle can naturally serve as the samples required by Assumption 2 for the current cycle.

2.2 Algorithm

Let $\{x_{j,n}^{off}\}_{j=1}^n$ be the optimal solution of the offline LP (1) and define

$$p_{j,n}(d) = \arg\min_{p\geq 0} d^T p + \frac{1}{n-j}\sum_{k=j+1}^{n} (u_{k,n} - {a_{k,n}}^T p)^+$$

Note that the optimization problem defining $p_{0,n}(d_0)$ is equivalent to the dual of the offline LP. Let $d_{j,n}^{off} = (b_{0,n} - \sum_{k=1}^{j} a_{k,n} x_{k,n}^{off})/(n-j)$, which is the sequence average remaining resources when the optimal solution of the offline LP is taken sequentially. By complementary slackness, for all j,

$$x_{j,n}^{off} = \begin{cases} 1 & \text{if } u_{j,n} > {a_{j,n}}^T p_{j-1,n}\left(d_{j-1,n}^{off}\right) \\ 0 & \text{if } u_{j,n} < {a_{j,n}}^T p_{j-1,n}\left(d_{j-1,n}^{off}\right) \end{cases} \tag{3}$$

(3) is a generalization of the dual decision rule studied in Li and Ye [2022], and it inspires us a special dynamic programming (DP) perspective of OLP problems. After making decisions on order j, we observe $d_{j,n} = b_{j,n}/(n-j)$ where $b_{j,n}$ is the remaining resources, and we can view $d_{j,n}$ to be the system state. Instead of learning the value function as in the standard dynamic programming, when $u_{j+1,n} = {a_{j+1,n}}^T p_{j,n}\left(d_{j,n}\right)$ happens rarely (ex. when the conditional distribution of $u_{j,n}$ given $a_{j,n}$ is continuous), (3) implies that we can directly learn the price function $p_{j,n}(\cdot)$, and then plug the learned price function into the threshold rule in (3) to accept or reject orders. To learn $p_{j,n}(\cdot)$, we can view it as an estimator of the following population price function

$$p_{j,n}^*(d) = \arg\min_{p\geq 0} d^T p + \frac{1}{n-j}\sum_{k=j+1}^{n} \mathbb{E}\left[(u_{k,n} - {a_{k,n}}^T p)^+\right]$$

If we replace $\{(u_{k,n}, a_{k,n}\}_{k=j+1}^{n}$ in the definition of $p_{j,n}(\cdot)$ with the samples $\{(\tilde{u}_{k,n}, \tilde{a}_{k,n}\}_{k=j+1}^{n}$ in Assumption 2, we create another identical and independent estimator of the population price function. Suppose the estimation error between $p_{j,n(\cdot)}$ and $p^*_{j,n}(\cdot)$ decreases fast when $n-j$ gets larger, price functions evaluated using the samples given in Assumption 2 should be a good choice for making decisions when n is large enough. Details of our algorithm motivated by the above DP perspective are given in Algorithm 1.

Algorithm 1

Input the initial average resource d_0 and a random sample $\left\{(\tilde{u}_{j,n}, \tilde{a}_{j,n})\right\}_{j=1}^{n}$.
Set $b_{0,n} = nd_0$
for $j = 1, \cdots, n$ **do**

$$p^{\pi}_{j-1,n} = \arg\min_{p \geq 0} {b_{j-1,n}}^T p + \sum_{k=j}^{n} \left(\tilde{u}_{k,n} - \tilde{a}_{k,n}^T p\right)^+$$

$$x^{\pi}_{j,n} = \mathbb{1}\left\{u_{j,n} > {a_{j,n}}^T p^{\pi}_{j-1,n}\right\} \mathbb{1}\left\{b_{j-1,n} \geq a_{j,n}\right\}$$

$$b_{j,n} = b_{j-1,n} - a_{j,n} x^{\pi}_{j,n}$$

end for

2.3 Regularity Assumptions on Distributions Classes

To make Algorithm 1 achieve the goal of $O((\log n)^2)$ regret as stated in (2), we need the following regularity assumptions on the distribution class $\mathcal{P}$.

Assumption 3. *There exist positive real numbers α, $\underline{a}$, λ, $\bar{L}$, $\bar{u}$, $\bar{\mu}$, $\underline{\mu}$, $\epsilon_{\mathcal{P}}$ and $\Xi \subseteq [0, \alpha]^m$ such that for any $P \in \mathcal{P}$,*

(a) $\left\{\left(u_{j,n}, a_{j,n}\right)\right\}_{j=1}^{n}$ *is a sequence of independent random pair in $\mathbb{R}_+ \times \mathbb{R}_+^m$ for all n.*

(b) Ξ is the support of $a_{j,n}$, the smallest entry of $\mathbb{E}[a_{j,n}]$ is greater than $\underline{a}$, and the smallest eigenvalue of $\mathbb{E}_{P_n}\left[a_{j,n}{a_{j,n}}^T\right]$ is greater than λ for all j and n.

(c) There exists a positive continuously differentiable function $v : [0,1] \times [0,\alpha]^m \to [0, \bar{\mu}]$ such that $\|\nabla v\|_2 \leq \bar{L}$ and the restriction of $v\left(\frac{j}{n}, \cdot\right)$ to Ξ is either the probability mass function or the probability density function of $a_{j,n}$ for all j and n.

(d) There exists a continuously differentiable function $f(t, a, u) : [0,1] \times [0,\alpha]^m \times [0, \bar{u}] \to [\underline{\mu}, \bar{\mu}]$ such that $\|\nabla f\|_2 \leq \bar{L}$ and $f\left(\frac{j}{n}, a, \cdot\right)$ is the conditional density of $u_{j,n}$ given $\left\{a_{j,n} = a\right\}$ for all $a \in \Xi$, j, and n.

(e) Let $\left\{(u_t, a_t)\right\}_{t\in[0,1]}$ be a collection of random variables such that $\left(u_{\frac{j}{n}}, a_{\frac{j}{n}}\right) \overset{d}{=} (u_{j,n}, a_{j,n})$ for all j and n. There exists $p^* \in \arg\min_{p\geq 0} d^T p + \int_0^1 \mathbb{E}\left[(u_s - {a_s}^T p)^+\right] ds$ such that $p^* > 0$ and $(\inf_{a\in\Xi} a^T p^*, \sup_{a\in\Xi} a^T p^*) \subseteq (\epsilon_{\mathcal{P}}, \bar{u} - \epsilon_{\mathcal{P}})$.

Assumption 3 (a) relaxes the standard i.i.d. assumptions in the literature to allow us consider nonstationary rewards and resource consumptions. We emphasize that the joint distribution of the reward and the resource consumption of the jth order may change with the total number of orders n under Assumption 3 (a). Assumption 3 (b), (c), (d) provide boundedness assumptions on the nonstationarity by requiring the support of $(u_{j,n}, a_{j,n})$ to be a subset of $[0, \bar{u}] \times [0, \alpha]^m$ for all j and n under all $P \in \mathcal{P}$, and they provide smoothness assumptions on the nonstationarity by restricting the change of distributions over the planning horizon through the requirements on the differentiability of the function v and f. The upper and lower bounds on the eigenvalue and probability densities in Assumption 3(b) and (c) together with Assumption 3(e) provide the non-degeneracy assumptions required by our regret analysis.

Compared with the assumptions made in the literature (ex. Jiang et al. [2025] and Ghuge et al. [2025]), the most important additional structural assumptions made in our paper are the smoothness assumptions on the nonstationarity. From modeling perspective, we believe these smoothness assumptions reflect the fact that, in many real-world applications, the distribution of order reward and resource consumption does not always remain identical though does not change too rapidly as well. From theory perspective, these smoothness assumptions allow us to derive a much better $O((\log n)^2)$ regret upper bound. Without these smoothness assumptions, the theory of Jiang et al. [2025] only provides a $O(T)$ regret upper bound because the deviation budget as defined in Jiang et al. [2025] of our single-sample setting is $O(T)$, and the theory of Ghuge et al. [2025] only provides a $O(n^{5/6})$ regret upper bound because Assumption 1 implies the total resource endowment is $\Theta(n)$. In fact, we can relax the smoothness assumption to make our polylogarithmic regret upper bound hold for more general distribution classes such as the one given in Example 1. A more general smoothness assumption is summarized in Assumption 4 in the appendix.

3 Regret Analysis

We will do regret analysis by fixing a sequence of distributions $P \in \mathcal{P}$, and the number of orders is included in the notation of the rewards and the resource consumptions. Thus, to simplify the notation, we will use $\mathbb{E}[\cdot]$ and $P\{\cdot\}$ to denote expectation and probability instead of $\mathbb{E}_{P_n}[\cdot]$ and $P_n\{\cdot\}$.

As mentioned in Section 2.2, the performance of Algorithm 1 largely depends on how well it can learn the price function $p_{j,n}(\cdot)$, and we can assess this by analyzing the difference between $p_{j,n}(\cdot)$ and $p^*_{j,n}(\cdot)$. The most straightforward way to measure the difference between these two functions is $\sup_{d\in\mathbb{R}^m_+}\left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2$. Unfortunately, it is technically difficult to analyze the difference if we take the supremum over the whole $\mathbb{R}^m_+$. However, a weaker result in which we take the supremum over a subset $\Omega_{j,n}$ of $\mathbb{R}^m_+$ is sufficient for the regret analysis. Recall $\{d^{off}_{j,n}\}_{j=0}^{n-1}$ defined in Section 2.2 is the sequence of average remaining resources following the optimal solutions of the offline LP. In addition, recall $b_{j,n}$ is the remaining resource after making the decision on order j when we implement Algorithm 1, and let $d_{j,n} = b_{j,n}/(n-j)$ be the average remaining resources of Algorithm 1. If Algorithm 1 has a low regret, $\{d_{j,n}\}_{j=0}^{n-1}$ should closely track $\{d^{off}_{j,n}\}_{j=0}^{n-1}$. Thus, we can just take $\Omega_{j,n}$ to be a neighborhood of $d^{off}_{j,n}$. However, $d^{off}_{j,n}$ is difficult to characterize, we instead consider a deterministic approximation it. Define

$$\delta_{j,n}(d) = \frac{nd - \sum_{k=1}^{j}\mathbb{E}\left[a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p^*_{0,n}(d)\}\right]}{n-j}.$$

We call $\{\delta_{j,n}(d)\}_{j=0}^{n-1}$ a δ-path, and we select $\Omega_{j,n}$ to be a neighborhood of $\delta_{j,n}(d_0)$. The question remained is how large $\Omega_{j,n}$ should be. Recall $\{x^{off}_{j,n}\}_{j=1}^{n}$ to be the optimal solutions of the offline LP. An interesting characteristic of the offline LP is that $p_{j,n}(d^{off}_{j,n})$ remains the same for all $0 \le j < n$. Thus, if Algorithm 1 has a low regret, the sequence of prices $\{p^\pi_{j,t}\}_{j=1}^{n}$ used by it should be relatively stationary. As $p^\pi_{j,t}$ is random and we prefer $\Omega_{j,n}$ to be a deterministic set in the regret analysis, we select $\Omega_{j,n}$ such that the set of population prices $\{p^*_{j,n}(d) : d \in \Omega_{j,n}\}$ remains the same for all j. Define

$$\bar{\delta}_{j,n}(p) = \frac{1}{n-j}\sum_{k=j+1}^{n}\mathbb{E}\left[a_{k,n}\mathbb{1}\left\{u_{k,n} > {a_{k,n}}^T p\right\}\right]$$

There are some interesting properties of $\bar{\delta}_{j,n}(\cdot)$. First, if $\bar{\delta}_{j,n}(\cdot)$ is invertible on a small enough neighborhood Ω_p of $p^*_{0,n}(d_0)$. Secondly, for each d in the range $\bar{\Omega}_{j,n}$ of $\bar{\delta}_{j,n}$ on Ω_p, $\bar{\delta}^{-1}_{j,n}(d) = p^*_{j,n}(d)$. Finally, $\delta_{j,n}(d_0)$ is in the range $\bar{\Omega}_{j,n}$. A formal statement of these properties is given in Lemma 4 in appendix C.1, and these properties allows us to define

$$\Omega_{j,n} = \left\{\bar{\delta}_{j,n}(p) : p \in B_{\frac{\epsilon}{2}}(p^*_{0,n}(d_0))\right\}.$$

where $B_{\frac{\epsilon}{2}}(p^*_{0,n}(d_0))$ is an open ball centered at $p^*_{0,n}(d_0)$ with radius $\frac{\epsilon}{2}$, and ϵ is some universally fixed small enough positive real number. Then, we have the following uniform dual convergence result.

Theorem 1. *Under Assumption 1, 3, there exists a constant C_{Dual} such that for all $P \in \mathcal{P}$, when n is large enough,*

$$\mathbb{E}\left[\sup_{d\in\Omega_{j,n}}\left\|p_{j,n}(d)-p^*_{j,n}(d)\right\|_2^2\right] \le \frac{C_{Dual}\log(n-j)}{n-j}$$

The complete and rigorous analysis on dual convergence and the proof of Theorem 1 are in appendix C.

Under Assumption 2, the uniform dual convergence result in Theorem 1 also holds for the prices $p^{\pi}_{j,n}$ used in Algorithm 1. However, the uniform dual convergence can be used only when $d_{j,n} \equiv b_{j,n}/(n-j)$ is in $\Omega_{j,n}$. Recall $b_{j,n}$ is the remaining resource after making the decision on order j when we implement Algorithm 1. When $d_{j,n}$ is not in $\Omega_{j,n}$ or we run out of resources, we have very little control over the quality of the decisions made by Algorithm 1. To prove the regret upper bound, we need to analyze how many times we cannot apply the uniform dual convergence result when we implement Algorithm 1. Let $b_{j,n}(i)$ be the ith entry of $b_{j,n}$. Define

$$\bar{\tau}_n = \inf\{j \ge 1 : \exists i \text{ s.t. } b_{j,n}(i) < \alpha \text{ or } d_{j,n} \notin \Omega_{j,n}\} \wedge n$$

A good news is that our construction of $\Omega_{j,n}$ makes the first time that $d_{j,n}$ leaves $\Omega_{j,n}$ or we run out of resources very close to the end of the planning horizon. The following theorem provides a characterization of $\bar{\tau}_n$ that reflects this fact.

Theorem 2. *Under Assumptions 1, 2, 3, there exists a constant C_{Res} such that for all $P \in \mathcal{P}$, when n is large enough,*

$$\mathbb{E}[n-\bar{\tau}_n] \le C_{Res}\log n.$$

The proof of Theorem 2 is the most difficult part of our regret analysis. When the underlying distributions are i.i.d., $\Omega_{j,n}$'s are all the same across j and n. In addition, the process $\{d_{j,n}\}_{j=0}^{n-1}$ is almost a martingale in the i.i.d. case. Thus, the standard concentration properties of martingales can be directly used to analyze the stopping time $\bar{\tau}_n$. However, when the underlying distributions become nonstationary, $\Omega_{j,n}$ changes over the planning horizon, and $\{d_{j,n}\}_{j=0}^{n-1}$ becomes far from a martingale. To prove Theorem 2, we decompose $\{d_{j,n}\}_{j=0}^{n-1}$ into a martingale component and a non-martingale component. We then upper the non-martingale component and derive a variation budget condition for the non-martingale component to be small enough. The smoothness assumptions on the nonstationarity together with Theorem 1 prove the variation budget condition holds. Then, the concentration property of the martingale component implies Theorem 2 to hold. Full details of the proof can be found in appendix D.

Theorem 1 and Theorem 2 help us deal with most of the challenging parts of the regret analysis of nonstationary OLP. In the following, we use the idea of myopic regret (Bray [2024]) to decompose the regret as in Proposition 1, and then we can apply the two theorems derived above to prove the polylogarithmic regret upper bound. Define $\Delta_{j,n}(p) = u_{j,n} - {a_{j,n}}^T p$ to be the net value of order j under price p.

Proposition 1. *Under Assumption 3, for all $P \in \mathcal{P}$ and n,*

$$\begin{aligned}
&Regret_n(\pi, P)\\
\le&\mathbb{E}\left[\sum_{j=1}^{\bar{\tau}_n-1}\Delta_{j,n}(p_{j,n}(d_{j,n}))\left(\mathbb{1}\left\{u_{j,n} > {a_{j,n}}^T p_{j,n}(d_{j,n})\right\} - \mathbb{1}\{u_{j,n} > {a_{j,n}}^T p^{\pi}_{j-1,n}\}\right)\right]\\
&+\left(\bar{u}+\frac{m\alpha\bar{u}}{\underline{d}}\right)\mathbb{E}[n-\tau_n+1]+2\bar{u}
\end{aligned}$$

The following Lemma 1 is a direct implication of Theorem 1 and the smoothness assumptions in Assumption 3. Lemma 1 implies the first term in the regret decomposition is $O((\log n)^2)$.

Lemma 1. *Under Assumptions 1, 2, 3, there exists a constant C_{Myopic} and $\tilde{\iota}$ such that for all $P \in \mathcal{P}$, when n is large enough,*

$$\begin{aligned}
&\mathbb{E}\left[\Delta_{j,n}(p_{j,n}(d_{j,n}))\left(\mathbb{1}\left\{u_{j,n} > {a_{j,n}}^T p_{j,n}(d_{j,n})\right\} - \mathbb{1}\{u_{j,n} > {a_{j,n}}^T p^{\pi}_{j-1,n}\}\right)\mathbb{1}\{j < \bar{\tau}_n\}\right]\\
\le&\frac{C_{Myopic}\log n}{n-j}
\end{aligned}$$

for all $1 \leq j < n - \tilde{\iota}$.

The proof of Lemma 1 is given in appendix F. In addition, we can directly apply Theorem 2 to upper bound the rest of the terms in the regret decomposition. Then, we get the following main theorem of our paper.

Theorem 3. *Under Assumptions 1, 2, 3, there exists a constant C such that for all $P \in \mathcal{P}$, when n is large enough,*

$$Regret_n(\pi, P) \leq C(\log n)^2$$

4 Numerical Experiments

We run numerical experiments to validate our $O((\log n)^2)$ regret upper bound in this section. We test Algorithm 1 in the following setting. We consider a sequence of independent random pairs $\{(u_{j,n}, a_{j,n})\}_{j=1}^n$. For each j and n, $u_{j,n} \sim Beta(1 + \frac{5(j-1)}{n-1}, 6 - \frac{5(j-1)}{n-1})$ and $a_{j,n}$ is a random vector with dimension m and the ith component $a_{j,n}(i) \sim Beta(3 + \frac{j-1}{n-1}, 4 - \frac{j-1}{n-1})$. $u_{j,n}$ and each component of $a_{j,n}$ are mutually independent. The initial average resource $d_0 = 0.25$. We run 10000 trials for each combination of $m \in \{1, 2, 3, 4\}$ and $n \in \{100, 200, 500, 1000, 2000, 5000, 10000\}$. The average regret of each combination is summarized in Table 1 and Figure 1(a). These numerical results justify our $O((\log n)^2)$ regret upper bound. In addition, for the single resource case ($m = 1$), the numerical results suggest that it is possible to improve the regret upper bound to $O(\log n)$.

We also test the piecewise stationary case in Example 1. $\{u_{j,n}\}_{j=1}^n$ is a sequence of independent random variables such that $u_{j,n} \sim Uniform(0, 1)$ if $j \leq \lceil \frac{n}{2} \rceil$, and $u_{j,n} \sim Uniform(2, 3)$ otherwise. For each j and n, $a_{j,n} = 1$ almost surely. Although it does not satisfy the smoothness assumptions in Assumption 3, it satisfies the more general smoothness requirements summarized in Assumption 4 in appendix. We run 10000 trials for each $n \in \{100, 200, 500, 1000, 2000, 5000, 10000\}$, and the average regret is summarized in Table 2 and Figure 1(b). We can see $O((\log n)^2)$ regret upper holds for this piecewise stationary case. This is also a single resource case, and the numerical results again suggest that it is possible to improve the regret upper bound to $O(\log n)$.

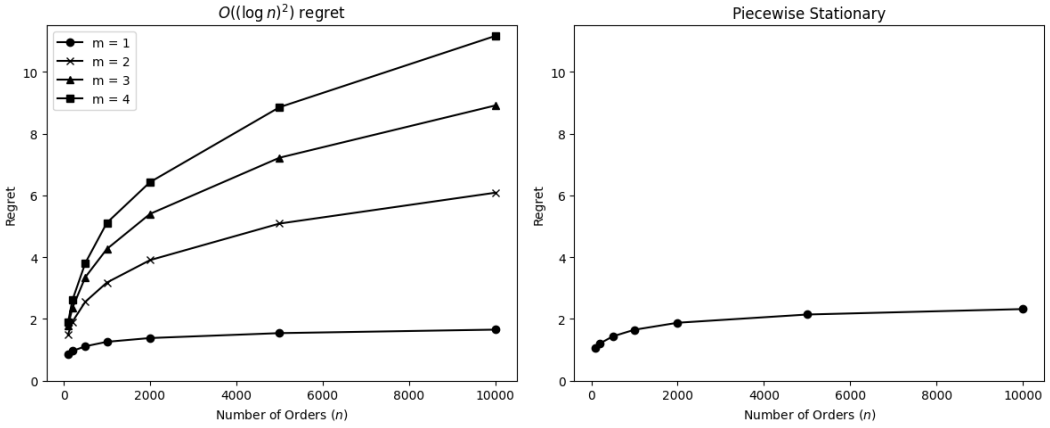


Figure 1: Numerical Results

Table 1: $O((\log n)^2)$ Regret (Entries represent Average $\pm$ 95% CI)

	n						
m	100	200	500	1000	2000	5000	10000
1	0.85±0.01	0.96±0.01	1.11±0.01	1.26±0.01	1.38±0.01	1.54±0.02	1.65±0.01
2	1.50±0.02	1.91±0.02	2.56±0.02	3.18±0.03	3.91±0.03	5.09±0.04	6.09±0.05
3	1.77±0.01	2.35±0.02	3.35±0.02	4.27±0.03	5.40±0.03	7.22±0.05	8.91±0.05
4	1.89±0.02	2.61±0.02	3.81±0.02	5.11±0.03	6.43±0.03	8.86±0.05	11.17±0.07

Table 2: Piecewise Stationary (Entries represent Average $\pm$ 95% CI)

n	100	200	500	1000	2000	5000	10000
Result	1.06$\pm$0.02	1.21$\pm$0.02	1.44$\pm$0.02	1.65$\pm$0.02	1.88$\pm$0.02	2.14$\pm$0.03	2.32$\pm$0.03

5 Conclusion

This paper investigates the nonstationary Online Linear Programming (OLP) problem, where reward-resource consumption pairs are drawn from a sequence of independent but non-identical distributions. We operate within the challenging single-sample setting, assuming the decision-maker has access to only one sample from each distribution at the start of the planning horizon. Based on a dynamic programming perspective, we propose a dual-based re-solving algorithm specifically tailored for this nonstationary environment. Our regret analysis is conducted under the regime where the resource endowment scales linearly with the total number of orders. We propose appropriate boundedness and smoothness assumptions on the nonstationarity under which we prove the proposed algorithm achieves $O((\log n)^2)$ regret. Our results demonstrate that polylogarithmic regret is attainable even under significant environmental shifts and minimal data availability, bridging the gap between stationary OLP and more volatile real-world resource allocation problems.

Our work has several limitations that also motivate future research directions. First, our polylogarithmic regret upper bound assumes access to samples drawn from the true distributions. A natural extension is to develop algorithms that use samples drawn from distributions that approximate the true distributions and to characterize how the regret depends on the approximation errors. Second, our analysis assumes that the reward and resource consumption pairs of different orders are independent. Extending the framework to accommodate dependence across orders is another important direction for future research. Third, our numerical results suggest that, at least in the single-resource case, the extra logarithmic factor in our regret upper bound may be removed, yielding an $O(\log n)$ regret bound. Such an improvement would require sharpening the dual convergence rate in Theorem 1 from $O\left(\frac{\log(n-j)}{n-j}\right)$ to $O\left(\frac{1}{n-j}\right)$. The latter rate is known to hold in the i.i.d. setting (Bray [2024]), and an immediate next step is to identify conditions under which it also holds in the nonidentically distributed setting. Finally, our current regret upper bound scales exponentially with m, whereas our numerical experiments suggest that the regret grows linearly with m. Improving the dependence of the regret upper bound on m is therefore another important direction for future research.

Funding and competing interests. This research received no external funding. The authors declare no competing interests.

References

Shipra Agrawal, Zizhuo Wang, and Yinyu Ye. A dynamic near-optimal algorithm for online linear programming. *Operations Research*, 62(4):876–890, 2014. doi: 10.1287/opre.2014.1289.

Santiago R. Balseiro, Haihao Lu, and Vahab Mirrokni. The best of many worlds: Dual mirror descent for online allocation problems. *Operations Research*, 71(1):101–119, 2023.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. In *Advances in Neural Information Processing Systems*, volume 27, 2014.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015. doi: 10.1287/opre.2015.1408.

Robert L. Bray. Logarithmic regret in multisecretary and online linear programs with continuous valuations. *Operations Research*, 73(4):2188–2203, 2024. doi: 10.1287/opre.2022.0036.

G. Chen, X. Li, and Y. Ye. An improved analysis of lp-based control for revenue management. *Operations Research*, 2022.

Wang Chi Cheung and Guodong Lyu. Online planning in nonstationary environments. *Operations Research*, 2025.

Yiding Feng, Jiashuo Jiang, and Yige Wang. Non-stationary online resource allocation: Learning from a single sample. *arXiv preprint*, arXiv:2602.18114, 2026.

Rohan Ghuge, Sahil Singla, and Yifan Wang. Single-sample and robust online resource allocation. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1442–1453, 2025.

Jiashuo Jiang, Xiaocheng Li, and Jiawei Zhang. Online stochastic optimization with wasserstein-based nonstationarity. *Management Science*, 71(11):9104–9122, 2025. doi: 10.1287/mnsc.2020.03850.

Xiaocheng Li and Yinyu Ye. Online linear programming: Dual convergence, new algorithms, and regret bounds. *Operations Research*, 70(5):2948–2966, 2022. doi: 10.1287/opre.2021.2164.

Xiaocheng Li., Chunlin Sun., and Yinyu Ye. Simple and fast algorithm for binary integer and online linear programming. *Mathematical Programming*, 200(2):831–875, 2023.

Shang Liu, Jiashuo Jiang, and Xiaocheng Li. Non-stationary bandits with knapsacks. In *Advances in Neural Information Processing Systems*, volume 35, pages 16522–16532, 2022.

Thodoris Lykouris, Max Simchowitz, Aleksandrs Slivkins, and Wen Sun. Corruption-robust exploration in episodic reinforcement learning. *Mathematics of Operations Research*, 50(2):1277–1304, 2025. doi: 10.1287/opre.2015.1408.

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

Xilin Zhang and Wang Chi Cheung. Piecewise-stationary bandits with knapsacks. In *Advances in Neural Information Processing Systems*, volume 37, pages 56001–56040, 2024.

Xilin Zhang and Wang Chi Cheung. Online allocation of reusable resources in nonstationary environments. *Mathematics of Operations Research*, 2025.

Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Adaptivity and non-stationarity: Problem-dependent dynamic regret for online convex optimization. *Journal of Machine Learning Research*, 25(98):1–52, 2024.

A Proof of Example 1

Proof. Denote $b^\pi(n)$ be the amount of resource left after making decisions on the order $\lceil \frac{n}{2} \rceil$. Take $\delta = \frac{1}{128}$ and $\epsilon = \frac{1}{32}$. Define $\mathcal{E}_n = \{\frac{(1-\epsilon)n}{4} \le b^\pi(n) \le \frac{n}{4}\}$, and denote $u_{(j),n}$ as the jth largest value among $\{u_{j,n}\}_{j=1}^n$ and $\tilde{u}_{(j),n}$ as the jth largest value among $\{u_{j,n}\}_{j=\lceil \frac{n}{2} \rceil+1}^n$. Since the distribution of the rewards is completely unknown, $\mathbb{E}_{P_{1,n}}[\mathbb{1}_{\mathcal{E}_n}] = \mathbb{E}_{P_{2,n}}[\mathbb{1}_{\mathcal{E}_n}]$. If $\mathbb{E}_{P_{1,n}}[\mathbb{1}_{\mathcal{E}_n}] \ge 1 - \delta$ then on event $\mathcal{E}_n$,

$$R_n(\pi) \le \frac{\epsilon}{4} n + \sum_{j=1}^{\lfloor \frac{n}{4} \rfloor} \tilde{u}_{(j),n}$$

Then, by the fact that the expectation of the sum of the k largest values among n i.i.d. $Uniform(0,1)$ random variables is $\frac{k(2n-k+1)}{2(n+1)}$,

$$\begin{aligned}
\mathbb{E}_{P_{1,n}}\left[R_n(\pi)\right] =& \mathbb{E}_{P_{1,n}}\left[R_n(\pi)\mathbb{1}_{\mathcal{E}_n^c}\right] + \mathbb{E}_{P_{1,n}}\left[R_n(\pi)\mathbb{1}_{\mathcal{E}_n}\right] \\
\le& \delta n + \frac{\epsilon}{4} n + \mathbb{E}_{P_{1,n}}\left[\sum_{j=1}^{\lfloor \frac{n}{4} \rfloor} \tilde{u}_{(j),n}\right] \\
\le& \delta n + \frac{\epsilon}{4} n + \frac{\lfloor \frac{n}{4} \rfloor \left(n - \lfloor \frac{n}{4} \rfloor + 1\right)}{n+1}
\end{aligned}$$

In addition,

$$\mathbb{E}_{P_{1,n}}[R_n^*] \geq \mathbb{E}_{P_{1,n}}\left[\sum_{j=1}^{\lfloor \frac{n}{4} \rfloor} u_{(j),n}\right] = \frac{\lfloor \frac{n}{4} \rfloor \left(2n - \lfloor \frac{n}{4} \rfloor + 1\right)}{2\,(n+1)}$$

Then,

$$Regret_n(\pi, P_{1,n}) \geq \frac{\lfloor \frac{n}{4} \rfloor \left(\lfloor \frac{n}{4} \rfloor - 1\right)}{2(n+1)} - \left(\delta + \frac{\epsilon}{4}\right) n \geq \frac{1}{64} n - \frac{3}{8} \tag{4}$$

If $\mathbb{E}_{P_{2,n}}[\mathbb{1}_{\mathcal{E}_n^c}] \geq \delta$, then on event $\mathcal{E}_n^c$,

$$R_n(\pi) \leq \frac{n}{4} - b^\pi(n) + \sum_{j=1}^{b^\pi(n)} \tilde{u}_{(j),n}$$

Let $\mathcal{F}_{\lceil \frac{n}{2} \rceil}$ be the sigma algebra generated by $\{u_{j,n}\}_{j=1}^{\lceil \frac{n}{2} \rceil}$.

$$\begin{aligned}
&\mathbb{E}_{P_{2,n}}\left[R_n(\pi)\mathbb{1}_{\mathcal{E}_n^c}\right] \\
&\leq \mathbb{E}_{P_{2,n}}\left[\left(\frac{n}{4} - b^\pi(n)\right)\mathbb{1}_{\mathcal{E}_n^c}\right] + \mathbb{E}_{P_{2,n}}\left[\mathbb{E}_{P_{2,n}}\left[\sum_{j=1}^{b^\pi(n)} \tilde{u}_{(j),n} \middle| \mathcal{F}_{\lceil \frac{n}{2} \rceil}\right]\mathbb{1}_{\mathcal{E}_n^c}\right] \\
&\leq \mathbb{E}_{P_{2,n}}\left[\left(\frac{n}{4} - b^\pi(n)\right)\mathbb{1}_{\mathcal{E}_n^c}\right] + \mathbb{E}_{P_{2,n}}\left[\left(\frac{\lfloor \frac{n}{4} \rfloor \left(2\lfloor \frac{n}{2} \rfloor - \lfloor \frac{n}{4} \rfloor + 1\right)}{2\left(\lfloor \frac{n}{2} \rfloor + 1\right)} + 2b^\pi(n)\right)\mathbb{1}_{\mathcal{E}_n^c}\right]
\end{aligned}$$

In addition,

$$\mathbb{E}_{P_{2,n}}[R_n^*\mathbb{1}_{\mathcal{E}^c}] \geq \mathbb{E}_{P_{2,n}}\left[\sum_{j=1}^{\lfloor \frac{n}{4} \rfloor} \tilde{u}_{(j),n}\mathbb{1}_{\mathcal{E}_n^c}\right] = \mathbb{E}_{P_{2,n}}\left[\left(\frac{\lfloor \frac{n}{4} \rfloor \left(2\lfloor \frac{n}{2} \rfloor - \lfloor \frac{n}{4} \rfloor + 1\right)}{2\left(\lfloor \frac{n}{2} \rfloor + 1\right)} + \left\lfloor \frac{n}{2} \right\rfloor\right)\mathbb{1}_{\mathcal{E}_n^c}\right]$$

Thus,

$$Regret(\pi, P_{2,n}) \geq \mathbb{E}_{P_{2,n}}\left[\left(\left\lfloor \frac{n}{2} \right\rfloor - \frac{n}{4} - b^\pi(n)\right)\mathbb{1}_{\mathcal{E}_n^c}\right] \geq \frac{\epsilon\delta}{4} n - 1 \tag{5}$$

(4) and (5) together imply that the $Regret_n(\pi)$ is $\Omega(n)$. □

B A More General Condition on Nonstationarity

In this section, we derive a more general condition on nonstationarity (Proposition 2) under which the $O((\log n)^2)$ regret upper bound holds. We start by introducing several definitions and lemmas that are used throughout the analysis. Note that we do regret analysis by fixing a sequence of distributions $P \in \mathcal{P}$, and the number of orders is included in the notation of the rewards and the resource consumptions. Thus, to simplify the notation, we will use $\mathbb{E}[\cdot]$ and $P\{\cdot\}$ to denote expectation and probability instead of $\mathbb{E}_{P_n}[\cdot]$ and $P_n\{\cdot\}$. Define

$$g_{j,n}^*(p, d) = d^T p + \frac{1}{n-j}\sum_{k=j+1}^{n} \mathbb{E}\left[\left(u_{k,n} - {a_{k,n}}^T p\right)^+\right]$$

and for a fixed d, define

$$p_{j,n}^*(d) \in \arg\min_{p \geq 0} g_{j,n}^*(p, d).$$

For simplicity, we write $g_{0,n}^*(p, d)$ as $g_n^*(p, d)$ and $p_{0,n}^*(d)$ as $p_n^*(d)$. As $g_{j,n}^*(\cdot, \cdot)$'s are functions only depend on the distributions of the rewards and the resource consumptions, we call the optimization problems defining $p_{j,n}^*(\cdot)$'s the *population problems* and call $p_{j,n}^*(\cdot)$'s the *population prices*. For now, we denote $p_{j,n}^*(d)$ as an arbitrary element in the optimal solution set and provide a boundedness result in the following Lemma. A uniqueness result will be given later. Let d^{min} be the smallest entry of the vector d.

Lemma 2. *Under Assumption 3, for any $P \in \mathcal{P}$, if $d^{min} > 0$, then $\left\|p^*_{j,n}(d)\right\|_1 \leq \frac{\bar{u}}{d^{min}}$ for all j and n.*

Proof. Since $p^*_{j,n}(d) \geq 0$,

$$d^{min} e^T p^*_{j,n}(d) \leq d^T p^*_{j,n}(d) \leq g^*_{j,n}(p^*_{j,n}(d), d) \leq g^*_{j,n}(0, d) \leq \bar{u}$$

□

Assumption 3 also guarantees some convergence properties of the population problems as n goes to infinity. Let $\{(u_t, a_t)\}_{t\in[0,1]}$ be a collection of random variables such that $\left(u_{\frac{j}{n}}, a_{\frac{j}{n}}\right) \overset{d}{=} (u_{j,n}, a_{j,n})$ for all j and n. Define

$$g^*_\infty(p, d) = d^T p + \int_0^1 \mathbb{E}\left[(u_s - {a_s}^T p)^+\right] ds$$

The following lemma shows a pointwise convergence property of the objective functions of the population problems defining $p^*_n(d)$.

Lemma 3. *Under Assumption 3, for any $P \in \mathcal{P}$, $\lim_{n\to+\infty} g^*_n(p, d_0) = g^*_\infty(p, d_0)$ for all $p \in \mathbb{R}^m$.*

Proof. Fix $p \in \mathbb{R}^m$, $\frac{1}{n}\sum_{j=1}^n \mathbb{E}\left[\left(u_{j,n} - {a_{j,n}}^T p\right)^+\right]$ is a Riemann sum of $\int_0^1 \mathbb{E}\left[(u_s - {a_s}^T p)^+\right]$. We assume the following that $a_{j,n}$'s have probability density function. For the case that $a_{j,n}$'s have probability mass function. In addition, for any $t_1, t_2 \in [0, 1]$,

$$\begin{aligned}
&\left|\mathbb{E}\left[\left(u_{t_1} - {a_{t_1}}^T p\right)^+\right] - \mathbb{E}\left[\left(u_{t_2} - {a_{t_2}}^T p\right)^+\right]\right| \\
\leq &\int_{\Xi_a}\int_0^{\bar{u}} (u - a^T p)^+ \left|f(t_1, u, a)v(t_1, a) - f(t_2, u, a)v(t_2, a)\right| du\, da \\
\leq &2\bar{\mu}\bar{L}|t_1 - t_2| \int_{\Xi_a}\int_0^{\bar{u}} (u - a^T p)^+ du\, da \\
\leq &2\bar{u}\bar{\mu}\bar{L}\alpha^m(\bar{u} + \alpha p)|t_1 - t_2|
\end{aligned}$$

where the Lipschitz continuity used above is given by Assumption 3 (c) and (d). Thus, $\mathbb{E}\left[(u_s - {a_s}^T p)^+\right]$ is Lipschitz continuous on $[0, 1]$ with respect to t and hence is Riemann integrable. Then,

$$\lim_{T\to+\infty} \frac{1}{T}\sum_{j=1}^T \mathbb{E}\left[\left(u_{j,T} - {a_j}^T p\right)^+\right] = \int_0^1 \mathbb{E}\left[(u_s - {a_s}^T p)^+\right] dt$$

which completes the proof. □

Thus, we can view $g^*_\infty(\cdot, \cdot)$ as the objective function of the limiting population problem. We are now ready to give the more general condition on nonstationarity.

Proposition 2. *Under Assumptions 1 and 3, there exists positive real numbers $\kappa < 1$, ϵ, ϵ_p, $\underline{\sigma}$, $\bar{\sigma}$, L such that for all $P \in \mathcal{P}$.*

- *(a) $\mathbb{E}\left[\left(u_{j,n} - {a_{j,n}}^T p\right)^+\right]$ is three times continuously differentiable and the absolute value of its third derivatives are bounded above by L on $B_{2\epsilon}(p^*) \equiv \Omega_p \subseteq \mathbb{R}^m_+$ for all j and n.*
- *(b) On Ω_p, the eigenvalues of the Hessian of $\mathbb{E}\left[\left(u_{j,n} - {a_{j,n}}^T p\right)^+\right]$ are in $[0, \bar{\sigma}]$ and the eigenvalues of the Hessian of $g^*_{j,n}(p, d_0)$ are in $[\underline{\sigma}, \bar{\sigma}]$ for all j and n.*
- *(c) There exists N_0 such that, when $n \geq N_0$, $p^*_n(d_0)$ is uniquely determined, $B_\epsilon(p^*_n(d_0)) \subseteq \Omega_p$, and each entry of $\mathbb{E}[a_{j,n}\mathbb{1}\{u_{j,n} > {a_{j,n}}^T p^*_n(d_0)\}]$ is greater than ϵ_p if $j \geq \kappa n$.*

(d) Let $\underline{s} = \inf_{a\in\Xi, p\in\Omega_p} a^T p$, $\bar{s} = \sup_{a\in\Xi, p\in\Omega_p} a^T p$, *and*

$$L_{j,n} = \frac{\bar{\mu}\alpha^{m+2}}{\underline{\sigma}(n-j)(n-j-1)} \sum_{k=j+2}^{n} \left(\sup_{a\in\Xi} \sup_{s\in[\underline{s},\bar{s}]} \left| f\left(\frac{k}{n}, a, s\right) - f\left(\frac{j+1}{n}, a, s\right) \right| + \sup_{a\in\Xi} \left| v\left(\frac{k}{n}, a\right) - v\left(\frac{j+1}{n}, a\right) \right| \right)$$

then for all $n \in \mathbb{N}_+$,

$$\prod_{j=0}^{n-2} \left(1 + L_{j,n}\right) \leq e^{\frac{\mu \bar{L} \alpha^{m+2}}{\underline{\sigma}}}$$

Proof. Let $\{e_i\}_{i=1}^m$ be the standard basis of $\mathbb{R}^m$. For part (a), for any $p \in \mathbb{R}^m$, by Assumption 3(d), $\mathbb{P}\left\{u_{j,n} = {a_{j,n}}^T p\right\} = 0$, then

$$\lim_{h\to 0} \frac{(u_{j,n} - {a_{j,n}}^T(p + he_i))^+ - (u_{j,n} - {a_{j,n}}^T p)^+}{h} = -{e_i}^T a_{j,n} \mathbb{1}\{u_{j,n} > {a_{j,n}}^T p\} \quad \text{a.s.}$$

and

$$\left| \frac{(u_{j,n} - {a_{j,n}}^T(p + he_i))^+ - (u_{j,n} - {a_{j,n}}^T p)^+}{h} \right| \leq \alpha \quad \forall h$$

By Bounded Convergence Theorem, $\mathbb{E}\left[\left(u_{j,n} - {a_{j,n}}^T p\right)^+\right]$ is differentiable on $\mathbb{R}^m$ and

$$\nabla \mathbb{E}\left[\left(u_{j,n} - {a_{j,n}}^T p\right)^+\right] = -\mathbb{E}\left[a_{j,n} \mathbb{1}\{u_{j,n} > {a_{j,n}}^T p\}\right]$$

By Assumption 3(e), there exists ϵ small enough such that for any $P \in \mathcal{P}$, $a^T p \in (0, \bar{u})$ if $p \in B_{2\epsilon}(p^*) \equiv \Omega_p \subseteq \mathbb{R}_m^+$ and $a \in \Xi$. Then, for $p \in \Omega_p$

$$\begin{aligned}
&\lim_{h\to 0} \frac{\mathbb{E}\left[{e_i}^T a_{j,n} \mathbb{1}\{u_{j,n} > {a_{j,n}}^T p\} - {e_i}^T a_{j,n} \mathbb{1}\{u_{j,n} > {a_{j,n}}^T (p + he_k)\}\right]}{h} \\
=&\lim_{h\to 0} \mathbb{E}\left[\frac{{e_i}^T a_{j,n} \left(P(u_{j,n} > {a_{j,n}}^T p | a_{j,n}) - P(u_{j,n} > {a_{j,n}}^T (p + he_k)) | a_{j,n})\right)}{h}\right] \\
=&\mathbb{E}\left[({e_i}^T a_{j,n})({e_k}^T a_{j,n}) f\left(\frac{j}{n}, {a_{j,n}}^T p, a_{j,n}\right)\right]
\end{aligned}$$

where the last line is by Assumption 3(d) and Bounded Convergence Theorem. Then, $\mathbb{E}\left[\left(u_{j,n} - {a_{j,n}}^T p\right)^+\right]$ is twice differentiable on Ω_p and its Hessian is

$$\mathbb{E}\left[(a_{j,n} {a_{j,n}}^T) f\left(\frac{j}{n}, {a_{j,n}}^T p, a_{j,n}\right)\right]$$

Similarly,

$$\begin{aligned}
&\lim_{h\to 0} \frac{1}{h} \left(\mathbb{E}\left[({e_i}^T a_{j,n})(e_k^T a_{j,n}) f\left(\frac{j}{n}, {a_{j,n}}^T (p + he_q), a_{j,n}\right)\right] \right. \\
&\qquad \left. - \mathbb{E}\left[({e_i}^T a_{j,n})({e_k}^T a_{j,n}) f\left(\frac{j}{n}, {a_{j,n}}^T p, a_{j,n}\right)\right] \right) \\
=&\mathbb{E}\left[({e_i}^T a_{j,n})({e_k}^T a_{j,n})({e_q}^T a_{j,n}) \frac{\partial f}{\partial u}\left(\frac{j}{n}, {a_{j,n}}^T p, a_{j,n}\right)\right]
\end{aligned}$$

Thus, $\mathbb{E}\left[\left(u_{j,n} - {a_{j,n}}^T p\right)^+\right]$ is three times continuously differentiable on Ω_p. In addition, Assumption 3 implies that

$$\left|\mathbb{E}\left[({e_i}^T a_{j,n})({e_k}^T a_{j,n})({e_q}^T a_{j,n})\frac{\partial f}{\partial u}\left(\frac{j}{n}, {a_{j,n}}^T p, a_{j,n}\right)\right]\right| \leq \bar{L}\alpha^3 \equiv L$$

For part (b), for each j and n, the largest eigenvalue of the Hessian of $\mathbb{E}[(u_{j,n} - {a_{j,n}}^T p)^+]$ is upper bounded by its trace, which is upper bounded by $m\alpha\bar{\mu} \equiv \bar{\sigma}$ by Assumption 3. For the smallest eigenvalue of the Hessian of $\mathbb{E}[(u_{j,n} - {a_{j,n}}^T p)^+]$, note that

$$\begin{aligned}
&\mathbb{E}\left[(a_{j,n}{a_{j,n}}^T) f\left(\frac{j}{n}, {a_{j,n}}^T p, a_{j,n}\right)\right] \\
=&\mathbb{E}\left[(a_{j,n}{a_{j,n}}^T)\underline{\mu}\right] + \mathbb{E}\left[(a_{j,n}{a_{j,n}}^T)\left(f\left(\frac{j}{n}, {a_{j,n}}^T p, a_{j,n}\right) - \underline{\mu}\right)\right]
\end{aligned}$$

Both of the two matrices on the right-hand-side are symmetric and positive semi-definite. Then, Assumption 3 implies that the smallest eigenvalue of $\mathbb{E}\left[a_{j,n}{a_{j,n}}^T\underline{\mu}\right]$ is greater than $\underline{\mu}\lambda_{min} \equiv \underline{\sigma} > 0$. Then, together with the Weyl's Inequality, the smallest eigenvalue of the Hessian of the Hessian of $\mathbb{E}[(u_{j,n} - {a_{j,n}}^T p)^+]$ is lower bounded by $\underline{\sigma}$. This immediately shows that the eigenvalues of the Hessian of $g^*_{j,n}(p, d_0)$ are in $[\underline{\sigma}, \bar{\sigma}]$.

For part (c), because $(u_{j,n} - {a_{j,n}}^T p)^+$ is convex in p almost surely for all $j$and n, $g^*_n(\cdot, d_0)$ and $g^*_\infty(\cdot, d_0)$ are all convex. Together with the pointwise convergence given by Lemma 3, $g^*_T(\cdot, d_0)$ converges to $g^*_\infty(\cdot, d_0)$ uniformly over $\left[0, \frac{2\bar{u}}{d_0^{min}}\right]^m$, where d_0^{min} is defined in Lemma 2. Similar arguments in the proof of Lemma 2 imply that $p^* \in \left[0, \frac{\bar{u}}{d_0^{min}}\right]^m$ for all $P \in \mathcal{P}$. Together with Assumption 3(e), we can select ϵ small enough such that Ω_p is in the interior of $\left[0, \frac{2\bar{u}}{d_0^{min}}\right]^m$ for all $P \in \mathcal{P}$. Following similar arguments in part (a), the gradient and the Hessian of $g^*_\infty(p, d_0)$ on Ω_p are $d_0 - \mathbb{E}\left[\int_0^1 a_s \mathbb{1}\{u_s > {a_s}^T p\} ds\right]$ and $\mathbb{E}\left[\int_0^1 (a_s {a_s}^T) f(s, {a_s}^T p, a_s) ds\right]$. In addition, the Hessian is positive definite on Ω_p with smallest eigenvalue greater than $\underline{\sigma}$. Thus, p^* is the unique optimal solution. Suppose that there exists $\nu > 0$ and $\{n_k\}_{k=1}^{+\infty}$ such that $\left\|p^*_{n_k}(d_0) - p^*\right\|_2 > \nu$ for all k and all possible selection of $p^*_{n_k}(d_0)$. We can assume $\nu < \epsilon$. Then, the strong convexity of $g^*_\infty(\cdot, d_0)$ on Ω_p implies that, for any $p \in \Omega_p \backslash B_\nu(p^*)$,

$$g(p) \geq \frac{\underline{\sigma}}{2}\nu^2$$

Together with the fact that the level set of $g^*_\infty(\cdot, d_0)$ is convex, we have

$$\{p : g_\infty(p, d_0) - g_\infty(p^*, d_0) \leq \frac{\underline{\sigma}}{4}\nu^2\} \subseteq B_\nu(p^*)$$

The uniform convergence implies that when k is large enough,

$$\left|g^*_{n_k}(p, d_0) - g^*_\infty(p, d_0)\right| < \frac{\underline{\sigma}}{12}\nu^2 \quad \forall p \in \left[0, \frac{2\bar{u}}{d_0^{min}}\right]^m$$

Since $p^*_{n_k}(d_0) \in \left[0, \frac{\bar{u}}{d_0^{min}}\right]^m$ by Lemma 2 and $p^*_{T_k}(d_0) \notin B_\nu(p^*)$, when k is large enough

$$g^*_{T_k}(p^*_{n_k}(d_0), d_0) - g^*_{n_k}(p^*, d_0) \geq \frac{\underline{\sigma}}{12}\nu^2$$

which means that $p^*_{n_k}$ is not an optimal solution. This is a contradiction. Thus, there exists a selection of $p^*_n(d_0)$'s such that $p^*_n(d_0) \to p^*$ as $n \to +\infty$. For this selection, there exists $N_0 \in \mathbb{N}_+$ such that $\left\|p^*_N(d_0) - p^*\right\|_2 < \epsilon$. Then, when $n \geq N_0$, $B_\epsilon(p^*_n(d_0)) \subseteq \Omega_p$. In addition, $p^*_n(d_0) > 0$,

and the Hessian of $g_n^*(\cdot, d_0)$ at $p_n^*(d_0)$ is positive definite by part (b), which implies that $p_n^*(d_0)$ is uniquely determined. Furthermore, pick κ to be any real number in $(0, 1)$, together with Assumption 3(e), we can select ϵ small enough such that there exists $\tilde{\epsilon}_p$ which holds for all $P \in \mathcal{P}$ such that $(\inf_{a\in\Xi} a^T p_n^*(d_0), \sup_{a\in\Xi} a^T p_n^*(d_0)) \subseteq (\tilde{\epsilon}_p, \bar{u} - \tilde{\epsilon}_p)$. Then, for all j, n

$$\begin{aligned}&\mathbb{E}[a_{j,n}\mathbb{1}\{u_{j,n} > {a_{j,n}}^T p_n^*(d_0)\}]\\ \geq&\mathbb{E}[a_{j,n}P\{u_{j,n} > {a_{j,n}}^T p_n^*(d_0)|a_{j,n}\}] \geq \mathbb{E}[a_{j,n}\underline{\mu}\epsilon_p] \geq \underline{a}\underline{\mu}\tilde{\epsilon}_p e \equiv \epsilon_p e\end{aligned}$$

For part (d), By Assumption 3(c) and (d),

$$\prod_{j=0}^{n-2}(1 + L_{j,n}) \leq \prod_{j=0}^{n-2}\left(1 + \frac{2\bar{\mu}\bar{L}\alpha^{m+2}\sum_{k=j+2}^{n}\frac{k-(j+1)}{n}}{\underline{\sigma}(n-j)(n-j-1)}\right) \leq \left(1 + \frac{\bar{\mu}\bar{L}\alpha^{m+2}}{\underline{\sigma}n}\right)^n \leq e^{\frac{\mu\bar{L}\alpha^{m+2}}{\underline{\sigma}}}$$

$$\prod_{k=0}^{n-2}\left(1 + \frac{\alpha^2\sum_{j=k+2}^{n}\sup_{a\in\Xi,s\in[\underline{s},\bar{s}]}\left|f(\frac{j}{n}, s, a) - f(\frac{k+1}{n}, s, a)\right|}{\underline{\sigma}(n-k)(n-k-1)}\right)$$

$$\leq \prod_{k=0}^{n-2}\left(1 + \frac{\alpha^2\bar{L}\sum_{j=k+2}^{n}\frac{j-(k+1)}{n}}{\underline{\sigma}(n-k)(n-k-1)}\right) \leq \left(1 + \frac{\alpha^2\bar{L}}{2\underline{\sigma}n}\right)^n \leq e^{\frac{\alpha^2\bar{L}}{2\underline{\sigma}}}$$

□

Proposition 2 (a), (b) and (c) gives the smoothness and strong convexity of $g_{j,n}(p, d)$ for p in a neighborhood of $p_n^*(d_0)$, and Proposition 2 (d) can be viewed as a variation budget requirement. We can view Proposition 2 as more general assumptions to make the regret bound in this paper hold. In fact, our regret analysis holds if Assumption 1, Assumption 2, independence and boundedness of the distributions in Assumption 3 and Proposition 2 hold. Thus, although the $\mathcal{P}$ in Example 1 does not exactly satisfies Assumption 3, the regret bound in this paper is still true for this $\mathcal{P}$ because Proposition 2 holds under it. To make it rigorous, we formally summarize the more general assumption as follows. Then, replacing Assumption 3 with Assumption 4, Theorem 3 still holds.

Assumption 4. *There exist positive real numbers $\kappa < 1$, ϵ, ϵ_p, $\underline{\sigma}$, $\bar{\sigma}$, $\bar{L}$, $\bar{\mu}$, $\bar{u}$, α, and C_{VB}, and a set $\Xi \subseteq [0, \alpha]^m$ such that for all $P \in \mathcal{P}$,*

(a) $\left\{\left(u_{j,n}, a_{j,n}\right)\right\}_{j=1}^{n}$ *is a sequence of independent random pairs in $\mathbb{R}_+ \times \mathbb{R}_+^m$ for all n.*

(b) Ξ *is the support of $a_{j,n}$, and $a_{j,n}$ has either a probability mass function or a probability density function $v\left(\frac{j}{n}, \cdot\right)$ such that $\sup_{a\in\Xi} v\left(\frac{j}{n}, a\right) \leq \bar{\mu}$ for all j and n.*

(c) $0 \leq u_{j,n} \leq \bar{u}$ *almost surely, and the conditional distribution of $u_{j,n}$ given $a_{j,n} = a$ is continuous and has a probability density function $f\left(\frac{j}{n}, a, \cdot\right)$ such that $\sup_{u\in[0,\bar{u}]} f\left(\frac{j}{n}, a, u\right) \leq \bar{\mu}$ for all $a \in \Xi$, j, and n.*

(d) There exists an open ball $\Omega_p \subseteq \mathbb{R}_+^m$ with radius 2ϵ such that $\mathbb{E}\left[\left(u_{j,n} - a_{j,n}^T p\right)^+\right]$ is three times continuously differentiable, and the absolute values of its third derivatives are bounded above by $\bar{L}$ on Ω_p for all j and n.

(e) On Ω_p, the eigenvalues of the Hessian of $\mathbb{E}\left[\left(u_{j,n} - a_{j,n}^T p\right)^+\right]$ are in $[0, \bar{\sigma}]$, and the eigenvalues of the Hessian of $g_{j,n}^*(p, d_0)$ are in $[\underline{\sigma}, \bar{\sigma}]$ for all j and n.

(f) There exists N_0 such that, when $n \geq N_0$, $p_n^*(d_0)$ is uniquely determined, $B_\epsilon\left(p_n^*(d_0)\right) \subseteq \Omega_p$, and each entry of $\mathbb{E}\left[a_{j,n}\mathbb{1}\left\{u_{j,n} > a_{j,n}^T p_n^*(d_0)\right\}\right]$ is greater than ϵ_p if $j \geq \kappa n$.

(g) Let $L_{j,n}$ be defined as in Proposition 2, then $\prod_{j=0}^{n-2}(1 + L_{j,n}) \leq C_{VB}$.

C Dual Convergence

In this section, we give the complete and rigorous analysis on dual convergence. Before giving the proof of Theorem 1, we need to introduce some definitions and useful Lemmas and Propositions.

We focus on the error of using $p^{\pi}_{j,n}$ to estimate $p_{j,n}\left(\frac{b_{j,n}}{n-j}\right)$. Define

$$g_{j,n}(p,d) = d^T p + \frac{1}{n-j}\sum_{k=j+1}^{n}\left(u_{k,n} - {a_{k,n}}^T p\right)^+$$

Then, $p_{j,n}(d) \in \arg\min_{p\geq 0} g_{j,n}(p;d)$. Note that both $p^{\pi}_{j,n}$ and $p_{j,n}(d)$ may not be unique; however, the selection of these prices from the corresponding optimal solution sets does not affect the regret analysis. Since $p^{\pi}_{j,n}$ and $p_{j,n}(\frac{b_{j,n}}{n-j})$ have the same distribution, and they are both a sample average approximation of the population price $p^{*}_{j,n}\left(\frac{b_{j,n}}{n-j}\right)$, we choose to focus on the mean squared error between $p_{j,n}(\frac{b_{j,n}}{n-j})$ and $p^{*}_{j,n}\left(\frac{b_{j,n}}{n-j}\right)$. Similar to the analysis of the i.i.d case as in Bray [2024], to deal with the randomness of $b_{j,n}$, our goal in this section is to prove a uniform dual convergence result by finding an upper bound of

$$\mathbb{E}\left[\sup_{d\in\Omega_{j,n}}\left\|p_{j,n}(d) - p^{*}_{j,n}(d)\right\|_2^2\right] \tag{6}$$

An interesting question to ask is: how should we select the set $\Omega_{j,n}$? If it is selected too small, then $\frac{b_{j,n}}{n-j}$ cannot fall into it with high probability which will make an upper bound of (6) useless. If it is selected too large, (6) will be too large to lead us to a polylogarithmic regret upper bound. We answer this question by introducing the concept of δ-*path* in the following subsection.

C.1 δ-path

Define $d^{\pi}_{j,n} = b^{\pi}_{j,n}/(n-j)$. Let $\{x^{off}_{j,n}\}_{j=1}^{n}$ be the optimal solution of the offline LP (1) and define $d^{off}_{j,n} = (b_{0,n} - \sum_{k=1}^{j} a_{k,n}x^{off}_{k,n})/(n-j)$. If Algorithm 1 has low regret, it is intuitive that the two sequences $\{d^{\pi}_{j,n}\}_{j=0}^{n-1}$ and $\{d^{off}_{j,n}\}_{j=0}^{n-1}$ are close with high probability. Thus, we may want $\Omega_{j,n}$ to be a neighborhood of $d^{off}_{j,n}$. However, $d^{off}_{j,n}$ is random which makes analyze the mean squared error over its neighborhood difficult. We instead construct a deterministic approximation of $d^{off}_{j,n}$ and let $\Omega_{j,n}$ be the neighborhood of the deterministic approximation.

Recall that d_0 is the initial average resource given in Assumption 1. By complementary slackness

$$x^{off}_{j,n} = \begin{cases} 1 & \text{if } u_{j,n} > {a_{j,n}}^T p_{0,n}(d_0) \\ 0 & \text{if } u_{j,n} < {a_{j,n}}^T p_{0,n}(d_0) \end{cases}$$

Assumption 3 implies that the conditional distribution of $u_{j,n}$ given $a_{j,n}$ is continuous for all j and n. Thus, $\{u_{j,n} = {a_{j,n}}^T p_{0,n}(d_0)\}$ happens for at most m orders. Thus, when $n >> m$, $\tilde{d}^{off}_{j,n} = (b_{0,n} - \sum_{k=1}^{j} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p_{0,n}(d_0)\})/(n-j)$ is almost the same as $d^{off}_{j,n}$. We then replace $p_{0,n}(d_0)$ with the population price $p^{*}_{n}(d_0)$ and take expectation to get the following definition of $\delta_{j,n}$. Define

$$\delta_{j,n}(d) = \frac{nd - \sum_{k=1}^{j}\mathbb{E}\left[a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p^{*}_{n}(d)\}\right]}{n-j}.$$

$\{\delta_{j,n}(d_0)\}_{j=0}^{n-1}$ is then our deterministic approximation of $\{d^{off}_{j,n}\}_{j=0}^{n-1}$, and we will make $\Omega_{j,n}$ a neighborhood of $\delta_{j,n}(d_0)$. The question remained is how large $\Omega_{j,n}$ should be. We answer this question by proving some useful properties of $\{\delta_{j,n}\}_{j=1}^{n}$.

We start from a more detailed discussion on the relationship between δ-path and population price. Let Ω_p be given in Proposition 2. For each j, n and $p \in \Omega_p$, denote $G^{*}_{j,n}(p)$ as the Hessian of $g^{*}_{j,n}(\cdot,d)$ at p and define

$$\bar{\delta}_{j,n}(p) = \frac{1}{n-j}\sum_{k=j+1}^{n}\mathbb{E}\left[a_{k,n}\mathbb{1}\left\{u_{k,n} > {a_{k,n}}^T p\right\}\right]$$

Define $\bar{\Omega}_{j,n} = \left\{\bar{\delta}_{j,n}(p) : p \in \Omega_p\right\}$ to be the range of $\bar{\delta}_{j,n}$. For simplicity, we write $\bar{\delta}_{0,n}$ as $\bar{\delta}_n$ and $\bar{\Omega}_{0,n}$ as $\bar{\Omega}_n$. The following Lemma relates the population prices and the δ-paths through $\bar{\delta}_{j,n}$'s.

Lemma 4. *Under Assumptions 1 and 3, for any $P \in \mathcal{P}$, when n is large enough, for all $0 \leq j < n$,*

(a) $p^*_{j,n}(\delta_{j,n}(d)) \in \Omega_p$ is uniquely determined and $p^*_{j,n}(\delta_{j,n}(d)) = p^*_n(d)$ for all $d \in \bar{\Omega}_n$.

(b) $\bar{\delta}_{j,n} : \Omega_p \to \bar{\Omega}_{j,n}$ is invertible and an open mapping, $\delta_{j,n}(d_0) \in \bar{\Omega}_{j,n}$, and $\bar{\delta}^{-1}_{j,n}(d) = p^*_{j,n}(d)$ for all $d \in \bar{\Omega}_{j,n}$.

(c) $p^*_{j,n}(\cdot)$ is continuously differentiable on $\bar{\Omega}_{j,n}$, and its Jacobian matrix at d is $-G^*_{j,n}(p^*_{j,n}(d))^{-1}$ for all $d \in \bar{\Omega}_{j,n}$

Proof. Let N_0 be given in Proposition 2(c), and fix $n \geq N_0$. For part (a), fix $d \in \bar{\Omega}_n$. By Proposition 2(b) and its proof, $g^*_{j,n}(\cdot, d)$ is differentiable and strictly convex on Ω_p for all j, and $\nabla g^*_{j,n}(p, d) = d - \bar{\delta}_{j,n}(p)$. Then, by definition, there exists $p \in \Omega_p$ such that $\nabla g^*_n(p, d) = 0$. Together with $\Omega_p \subseteq \mathbb{R}^m_+$, $\arg\min_{p \geq 0} g^*_n(p, d)$ has an unique optimal solution in Ω_p, which proves that $p^*_n(d) \in \Omega_p$ is uniquely determined. Fix $1 \leq j < n$, by definition of $\delta_{j,n}(d)$,

$$\begin{aligned}\delta_{j,n}(d) =& \frac{\sum_{k=1}^{n} \mathbb{E}\left[a_{k,n} \mathbb{1}\{u_{k,n} > {a_{k,n}}^T p^*_n(d)\}\right] - \sum_{k=1}^{j} \mathbb{E}\left[a_{k,n} \mathbb{1}\{u_{k,n} > {a_{k,n}}^T p^*_n(d)\}\right]}{n - j} \\ =& \bar{\delta}_{j,n}(p^*_n(d))\end{aligned} \tag{7}$$

which implies that $\nabla g^*_{j,n}(p^*_n(d), \delta_{j,n}(d)) = 0$. Following the same arguments, $p^*_n(d)$ is the unique optimal solution of $\min_{p \geq 0} g^*_{j,n}(p, \delta_{j,n}(d))$. Thus, $p^*_{j,n}(\delta_{j,n}(d)) \in \Omega_p$ is uniquely determined, and $p^*_n(d) = p^*_{j,n}(\delta_{j,n}(d))$. This completes the proof of part (a).

For part (b) and (c), it is sufficient to only prove the case with $j = 0$. By Proposition 2 (a) and (b), $\bar{\delta}_n(\cdot)$ is a continuously differentiable function on Ω_p, and its derivative is invertible at each point in Ω_p. Then, by Inverse Function Theorem, $\bar{\delta}_n$ is invertible and an open mapping on Ω_p, and (7) implies that $\bar{\delta}^{-1}_n(d) = p^*_n(d)$. Proposition 2(c) implies that $p^*_n(d_0) \in \Omega_p$. Therefore, $d_0 \in \bar{\Omega}_n$. Since the Jacobian matrix of $\bar{\delta}_n$ on Ω_p is $-G^*_{0,n}(p)$. Then, part (c) is directly given by Inverse Function Theorem. □

Lemma 4 shows that $\bar{\Omega}_{j,n}$ is a neighborhood of $\delta_{j,n}(d_0)$. In addition, Proposition 2 and Lemma 4 together show that the population prices $p_{j,n}$'s have good smoothness properties over $\bar{\Omega}_{j,n}$, and $g^*_{j,n}$'s have good smoothness properties over the neighborhood of the population prices evaluated on $\bar{\Omega}_{j,n}$. These smoothness properties allow us to simplify the uniform dual convergence problem to an easier empirical process problem. However, to achieve this simplification, we need to further narrow down $\bar{\Omega}_{j,n}$ to get the neighborhood $\Omega_{j,n}$ required in the uniform convergence result.

C.2 Uniform Convergence

We prove the uniform convergence result by generalizing the method in Bray [2024] to the nonstationary case. Consider the following subset of $\bar{\Omega}_{j,n}$. Let ϵ be given in Proposition 2, define

$$\Omega_{j,n} = \left\{\bar{\delta}_{j,n}(p) : p \in B_{\frac{\epsilon}{2}}(p^*_n(d_0))\right\}.$$

Lemma 4 still holds if Ω_p is replaced by $B_{\frac{\epsilon}{2}}(p^*_n(d_0))$ and $\bar{\Omega}_{j,n}$ is replaced by $\Omega_{j,n}$. Thus, $\Omega_{j,n}$ is still a neighborhood of $\delta_{j,n}(d_0)$. We start from an upper bound on the tail probability.

Proposition 3. *Under Assumptions 1 and 3, there exists C_1, C_2 and $\nu < \frac{\epsilon}{2}$ such that for all $P \in \mathcal{P}$, when n is large enough,*

$$\mathbb{P}\left\{\sup_{d \in \Omega_{j,n}} \left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2 > \nu\right\} \leq C_1 \exp\left(-C_2 \nu^2 (n - j)\right) \quad \forall 0 \leq j < n$$

The proof of Proposition 3 is long and we defer it to Section C.2.2. Proposition 3 shows that the probability of $p_{j,n}(d)$ falling outside a small neighborhood of $p^*_{j,n}(d)$ is exponentially small uniformly over $\Omega_{j,n}$. In addition, $p_{j,n}$ is bounded by the following lemma,

Lemma 5. *Under Assumption 1 and 3, there exists $\underline{d} > 0$ such that for all $P \in \mathcal{P}$, when n is large enough, for all $0 \le j < n$ and $d \in \Omega_{j,n}$, each entry of d is in $(\underline{d}, \alpha]$ and $\left\|p^*_{j,n}(d)\right\|_2 \le \frac{\bar{u}}{\underline{d}}$ almost surely.*

Proof. Fix $P \in \mathcal{P}$. Let N_0 be given by Proposition 2(c) and fix $n \ge N_0$. By definition and Assumption 3,

$$\left\|\bar{\delta}_{j,n}(p)\right\|_\infty \le \frac{1}{n-j}\sum_{k=j+1}^{n}\left\|\mathbb{E}[a_{k,n}]\right\|_\infty \le \alpha$$

Proposition 2(c) and Lemma 4(a) together show that $\delta_{j,n}(d_0) > \frac{1-\kappa}{2}\epsilon_p \cdot e$ for all j. By Proposition 2(b) and Mean Value Theorem, for any $p \in B_{\frac{\epsilon}{2}}(p^*_n(d_0))$,

$$\left\|\bar{\delta}_{j,n}(p) - \bar{\delta}_{j,n}(p^*_n(d_0))\right\| \le \frac{\bar{\sigma}\epsilon}{2}$$

We can assume ϵ is small enough such that $\frac{\bar{\sigma}\epsilon}{2} < \frac{1-\kappa}{2}\epsilon_p$. Thus, there exists $\underline{d} > 0$ such that each entry of d is greater than $\underline{d}$ when $d \in \Omega_{j,n}$ for all j. Then, by the same argument in the proof of Lemma 2, we can show that $\left\|p_{j,n}(d)\right\|_2 \le \left\|p_{j,n}(d)\right\|_1 \le \frac{\bar{u}}{\underline{d}}$ almost surely for all $d \in \Omega_{j,n}$ and j. □

Thus, we only need to focus on the case $\sup_{d\in\Omega_{j,n}}\left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2 \le \nu$, which allow us to use the smoothness properties given by Lemma 4 and Proposition 2. We can now give the complete proof of Theorem 1 in the following subsection.

C.2.1 Proof of Theorem 1

Fix $P \in \mathcal{P}$. Let N_0 be given by Proposition 2(c) and fix $n \ge N_0$. Then, by Proposition 2(a), (b) and Taylor's theorem, we can assume ϵ is small (so that Ω_p is small enough) enough such that for any $p \in \Omega_p$, $d \in \Omega_{j,n}$ and $0 \le j < n$,

$$(p - p^*_{j,n}(d))^T G^*_{j,n}(p^*_{j,n}(d))(p - p^*_{j,n}(d)) \ge \underline{\sigma}\left\|p - p^*_{j,n}(d)\right\|_2^2 \tag{8}$$

and

$$\begin{aligned}
&\left\|\nabla g^*_{j,n}(p;d) - G^*_{j,n}(p^*_{j,n}(d))(p - p^*_{j,n}(d))\right\|_2 \\
=&\left\|\nabla g^*_{j,n}(p;d) - \nabla g^*_{j,n}(p^*_{j,n}(d);d) - G^*_{j,n}(p^*_{j,n}(d))(p - p^*_{j,n}(d))\right\|_2 \\
=&\left\|G^*_{j,n}(p^*_{j,n}(d))(p - p^*_{j,n}(d)) + \mathcal{E}(p - p^*_{j,n}(d);j,n,d) - G^*_{j,n}(p^*_{j,n}(d))(p - p^*_{j,n}(d))\right\|_2 \\
=&\left\|\mathcal{E}(p - p^*_{j,n}(d);j,n,d)\right\|_2 \\
\le&\frac{\underline{\sigma}\left\|p - p^*_{j,n}(d)\right\|}{2}
\end{aligned} \tag{9}$$

(8) and (9) then imply that

$$\begin{aligned}
&(p - p^*_{j,n}(d))^T \nabla g^*_{j,n}(p;d) \\
=&(p - p^*_{j,n}(d))^T G^*_{j,n}(p^*_{j,n}(d))(p - p^*_{j,n}(d)) \\
&+ (p - p^*_{j,n}(d))^T(\nabla g^*_{j,n}(p;d) - G^*_{j,n}(p^*_{j,n}(d))(p - p^*_{j,n}(d))) \\
\ge&\underline{\sigma}\left\|p - p^*_{j,n}(d)\right\|_2^2 - \left\|\nabla g^*_{j,n}(p;d) - G^*_{j,n}(p^*_{j,n}(d))(p - p^*_{j,n}(d))\right\|_2\left\|p - p^*_{j,n}(d)\right\|_2 \\
\ge&\frac{\underline{\sigma}\left\|p - p^*_{j,n}(d)\right\|_2^2}{2}
\end{aligned} \tag{10}$$

Let $\hat{p}_{j,n}(d) = \frac{p_{j,n}(d) + p^*_{j,n}(d)}{2}$. By Lemma 6,

$$(\hat{p}_{j,n}(d) - p^*_{j,n}(d))^T \nabla g_{j,n}(\hat{p}_{j,n}(d), d) = (p_{j,n}(d) - \hat{p}_{j,n}(d))^T \nabla g_{j,n}(\hat{p}_{j,n}(d), d) \le 0 \tag{11}$$

Fix $0 \leq j < n$ and $d \in \Omega_{j,n}$. Then, if $p_{j,n}(d) \in \Omega_P$, $\hat{p}_{j,n}(d) \in \Omega_p$. Together with (10) and (11), if $p_{j,n}(d) \in \Omega_P$, then

$$\begin{aligned}
&\left\|\hat{p}_{j,n}(d) - p^*_{j,n}(d)\right\|_2 \left\|\nabla g^*_{j,n}(\hat{p}_{j,n}(d); d) - \nabla g_{j,n}(\hat{p}_{j,n}(d); d)\right\|_2 \\
\geq &(\hat{p}_{j,n}(d) - p^*_{j,n}(d))^T (\nabla g^*_{j,n}(\hat{p}_{j,n}(d); d) - \nabla g_{j,n}(\hat{p}_{j,n}(d); d)) \\
\geq &(\hat{p}_{j,n}(d) - p^*_{j,n}(d))^T \nabla g^*_{j,n}(\hat{p}_{j,n}(d); d) \\
\geq &\frac{\underline{\sigma}\left\|\hat{p}_{j,n}(d) - p^*_{j,n}(d)\right\|_2^2}{2}
\end{aligned}$$

which implies that

$$\left\|\nabla g^*_{j,n}(\hat{p}_{j,n}(d); d) - \nabla g_{j,n}(\hat{p}_{j,n}(d); d)\right\|_2 \geq \frac{\underline{\sigma}\left\|\hat{p}_{j,n}(d) - p^*_{j,n}(d)\right\|_2}{2} = \frac{\underline{\sigma}\left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2}{4} \tag{12}$$

Let ν be given by Proposition 3. By the definition of $\Omega_{j,n}$, $p^*_{j,n}(d) \in B_{\frac{\epsilon}{2}}(p^*_n(d_0))$. Together with the fact $\nu < \frac{\epsilon}{2}$ and Proposition 2(c), if $p_{j,n}(d) \in B_\nu(p^*_{j,n}(d))$, then $p_{j,n}(d) \in \Omega_p$. Thus, (12) and Lemma 8(b) imply that, when $n - j$ is greater than some constant N_1 that holds for all $P \in \mathcal{P}$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{d \in \Omega_{j,n}} \mathbb{1}\{p_{j,n}(d) \in B_\nu(p^*_{j,n}(d))\}\left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2^2\right] \\
\leq &\frac{16}{\underline{\sigma}^2}\mathbb{E}\left[\sup_{d \in \Omega_{j,n}}\left\|\nabla g^*_{j,n}(\hat{p}_{j,n}(d), d) - \nabla g_{j,n}(\hat{p}_{j,n}(d), d)\right\|_2^2\right] \\
\leq &\frac{16}{\underline{\sigma}^2}\mathbb{E}\left[\sup_{p \in \Omega_p}\left\|\frac{1}{n-j}\sum_{k=j+1}^{n} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{k,n}(p)\right\|_2^2\right] \leq \frac{C_3 \log(n-j)}{n-j}
\end{aligned} \tag{13}$$

where C_3 is a constant that holds for all $P \in \mathcal{P}$ and does not depend on j and n. By Proposition 3 and Lemma 5,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{d \in \Omega_{j,n}} \mathbb{1}\{p_{j,n}(d) \notin B_\nu(p^*_{j,n}(d))\}\left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2^2\right] \\
\leq &\frac{4\bar{u}^2}{\underline{d}^2}\mathbb{P}\left\{\sup_{d \in \Omega_{j,n}}\left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2 > \nu\right\} \\
\leq &\frac{4\bar{u}^2}{\underline{d}^2} C_1 \exp\left(-C_2 \nu^2 (n-j)\right)
\end{aligned} \tag{14}$$

Lemma 5 also implies that $\mathbb{E}\left[\sup_{d \in \Omega_{j,n}}\left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2^2\right] \leq \frac{4\bar{u}^2}{\underline{d}^2}$ if $n - j \leq N_1$. Together with (13) and (14), the proof is complete.

C.2.2 Proof of Proposition 3

Fix $P \in \mathcal{P}$ satisfying Assumption 3 throughout this section. We first introduce some notations. Define

$$\nabla g_{j,n}(p, d) = d - \frac{1}{n-j}\sum_{k=j+1}^{n} a_{k,n}\mathbb{1}\left\{u_{k,n} > {a_{k,n}}^T p\right\}$$

which is a subgradient of $g_{j,n}(\cdot, d)$ at p, and recall $G^*_{j,n}(p)$ is the Hessian of $g^*_{j,n}(\cdot, d)$ at p. By Proposition 2(b), for any $p \in \Omega_p$, $G^*_{j,n}(p)$ has orthonormal eigenvectors $\left\{\omega^i_{j,n}(p)\right\}_{i=1}^m$ and the corresponding eigenvalues $\left\{\sigma^i_{j,n}(p)\right\}_{i=1}^m$ satisfying $\underline{\sigma} < \sigma^m_{j,n}(p) \leq \cdots \leq \sigma^1_{j,n}(p) < \bar{\sigma}$. The following are three useful technical Lemmas.

Lemma 6. $\nabla g_{j,n}(p,d)^T(p_{j,n}(d)-p)\le 0$ *for all* $d,\ p\in\mathbb{R}^m_+$ *and* $j,\ n$.

Proof. Fix $d,\ p\in\mathbb{R}^m_+$ and $j,\ n$, because $\nabla g_{j,n}(p,d)$ is a subgradient of $g_{j,n}(\cdot,d)$ at p,

$$g_{j,n}(p_{j,n}(d),d)-g_{j,n}(p,d)\ge\nabla g_{j,n}(p,d)^T(p_{j,n}(d)-p)$$

Since $p_{j,n}(d)$ minimizes $g_{j,n}(\cdot,d)$ over $p\ge 0$, $g_{j,n}(p_{j,n}(d),d)-g_{j,n}(p,d)\le 0$, which completes the proof. □

Lemma 7. *When* n *is large enough, for all* j *and* $d\in\bar{\Omega}_n$, *if* $p\in\mathbb{R}^m_+$ *and* $\eta>0$ *satisfy* $p+k\eta\omega^i_{j,n}(p^*_n(d))\in\mathbb{R}^m_+$ *and*

$$\left\|\nabla g_{j,n}(p+k\eta\omega^i_{j,n}(p^*_n(d)),\delta_{j,n}(d))-k\eta\sigma^i_{j,n}(p^*_n(d))\omega^i_{j,n}(p^*_n(d))\right\|_2\le\frac{\eta\sigma^m_{j,n}(p^*_n(d))}{2\sqrt{m}}$$

for all $1\le i\le m$ *and* $k\in\{-1,1\}$, *then*

$$\left\|p_{j,n}(\delta_{j,n}(d))-p\right\|_2\le\eta\left(1+\frac{2\sqrt{m}\sigma^1_{j,n}(p^*_n(d))}{\sigma^m_{j,n}(p^*_n(d))}\right)$$

Proof. Let N_0 be given by Proposition 2. Fix $n\ge N_0$ and $d\in\bar{\Omega}_n$. By Lemma 4, $p^*_n(d)\in\Omega_p$, which makes $\left\{\omega^i_{j,n}(p^*_n(d))\right\}_{i=1}^m$ and $\left\{\sigma^i_{j,n}(p^*_n(d))\right\}_{i=1}^m$ well defined. Let p satisfy the required conditions, by Lemma 6,

$$\begin{aligned}
0\ge&\left(p_{j,n}(\delta_{j,n}(d))-p-k\eta\omega^i_{j,n}(p^*_n(d))\right)^T\nabla g_{j,n}(p+k\eta\omega^i_{j,n}(p^*_n(d)),\delta_{j,n}(d))\\
=&\left(p_{j,n}(\delta_{j,n}(d))-p-k\eta\omega^i_{j,n}(p^*_n(d))\right)^T k\eta\sigma^i_{j,n}(p^*_n(d))\omega^i_{j,n}(p^*_n(d))\\
&+\left(p_{j,n}(\delta_{j,n}(d))-p-k\eta\omega^i_{j,n}(p^*_n(d))\right)^T\left(\nabla g_{j,n}(p+k\eta\omega^i_{j,n}(p^*_n(d)),\delta_{j,n}(d))-\right.\\
&\qquad\left.k\eta\sigma^i_{j,n}(p^*_n(d))\omega^i_{j,n}(p^*_n(d))\right)\\
\ge&k\eta\sigma^i_{j,n}(p^*_n(d))\left(p_{j,n}(\delta_{j,n}(d))-p\right)^T\omega^i_{j,n}(p^*_n(d))-\eta^2\sigma^i_{j,n}(p^*_n(d))\omega^i_{j,n}(p^*_n(d))^T\omega^i_{j,n}(p^*_n(d))\\
&-\left(\left\|p_{j,n}(\delta_{j,n}(d))-p-k\eta\omega^i_{j,n}(p^*_n(d))\right\|_2\cdot\right.\\
&\qquad\left.\left\|\nabla g_{j,n}(p+k\eta\omega^i_{j,n}(p^*_n(d)),\delta_{j,n}(d))-k\eta\sigma^i_{j,n}(p^*_n(d))\omega^i_{j,n}(p^*_n(d))\right\|_2\right)
\end{aligned}$$

where the last line is by Cauchy Inequality. Since $\omega^i_{j,n}(p^*_n(d))$'s are orthonormal and the above inequality is true for both $k=1$ and $k=-1$,

$$\begin{aligned}
0\ge&\eta\sigma^m_{j,n}(p^*_n(d))\left|\left(p_{j,n}(\delta_{j,n}(d))-p\right)^T\omega^i_{j,n}(p^*_n(d))\right|-\eta^2\sigma^1_{j,n}(p^*_n(d))\\
&-\left(\left\|p_{j,n}(\delta_{j,n}(d))-p\right\|_2+\eta\right)\frac{\eta\sigma^m_{j,n}(p^*_n(d))}{2\sqrt{m}}
\end{aligned}$$

$\omega^i_{j,n}(p^*_n(d))$'s are orthonormal also implies that there exists i such that

$$\left|\left(p_{j,n}(\delta_{j,n}(d))-p\right)^T\omega^i_{j,n}(p^*_n(d))\right|\ge\frac{\left\|p_{j,n}(\delta_{j,n}(d))-p\right\|_2}{\sqrt{m}}$$

Then,

$$\begin{aligned}
0\ge&\frac{\eta\sigma^m_{j,n}(p^*_n(d))\left\|p_{j,n}(\delta_{j,n}(d))-p\right\|_2}{\sqrt{m}}-\eta^2\sigma^1_{j,n}(p^*_n(d))\\
&-\left(\left\|p_{j,n}(\delta_{j,n}(d))-p\right\|_2+\eta\right)\frac{\eta\sigma^m_{j,n}(p^*_n(d))}{2\sqrt{m}}
\end{aligned}$$

which is equivalent to

$$\left\|p_{j,n}(\delta_{j,n}(d)) - p\right\|_2 \leq \eta\left(1 + \frac{2\sqrt{m}\sigma^1_{j,n}(p^*_n(d))}{\sigma^m_{j,n}(p^*_n(d))}\right)$$

□

Lemma 8. *Under Assumption 1, 3, for any $P \in \mathcal{P}$,*

(a)

$$\mathbb{P}\left\{\sup_{p\in\Omega_p}\left\|\frac{1}{n-j}\sum_{k=j+1}^{n} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{k,n}(p)\right\|_2 > \nu\right\}$$
$$\leq \left(1 + \left\lceil\frac{6m\bar{u}\bar{\sigma}}{d_0^{min}\nu}\right\rceil\right)^m m\exp\left(-\frac{2(n-j)\nu^2}{9m\alpha^2}\right)$$

(b) When $n-j$ is large enough,

$$\mathbb{E}\left[\sup_{p\in\Omega_p}\left\|\frac{1}{n-j}\sum_{k=j+1}^{T} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{k,n}(p)\right\|_2^2\right]$$
$$\leq m\left(2^{m-1} + \frac{1}{2}\left(\frac{18m\bar{u}\bar{\sigma}}{d^0_{min}}\right)^m\right)\frac{9m^2\alpha^2\log(n-j)}{2(n-j)}$$

Proof. For part (a), by Hoeffding Inequality, for any $p \in \mathbb{R}^m$,

$$\mathbb{P}\left\{\left\|\frac{1}{n-j}\sum_{k=j+1}^{n} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{k,n}(p)\right\|_2 > \frac{\nu}{3\sqrt{m}}\right\} \leq m\exp\left(-\frac{2(n-j)\nu^2}{9m\alpha^2}\right)$$

We can assume Ω_p is small enough such that it is a subset of a hyper cube $\bar{\Omega}_p$, and Proposition 2 holds if we replace Ω_p in its statements by $\bar{\Omega}_p$. We can also assume $\bar{\Omega}_p \subseteq \left[0, \frac{2\bar{u}}{d_0^{min}}\right]^m$ where d_0^{min} is the smallest entry of d_0. We can denote $\bar{\Omega}_p$ as $\otimes_{i=1}^m \left[\underline{p_i}, \bar{p}_i\right]$ Then, we construct the following grid $\mathcal{G}$

$$\mathcal{G} = \left\{\left(\left(\underline{p_1} + \frac{\nu}{3m\bar{\sigma}}k_1\right)\wedge\bar{p}_1, \cdots, \left(\underline{p_m} + \frac{\nu}{3m\bar{\sigma}}k_m\right)\wedge\bar{p}_m\right) :\right.$$
$$\left. 0 \leq k_i \leq \left\lceil\frac{3m(\bar{p}_i - \underline{p_i})\bar{\sigma}}{\nu}\right\rceil,\ 1 \leq i \leq m\right\}$$

Then, for any $p \in \Omega_p$, there exists $p_1, p_2 \in \mathcal{G}$ such that $p_1 \leq p \leq p_2$ and $\|p_1 - p_2\|_2 \leq \frac{\nu}{3\sqrt{m}\bar{\sigma}}$. "$\leq$" between vectors element-wise smaller than or equal to. By Mean Value Theorem, $\left\|\bar{\delta}_{j,n}(p_1) - \bar{\delta}_{j,n}(p_2)\right\|_2 \leq \frac{\nu}{3\sqrt{m}}$. By monotonicity, $\bar{\delta}_{j,n}(p_2) \leq \bar{\delta}_{j,n}(p) \leq \bar{\delta}_{j,n}(p_1)$ and

$$\frac{1}{n-j}\sum_{k=j+1}^{n} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p_2\}$$
$$\leq \frac{1}{n-j}\sum_{k=j+1}^{n} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\}$$
$$\leq \frac{1}{n-j}\sum_{k=j+1}^{n} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p_1\} \quad \text{a.s.}$$

Define

$$\mathcal{E} = \left\{ \left\| \frac{1}{n-j} \sum_{k=j+1}^{T} a_{k,n} \mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{j,n}(p) \right\|_2 \leq \frac{\nu}{3\sqrt{m}} : p \in \mathcal{G} \right\}$$

Let $\{e_i\}_{i=1}^m$ be the standard basis, then on $\mathcal{E}$, for any $p \in \Omega_p$,

$$\begin{aligned} &{e_i}^T \bar{\delta}_{j,n}(p), {e_i}^T \left(\frac{1}{n-j} \sum_{k=j+1}^{n} a_{k,n} \mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} \right) \\ \in &\left({e_i}^T \bar{\delta}_{j,n}(p_2) - \frac{\nu}{3\sqrt{m}}, {e_i}^T \bar{\delta}_{j,n}(p_1) + \frac{\nu}{3\sqrt{m}} \right) \end{aligned}$$

which implies that

$$\left\| \frac{1}{n-j} \sum_{k=j+1}^{n} a_{k,n} \mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{j,n}(p) \right\|_2 \leq \nu \quad \forall p \in \Omega_p$$

Thus,

$$\begin{aligned} &\mathbb{P}\left\{ \sup_{p \in \Omega_p} \left\| \frac{1}{n-j} \sum_{k=j+1}^{n} a_{k,n} \mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{j,n}(p) \right\|_2 > \nu \right\} \\ \leq &1 - \mathbb{P}\{\mathcal{E}\} \leq |\mathcal{G}|\, m \exp\left(-\frac{2(n-j)\nu^2}{9m\alpha^2} \right) \leq \left(1 + \left\lceil \frac{6m\bar{u}\bar{\sigma}}{d_{min}^0 \nu} \right\rceil \right)^m m \exp\left(-\frac{2(n-j)\nu^2}{9m\alpha^2} \right) \end{aligned}$$

For part (b),

$$\begin{aligned} &\mathbb{E}\left[\sup_{p \in \Omega_p} \left\| \frac{1}{n-j} \sum_{k=j+1}^{n} a_{k,n} \mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{k,n}(p) \right\|_2^2 \right] \\ \leq &\frac{1}{2} \int_0^{+\infty} 1 \wedge \left\{ \left(1 + \left\lceil \frac{6m\bar{u}\bar{\sigma}}{d_{min}^0 \sqrt{\nu}} \right\rceil \right)^m m \exp\left(-\frac{2(n-j)\nu}{9m\alpha^2} \right) \right\} dv \\ \leq &2^{m-1} m \int_0^{+\infty} \exp\left(-\frac{2(n-j)\nu}{9m\alpha^2} \right) dv \\ &+ \frac{1}{2} \left(\frac{18m\bar{u}\bar{\sigma}}{d_{min}^0} \right)^m m \int_0^{+\infty} 1 \wedge \left\{ v^{-\frac{m}{2}} \exp\left(-\frac{2(n-j)\nu}{9m\alpha^2} \right) \right\} dv \end{aligned}$$

where the second inequality is by the fact that $1 + \lceil x \rceil \leq 3x$ if $x \geq 1$. To evaluate the right-hand-side. First,

$$\int_0^{+\infty} \exp\left(-\frac{2(n-j)\nu}{9m\alpha^2} \right) dv = \frac{9m\alpha^2}{2(n-j)}$$

Secondly, let $v = \frac{9m^2\alpha^2 \log(n-j)}{4(n-j)} v'$

$$\begin{aligned} &\int_0^{+\infty} 1 \wedge v^{-\frac{m}{2}} \exp\left(-\frac{2(n-j)\nu}{9m\alpha^2} \right) dv \\ = &\int_0^{+\infty} 1 \wedge \exp\left(-\frac{2(n-j)\nu}{9m\alpha^2} + \frac{m}{2} \log\left(\frac{1}{v} \right) \right) dv \\ = &\frac{9m^2\alpha^2 \log(n-j)}{4(n-j)} \int_0^{+\infty} 1 \wedge \exp\left(-\frac{m}{2} \log(n-j) v' + \frac{m}{2} \log\left(\frac{4(n-j)}{9m^2\alpha^2 \log(n-j)\alpha^2 v'} \right) \right) dv' \end{aligned}$$

When $n-j$ is large enough,

$$\begin{aligned}
&\int_0^{+\infty} 1 \wedge \exp\left(-\frac{m}{2}\log(n-j)v' + \frac{m}{2}\log\left(\frac{4(n-j)}{9m^2\alpha^2\log(n-j)\alpha^2 v'}\right)\right) dv' \\
\leq & 1 + \int_1^{+\infty} \exp\left(-\frac{m}{2}\log(n-j)v' + \frac{m}{2}\log\left(\frac{n-j}{v'}\right)\right) dv' \\
\leq & 1 + \int_1^{+\infty} \exp\left(-\frac{m}{2}\log(n-j)v' + \frac{m}{2}\log\left(n-j\right)\right) dv' \\
\leq & 1 + \frac{2}{m\log(n-j)} \leq 2
\end{aligned}$$

Summarize all the above, when $n-j$ is large enough,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{p\in\Omega_p}\left\|\frac{1}{n-j}\sum_{k=j+1}^{n} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{j,n}(p)\right\|_2^2\right] \\
\leq & m\left(2^{m-1} + \frac{1}{2}\left(\frac{18m\bar{u}\bar{\sigma}}{d_{min}^0}\right)^m\right)\frac{9m^2\alpha^2\log(n-j)}{2(n-j)}
\end{aligned}$$

□

We can now give the complete proof of Proposition 3. Fix $P \in \mathcal{P}$ and let N_0 be given by Proposition 2(c). Fix $n \geq N_0$, by the same arguments in the proof of Lemma 4, $p_{j,n}^*(d) \in B_{\frac{\epsilon}{2}}(p_n^*(d_0))$ for all $d \in \Omega_{j,n}$. Together with Proposition 2(c), there exists $\nu < \frac{\epsilon}{2}$ that holds for all $P \in \mathcal{P}$ such that,

$$p_{j,n}^*(d) + \frac{\nu k \omega_{j,n}^i(p_{j,n}^*(d))}{1 + \frac{8\sqrt{m}\bar{\sigma}}{\underline{\sigma}}} \in \Omega_p \quad \forall d \in \Omega_{j,n},\ 1 \leq i \leq m,\ 0 \leq j < n$$

Denote $\eta(\nu) = \frac{\nu}{1+\frac{8\sqrt{m}\bar{\sigma}}{\underline{\sigma}}}$ By Lemma 4 and Lemma 7, for all $0 \leq j < n$,

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{d\in\Omega_{j,n}}\left\|p_{j,n}(d) - p_{j,n}^*(d)\right\|_2 > \nu\right\} \\
\leq & \mathbb{P}\left\{\sup_{d\in\Omega_{j,n}}\left\|p_{j,n}(d) - p_{j,n}^*(d)\right\|_2 > \left(1 + \frac{2\sqrt{m}\bar{\sigma}}{\underline{\sigma}}\right)\eta(\nu)\right\} \\
\leq & \mathbb{P}\left\{\sup_{d\in\Omega_{j,n}}\max_{1\leq i\leq m}\max_{k\in\{-1,1\}}||\nabla g_{j,n}(p_{j,n}^*(d) + k\eta(\nu)\omega_{j,n}^i(p_{j,n}^*(d)); d)) \right.\\
& \qquad \left. -k\eta(\nu)\sigma_{j,n}^i(p_{j,n}^*(d))\omega_{j,n}^i(p_{j,n}^*(d))||_2 > \frac{\underline{\sigma}\eta(\nu)}{4\sqrt{m}}\right\}
\end{aligned} \tag{15}$$

By Proposition 2(b),

$$\begin{aligned}
&\left\|\nabla g_{j,n}^*(p_{j,n}^*(d) + k\eta(\nu)\omega_{j,n}^i(p_{j,n}^*(d)); d)) - k\eta(\nu)\sigma_{j,n}^i(p_{j,n}^*(d))\omega_{j,n}^i(p_{j,n}^*(d))\right\|_2 \\
=&||\nabla g_{j,n}^*(p_{j,n}^*(d) + k\eta(\nu)\omega_{j,n}^i(p_{j,n}^*(d)); d)) \\
&- \nabla g_{j,n}^*(p_{j,n}^*(d); d)) - k\eta(\nu)\sigma_{j,n}^i(p_{j,n}^*(d))\omega_{j,n}^i(p_{j,n}^*(d))||_2 \\
=&\left\|G_{j,n}^*(p_{j,n}^*(d))(k\eta(\nu)\omega_{j,n}^i(p_{j,n}^*(d))) + \mathcal{E}_{j,n}(\nu; i, k, d) - k\eta(\nu)\sigma_{j,n}^i(p_{j,n}^*(d))\omega_{j,n}^i(p_{j,n}^*(d))\right\|_2 \\
=&\left\|\mathcal{E}_{j,n}(\nu; i, k, d)\right\|_2
\end{aligned}$$

where $\mathcal{E}(\cdot)$ denotes the residual term in Taylor's Theorem. Proposition 2(a) guarantees that $\mathcal{E}(\cdot)$ is $o(\nu)$ uniformly over j, $n \geq N_0$, i, k, and $d \in \Omega_{j,n}$. In addition, we can assume that ν are small

enough such that, for all $0 \le j < n$

$$\sup_{d\in\Omega_{j,n}} \max_{1\le i\le m} \max_{k\in\{-1,1\}} ||\nabla g^*_{j,n}(p^*_{j,n}(d) + k\eta(\nu)\omega^i_{j,n}(p^*_{j,n}(d)); d)) - k\eta(\nu)\sigma^i_{j,n}(p^*_{j,n}(d))\omega^i_{j,n}(p^*_{j,n}(d))||_2 \le \frac{\eta(\nu)\underline{\sigma}}{8\sqrt{m}} \tag{16}$$

Then, by (15), (16), and Lemma 8, there exists C_1 and C_2 that holds for all $P \in \mathcal{P}$ such that

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{d\in\Omega_{j,n}(\nu)} \left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2 > \nu\right\} \\
\le&\mathbb{P}\left\{\sup_{d\in\Omega_{j,n}(\nu)} \max_{1\le i\le m} \max_{k\in\{-1,1\}} ||\nabla g_{j,n}(p^*_{j,n}(d) + k\eta(\nu)\omega^i_{j,n}(p^*_{j,n}(d)); d)) \right.\\
&\qquad \left. -\nabla g^*_{j,n}(p^*_{j,n}(d) + k\eta(\nu)\omega^i_{j,n}(p^*_{j,n}(d)); d))||_2 > \frac{\underline{\sigma}\eta(\nu)}{8\sqrt{m}}\right\} \\
\le&\mathbb{P}\left\{\sup_{p\in\Omega_p} \left\|\frac{1}{n-j}\sum_{k=j+1}^{T} a_{k,n}\mathbb{1}\{u_{k,n} > {a_{k,n}}^T p\} - \bar{\delta}_{k,n}(p)\right\|_2 > \frac{\underline{\sigma}\eta(\nu)}{8\sqrt{m}}\right\} \\
\le&C_1 \exp\left(-C_2\nu^2(n-j)\right)
\end{aligned}$$

D Average Remaining resource

Under Assumption 2, the uniform dual convergence result in Theorem 1 also holds for the prices $p^{\pi}_{j,n}$ used in Algorithm 1. However, the uniform dual convergence can be used only when the average remaining resource $d_{j,n} \equiv b_{j,n}/(n-j)$ is in $\Omega_{j,n}$. When $d_{j,n}$ is not in $\Omega_{j,n}$ or we run out of resources, we have very little control over the quality of the decisions made by Algorithm 1. To prove the regret upper bound, we need to analyze how many times we cannot apply the uniform dual convergence result when we implement Algorithm 1. Let $b_{j,n}(i)$ be the ith entry of $b_{j,n}$. Define

$$\bar{\tau}_n = \inf\{j \ge 1 : \exists i \text{ s.t. } b_{j,n}(i) < \alpha \text{ or } d_{j,n} \notin \Omega_{j,n}\} \wedge n$$

Let $\{\mathcal{F}_{j,n}\}_{j=0}^n$ be the natural filtration of $\{(u_{j,n}, a_{j,n})\}_{j=1}^n$, i.e., $\mathcal{F}_{0,n} = \{\emptyset, \Omega\}$ and $\mathcal{F}_{j,n} = \sigma(\{(u_{k,n}, a_{k,n})\}_{k=1}^j)$ for $1 \le j \le n$. Then, $\bar{\tau}_n$ is a stopping time adapted to $\{\mathcal{F}_{j,n}\}_{j=0}^n$. Our goal of this section is to show $\mathbb{E}[n - \bar{\tau}_n]$ is small. The following Lemma provides a characterization of $\Omega_{j,n}$ which allow us to constructing stopping times that are easier to analyze and smaller than or equal to $\bar{\tau}_n$.

Lemma 9. *Under Assumptions 1 and 3, there exists $\epsilon_d > 0$ such that for all $P \in \mathcal{P}$, when n is large enough,*

$$B_{\epsilon_d}(\delta_{j,n}(d_0)) \subseteq \Omega_{j,n} \quad \forall 0 \le j < n$$

We defer the proof of Lemma 9 to Section D.2. Let $\underline{d}$ be given in Lemma 5. Then, Lemma 5 and 9 imply that if

$$\iota \ge \left\lceil \frac{\alpha}{\underline{d}} \right\rceil \text{ and } 0 < \epsilon_{j,n} < \epsilon_d \ \forall 0 \le j \le n - \iota - 1 \tag{17}$$

then

$$\bar{\tau}_n \ge \tau_n \equiv \inf\left\{1 \le j \le n - \iota - 1 : d_{j,n} \notin B_{\epsilon_{j,n}}(\delta_{j,n}(d_0))\right\} \wedge n - \iota \quad \text{a.s.}$$

In the rest of this section, we focus on how to select appropriate ι and $\{\epsilon_{j,n}\}_{j=1}^n$ so that $\mathbb{E}[n - \tau_n]$ is $O((\log n))$. Let $e \in \mathbb{R}^m$ be the vector that every entry is 1. Define

$$d'_{j,n} = d_{j,n}\mathbb{1}\{j \le \tau_n\} + 2\alpha e\mathbb{1}\{j > \tau_n\}$$

In other words, $d'_{j,n}$ is the average remaining resource up to the stopping time τ_n and is fixed to a relatively large value $2\alpha e$ after τ_n. In addition, let $X_{j,n} = d'_{j,n} - \delta_{j,n}(d_0)$, which is $d'_{j,n}$ centered by $\delta_{j,n}(d_0)$. Then, the definition of τ_n together with Lemma 5 implies that $P(\tau_n \le j) \le P(\left\|X_{j,n}\right\|_2 >$

$\epsilon_{j,n}$) for all $j < n - \iota$. Thus, to get an upper bound of $\mathbb{E}[n - \tau_n]$, we can focus on the tail probability of $X_{j,n}$. Define

$$Z_{j,n} = \mathbb{E}\left[X_{j+1,n}\middle|\mathcal{F}_{j,n}\right] - X_{j,n}$$

Then $M_{j,n} = X_{j,n} - \sum_{k=0}^{j-1} Z_{k,n}$ forms a martingale adapted to $\{\mathcal{F}_{j,n}\}_{j=0}^{n}$. Note that $d'_{j,n}$, $X_{j,n}$, $Z_{j,n}$ and $M_{j,n}$ all depend on ι and $\epsilon_{j,n}$'s selected for the stopping time τ_n, but we do not include them into the notation for simplicity. The following lemma gives an upper bound on the tail probability of $M_{j,n}$.

Lemma 10. *Under Assumption 3, for any $P \in \mathcal{P}$, if ι and $\{\epsilon_{j,n}\}_{j=0}^{n-\iota-1}$ satisfy (17), then*

$$P\left\{\left\|M_{j,n}\right\|_2 \geq \eta\right\} \leq 2m \exp\left(-\frac{\eta^2(n-j-1)}{(8\alpha + 2\epsilon_d)^2}\right)$$

for all $n \in \mathbb{N}_+$, $0 \leq j \leq n$ and $\eta > 0$.

Proof. Because $X_{j,n}$ and $Z_{j,n}$ are both measurable with respect to $\mathcal{F}_{j,n}$, $M_{j,n}$ is measurable with respect to $\mathcal{F}_{j,n}$. Also,

$$\begin{aligned}
\mathbb{E}\left[M_{j+1,n}\middle|\mathcal{F}_{j,n}\right] =& \mathbb{E}\left[X_{j+1,n} - \sum_{k=0}^{j} Z_{k,n}\middle|\mathcal{F}_{j,n}\right] \\
=& \mathbb{E}\left[X_{j+1,n}\middle|\mathcal{F}_{j,n}\right] - \sum_{k=0}^{j} Z_{k,n} \\
=& X_{j,n} + Z_{j,n} - \sum_{k=0}^{j} Z_{k,n} = X_{j,n} - \sum_{k=0}^{j-1} Z_{k,n} = M_{j,n}
\end{aligned}$$

Thus, $\{M_{j,n}\}_{j=0}^{n}$ is a martingale adapted to $\{\mathcal{F}_{j,n}\}_{j=0}^{n}$. In addition,

$$\begin{aligned}
(M_{j+1,n} - M_{j,n})\mathbb{1}\{j \geq \tau_n\} =& \left(X_{j+1,n} - X_{j,n} - Z_{j,n}\right)\mathbb{1}\{j \geq \tau_n\} \\
=& \left(X_{j+1,n} - \mathbb{E}[X_{j+1,n}\middle|\mathcal{F}_{j,n}]\right)\mathbb{1}\{j \geq \tau_n\} \\
=& X_{j+1,n}\mathbb{1}\{j \geq \tau_n\} - \mathbb{E}[X_{j+1,n}\mathbb{1}\{j \geq \tau_n\}\middle|\mathcal{F}_{j,n}] \\
=& 2\alpha e\mathbb{1}\{j \geq \tau_n\} - 2\alpha e\mathbb{1}\{j \geq \tau_n\} = 0
\end{aligned} \tag{18}$$

and let $M_{j,n}(i)$ and $X_{j,n}(i)$ be their ith entry, then

$$\begin{aligned}
&(M_{j+1,n}(i) - M_{j,n}(i))\mathbb{1}\{j < \tau_n\} \\
\leq& \left|X_{j+1,n}(i) - X_{j,n}(i)\right|\mathbb{1}\{j < \tau_n\} + \mathbb{E}\left[\left|X_{j+1,n}(i) - X_{j,n}(i)\right|\mathbb{1}\{j < \tau_n\}\middle|\mathcal{F}_{j,n}\right]
\end{aligned}$$

By Assumption 3 and Lemma 5

$$\begin{aligned}
&\left|X_{j+1,n}(i) - X_{j,n}(i)\right|\mathbb{1}\{j < \tau_n\} \\
\leq& \left(\left|d_{j+1,n}(i) - d_{j,n}(i)\right| + \left|\delta_{j+1,n}(d_0)(i) - \delta_{j,n}(d_0)(i)\right|\right)\mathbb{1}\{j < \tau_n\} \\
\leq& \left(\frac{a_{j+1,n}(i) + d_{j,n}(i)}{n-j-1} + \frac{\mathbb{E}[a_{j+1,n}(i)] + \delta_{j,n}(d_0)(i)}{n-j-1}\right)\mathbb{1}\{t < \tau_T\} \leq \frac{4\alpha + \epsilon_d}{n-j-1}
\end{aligned}$$

Then, by Azuma's Inequality and Union Bound,

$$P\left\{\left\|M_{j,n}\right\|_2 \geq \eta\right\} \leq 2m \exp\left(-\frac{\eta^2}{(8\alpha + 2\epsilon_d)^2 \sum_{k=1}^{j} \frac{1}{(n-k)^2}}\right) \leq 2m \exp\left(-\frac{\eta^2(n-j-1)}{(8\alpha + 2\epsilon_d)^2}\right)$$

□

The fast decay rate of the tail probability of $M_{j,n}$'s given in Lemma 10 suggests that, if we can select $\epsilon_{j,n}$'s such that $P(\|X_{j,n}\|_2 > \epsilon_{j,n}) \leq P(\|M_{j,n}\|_2 \geq \eta)$ for all $0 \leq j < n - \iota$ under some small enough ι and η, then we can achieve the desired upper bound of $\mathbb{E}[n - \bar{\tau}_n]$. To find ι, η and construct $\epsilon_{j,n}$'s that satisfy this condition, we characterize $Z_{j,n}$'s in the following way. Define

$$Z_n(j,d) = \mathbb{E}\left[d_{j+1,n} - \delta_{j+1,n}(d_0) \middle| d_{j,n} = d\right] - (d - \delta_{j,n}(d_0))$$

By the independence property given in Assumption 2 and Assumption, 3

$$Z_{j,n}\mathbb{1}\{j < \tau_n\} = Z_n(j, d_{j,n})\mathbb{1}\{j < \tau_n\} \quad \text{a.s.} \tag{19}$$

In the setting that $\{u_{j,n}, a_{j,n}\}_{j=1}^n$ are i.i.d. and satisfy Assumptions 3, $\delta_{j,n}(d) = d$ for all $d \in B_{\epsilon_d}(d_0)$, then

$$\begin{aligned} Z_{j,n}\mathbb{1}\{j < \tau_n\} &= \frac{d_{j,n} - \mathbb{E}\left[a_{j+1,n}\mathbb{1}\left\{u_{j+1,n} > {a_{j+1,n}}^T p^{\pi}_{j,n}\right\} \middle| d_{j,n}\right]}{n - j - 1}\mathbb{1}\{j < \tau_n\} \\ &= \frac{1}{n-j-1}\left(\mathbb{E}\left[a_{j+1,n}\left(\mathbb{1}\left\{u_{j+1,n} > {a_{j+1,n}}^T p^{*}_{j,n}(d_{j,n})\right\}\right.\right.\right. \\ &\qquad \left.\left.\left. -\mathbb{1}\left\{u_{j+1,n} > {a_{j+1,n}}^T p^{\pi}_{j,n}\right\}\right) \middle| d_{j,n}\right]\right)\mathbb{1}\{j < \tau_n\} \end{aligned}$$

Thus, in the i.i.d. case, the uniform dual convergences results imply $Z_{j,n}$'s are small enough so that there is no big difference between $X_{j,n}$ and $M_{j,n}$. Then, we can directly take $\epsilon_{j,n} = \epsilon_d$, which will make our analysis of the stopping time almost the same as the analysis in the literature of the i.i.d. case (ex. Bray [2024]). However, in the nonstationary setting, $\{\delta_{j,n}(d)\}_{j=1}^n$ is no longer a constant sequence, and $X_{j,n}$ and $M_{j,n}$ are different significantly. We need to handle this difference by appropriately constructing $\epsilon_{j,n}$. The following theorem is a formal version of Theorem 2 and summarizes our approach and provide the main result of this section.

Theorem 4. *Under Assumptions 1, 2, 3, there exist constants η, C_{Res} and $\{\iota_n\}_{n=1}^{+\infty}$ such that for all $P \in \mathcal{P}$, when n is large enough, if $\epsilon_{0,n} = \eta$, and*

$$\epsilon_{j+1,n} = \epsilon_{j,n} + \max_{d \in B_{\epsilon_{j,n}}(\delta_{j,n}(d_0))} \|Z_n(j,d)\|_2 \quad 0 \leq j < n - \iota_n$$

then

$$\bar{\tau}_n \geq \tau_n \equiv \inf\left\{1 \leq j \leq n - \iota_n - 1 : d_{j,n} \notin B_{\epsilon_{j,n}}(\delta_{j,n}(d_0))\right\} \wedge n - \iota_n \quad \text{a.s.}$$

and

$$\mathbb{E}[n - \bar{\tau}_n] \leq C_{Res} \log n.$$

A key step in the proof of Theorem 4 is to show that there exists η small enough such that $\epsilon_{j,n} < \epsilon_d$ for all j and n. This step is based on the fact that Assumption 3 guarantees that the distributions of the rewards and the resource consumptions do not change too rapidly over the planning horizon. Proposition 2(d) gives the precise mathematical description of this fact. Thus, we can view Proposition 2(d) as a variation budget requirement for achieving single-sample polylogarithmic regret upper bound for nonstationary OLP. The complete proof of Theorem 4 (and Theorem 2) is given the following subsection.

D.1 Proof of Theorem 4

We start from the following Lemma which can be viewed as a sufficient condition for Theorem 4. Let $\epsilon_{j,n}$'s and τ_n be defined in Theorem 4.

Lemma 11. *Under Assumptions 1, 2, 3, if there exists $\eta > 0$ and $\{\iota_n\}_{n=1}^{+\infty}$ such that for all $P \in \mathcal{P}$, $\epsilon_{n-\iota-1,n} \in (0, \epsilon_d)$ when n is large enough, then*

$$\mathbb{E}[n - \tau_n] \leq 2m\left(\exp\left(\frac{\eta^2}{(16\alpha + 4\epsilon_d)^2}\right) - 1\right)^{-1} + \iota_n + 1$$

for all n large enough.

Proof. Fix $P \in \mathcal{P}$. Let N_0 be given in Proposition 2(c). Fix $n \geq N_0$. If $n \leq \iota_n + 1$, the result is trivial. Assume $n > \iota_n + 1$. By the construction of $\{\epsilon_{j,n}\}_{j=0}^{n-\iota_n-1}$, it is an increasing sequence. Thus, $0 < \epsilon_{j,n} < \epsilon_d$ for all $0 \leq j \leq n - \iota_n - 1$. In addition, the construction of $\{\epsilon_{j,n}\}_{t=0}^{T-1}$ implies that

$$\epsilon_{j,n} - \sum_{k=0}^{j-1} \max_{d \in B_{\epsilon_{k,n}}(\delta_{k,n}(d_0))} \left\| Z_n(k,d) \right\|_2 \geq \frac{\eta}{2} \quad \forall j = 0, \cdots, n - \iota_n - 1$$

Then, by (19),

$$P\left(\left\{ \sum_{j=1}^{n-\iota_n-1} \left(\epsilon_{j,n} - \sum_{k=0}^{j-1} \left\| Z_{k,n} \right\|_2 \right) \mathbb{1}\{\tau_n = j\} \geq \frac{\eta}{2} \right\} \bigcup \{\tau_n = n - \iota_n\} \right) = 1$$

which implies that

$$\mathbb{P}\left(\left\{ \epsilon_{\tau_n,n} - \sum_{j=0}^{\tau_n-1} \left\| Z_{j,n} \right\|_2 \geq \frac{\eta}{2} \right\} \bigcup \{\tau_n = n - \iota_n\} \right) = 1$$

Because $\left\| X_{\tau_n,n} \right\|_2 \geq \epsilon_{\tau_n,n}$ almost surely, then

$$\left\| M_{\tau_n,n} \right\|_2 = \left\| X_{\tau_n,n} - \sum_{j=0}^{\tau_n-1} Z_{j,n} \right\|_2 \geq \left\| X_{\tau_n,n} \right\|_2 - \left\| \sum_{j=0}^{\tau_n-1} Z_{j,n} \right\|_2 \geq \epsilon_{\tau_n,n} - \sum_{j=0}^{\tau_n-1} \left\| Z_{j,n} \right\|_2$$

Thus,

$$\mathbb{P}\left(\left\{ \left\| M_{\tau_n,n} \right\| \geq \frac{\eta}{2} \right\} \bigcup \{\tau_n = n - \iota_n\} \right) = 1 \tag{20}$$

For any $1 \leq j < n - \iota_n$, $\left\| X_{j,n} \right\|_2 \geq \epsilon_{j,n}$ if and only if $\tau_n \leq j$ because $0 < \epsilon_{j,n} < \epsilon_d$. Then, by (18), $\left\| X_{j,n} \right\|_2 \geq \epsilon_{j,n}$ implies that $M_{\tau_n,n} = M_{j,n}$. Together with (20) and Lemma 10,

$$\mathbb{P}\left\{ \left\| X_{j,n} \right\|_2 \geq \epsilon_{j,n} \right\} \leq \mathbb{P}\left\{ \left\| M_{j,n} \right\|_2 \geq \frac{\eta}{2} \right\} \leq 2m \exp\left(-\frac{\eta^2 (n - j - 1)}{(16\alpha + 4\epsilon_d)^2} \right)$$

Then,

$$\begin{aligned}
\mathbb{E}[n - \tau_n] &= \sum_{j=1}^{n-1} \mathbb{P}\{n - \tau_n \geq j\} \\
&\leq \sum_{j=1}^{n-\iota_n-1} \mathbb{P}\left\{ \left\| X_{j,n} \right\|_2 \geq \epsilon_{j,n} \right\} + \iota_n + 1 \\
&\leq \sum_{j=1}^{n-\iota_n-1} 2m \exp\left(-\frac{\eta^2 (n - j - 1)}{(16\alpha + 4\epsilon_d)^2} \right) + \iota_n + 1 \leq 2m \left(\exp\left(\frac{\eta^2}{(16\alpha + 4\epsilon_d)^2} \right) - 1 \right)^{-1} \\
&\quad + \iota_n + 1
\end{aligned}$$

□

We are now ready to give the complete proof of Theorem 4. We start by showing the existence of η and $\{\iota_n\}_{n=1}^{+\infty}$ that satisfy the hypothesis in Lemma 11. Fix $P \in \mathcal{P}$, let N_0 be given in Proposition 2(c) and fix $n \geq N_0$. Given $d \in B_{\epsilon_d}(\delta_{j,n}(d_0))$ and $0 \leq j < n$. Define

$$\tilde{p}_{j,n}(d) = \arg\min_{p \geq 0} d^T p + \frac{1}{n - j} \sum_{k=j+1}^{n} \left(\tilde{u}_{k,n} - \tilde{a}_{k,n}^T p \right)^+$$

Then, by Lemma 4

$$
\begin{aligned}
&Z_n(j,d)\\
=&\frac{1}{n-j-1}\mathbb{E}\left[(n-j)d - a_{j+1,n}\mathbb{1}\left\{u_{j+1,n} > {a_{j+1,n}}^T \tilde{p}_{j,n}(d)\right\}\right]\\
&-\bar{\delta}_{j+1,n}(p_n^*(d_0)) - \left(d - \bar{\delta}_{j,n}(p_n^*(d_0))\right)\\
=&\frac{1}{n-j-1}\mathbb{E}\left[a_{j+1,n}\left(\mathbb{1}\{u_{j+1,n} > {a_{j+1,n}}^T p_{j,n}^*(d)\} - \mathbb{1}\{u_{j+1,n} > {a_{j+1,n}}^T \tilde{p}_{j,n}(d)\}\right)\right]\\
&+\frac{1}{n-j-1}\mathbb{E}\left[(n-j)d - a_{j+1,n}\mathbb{1}\left\{u_{j+1,n} > {a_{j+1,n}}^T p_{j,n}^*(d)\right\}\right]\\
&-\bar{\delta}_{j+1,n}(p_n^*(d_0)) - \left(d - \bar{\delta}_{j,n}(p_n^*(d_0))\right)
\end{aligned}
$$

For simplicity of the notation, we define $A_{j,n}(p) \equiv a_{j,n}\mathbb{1}\{u_{j,n} > {a_{j,n}}^T p\}$. Together with the fact that $\nabla g_{j,n}^*(p_{j,n}^*(d); d) = 0$ and $p_n^*(d_0) = p_{j,n}^*(\delta_{j,n}(d_0))$,

$$
\begin{aligned}
&Z_n(j,d)\\
=&\frac{\mathbb{E}\left[A_{j+1,n}(p_{j,n}^*(d)) - A_{j+1,n}(\tilde{p}_{j,n}(d))\right]}{n-j-1}\\
&+\left(\bar{\delta}_{j+1,n}(p_{j,n}^*(d)) - \bar{\delta}_{j+1,n}(p_n^*(d_0))\right) - \left(\bar{\delta}_{j,n}(p_{j,n}^*(d)) - \bar{\delta}_{j,n}(p_n^*(d_0))\right)\\
=&\frac{\bar{\delta}_{j+1,n}(p_{j,n}^*(d)) - \bar{\delta}_{j+1,n}(p_n^*(d_0))}{n-j} - \frac{\mathbb{E}\left[A_{j+1,n}(p_{j,n}^*(d)) - A_{j+1,n}(p_n^*(d_0))\right]}{n-j}\\
&+\frac{\mathbb{E}\left[A_{j+1,n}(p_{j,n}^*(d)) - A_{j+1,n}(\tilde{p}_{j,n}(d))\right]}{n-j-1}
\end{aligned}
\tag{21}
$$

We here assume $a_{j,n}$'s have probability density function. The proof for the case that $a_{j,n}$'s have probability mass function is the same. We first upper bound the difference between the first two terms in the last line of (21). For any p_1, p_2 in Ω_p and $k \geq j+2$,

$$
\mathbb{E}[A_{k,n}(p_1) - A_{k,n}(p_2)] = \int_\Xi a \int_{a^T p_1}^{a^T p_2} f\left(\frac{k}{n}, a, s\right) ds v\left(\frac{k}{n}, a\right) da
$$

Recall $f\left(\frac{k}{n}, a, \cdot\right)$ is the conditional density of $u_{k,n}$ given $a_{k,n} = a$ and $v\left(\frac{k}{n}, a\right)$ is the density of $a_{k,n}$. Then, for any $k \geq j+2$,

$$
\begin{aligned}
&\left\|\mathbb{E}[A_{k,n}(p_{j,n}^*(d)) - A_{k,n}(p_n^*(d_0))] - \mathbb{E}[A_{j+1,n}(p_{j,n}^*(d)) - A_{j+1,n}(p_n^*(d_0))]\right\|_2\\
=&\left\| \int_\Xi a \left(v\left(\frac{k}{n}, a\right) \int_{a^T p_{j,n}^*(d)}^{a^T p_n^*(d_0)} f\left(\frac{k}{n}, a, s\right) ds \right.\right.\\
&\left.\left. - v\left(\frac{j+1}{n}, a\right) \int_{a^T p_{j,n}^*(d)}^{a^T p_n^*(d_0)} f\left(\frac{j+1}{n}, a, s\right) ds \right) da \right\|_2\\
\leq&\bar{\mu}\alpha^{m+2}\left\|p_{j,n}^*(d) - p_n^*(d_0)\right\|_2 \left(\sup_{a\in\Xi}\sup_{s\in[\underline{s},\bar{s}]} \left| f\left(\frac{k}{n}, a, s\right) - f\left(\frac{j+1}{n}, a, s\right)\right| \right.\\
&\left. + \sup_{a\in\Xi}\left| v\left(\frac{k}{n}, a\right) - v\left(\frac{j+1}{n}, a\right)\right| \right)
\end{aligned}
$$

where $\underline{s} = \inf_{a\in\Xi_a, p\in\Omega_p} a^T p$ and $\bar{s} = \sup_{a\in\Xi_a, p\in\Omega_p} a^T p$. By Lemma 4, Proposition 2(b) and by Mean Value Theorem, there exists $\tilde{d}$ in between d and $\delta_{j,n}(d_0)$ such that

$$\begin{aligned}\left\|p^*_{j,n}(d) - p^*_n(d_0)\right\|_2 &= \left\|p^*_{j,n}(d) - p^*_{j,n}(\delta_{j,n}(d_0))\right\|_2 \\ &\le \left\|G^*_{j,n}(p^*_{j,n}(\tilde{d}))^{-1}\right\|_2 \left\|d - \delta_{j,n}(d_0)\right\|_2 \le \frac{\left\|d - \delta_{j,n}(d_0)\right\|_2}{\underline{\sigma}}\end{aligned}$$

Thus,

$$\begin{aligned}&\left\|\frac{\bar{\delta}_{j+1,n}(p^*_{j,n}(d)) - \bar{\delta}_{j+1,n}(p^*_n(d_0))}{n-j} - \frac{\mathbb{E}\left[A_{j+1,n}(p^*_{j,n}(d)) - A_{j+1,n}(p^*_n(d_0))\right]}{n-j}\right\|_2 \\ \le& L_{j,n}\left\|d - \delta_{j,n}(d_0)\right\|_2\end{aligned}$$

where

$$\begin{aligned}L_{j,n} = \frac{\bar{\mu}\alpha^{m+2}}{\underline{\sigma}(n-j)(n-j-1)} \sum_{k=j+2}^{n} \Bigg(&\sup_{a\in\Xi}\sup_{s\in[\underline{s},\bar{s}]}\left|f\left(\frac{k}{n}, a, s\right) - f\left(\frac{j+1}{n}, a, s\right)\right| \\ &+ \sup_{a\in\Xi}\left|v\left(\frac{k}{n}, a\right) - v\left(\frac{j+1}{n}, a\right)\right|\Bigg)\end{aligned}$$

For the last term in (21), let ν be given in Proposition 3,

$$\begin{aligned}&\left\|\mathbb{E}\left[A_{j+1,n}(p^*_{j,n}(d)) - A_{j+1,n}(\tilde{p}_{j,n}(d))\right]\right\|_2 \\ \le&\mathbb{E}\left[\left\|\mathbb{E}\left[A_{j+1,n}(p^*_{j,n}(d)) - A_{j+1,n}(\tilde{p}_{j,n}(d))\middle|\tilde{p}_{j,n}(d)\right]\right\|_2 \mathbb{1}\{\left\|\tilde{p}_{j,n}(d) - p^*_{j,n}(d)\right\|_2 \le \nu\}\right] \\ &+ \alpha\mathbb{P}\left\{\left\|\tilde{p}_{j,n}(d) - p^*_{j,n}(d)\right\|_2 > \nu\right\}\end{aligned}$$

By Proposition 3,

$$\alpha\mathbb{P}\left\{\left\|\tilde{p}_{j,n}(d) - p^*_{j,n}(d)\right\|_2 > \nu\right\} \le \alpha C_1 \exp\left(-C_2\nu^2(n-j)\right)$$

Because $(u_{j+1,n}, a_{j+1,n})$ and $\tilde{p}_{j,n}(d)$ are independent by Assumption 2, Proposition 2(b) and Mean Value Theorem imply that,

$$\begin{aligned}&\left\|\mathbb{E}\left[a_{j+1,n}\left(\mathbb{1}\{u_{j+1,n} > {a_{j+1,n}}^T p^*_{j,n}(d)\} - \mathbb{1}\{u_{j+1,n} > {a_{j+1,n}}^T \tilde{p}_{j,n}(d)\}\right)\middle|\tilde{p}_{j,n}(d)\right]\right\|_2 \cdot \\ &\qquad\qquad \mathbb{1}\left\{\left\|\tilde{p}_{j,n}(d) - p^*_{j,n}(d)\right\|_2 \le \nu\right\} \\ \le&\bar{\sigma}\left\|p^*_{j,n}(d) - \tilde{p}_{j,n}(d)\right\|_2\end{aligned}$$

Together with Theorem 1, there exists a constant C_3 that holds for all $P \in \mathcal{P}$ and does not depend on j, n and d such that

$$\left\|\mathbb{E}\left[A_{j+1,n}(p^*_{j,n}(d)) - A_{j+1,n}(\tilde{p}_{j,n}(d))\right]\right\|_2 \le C_3\sqrt{\frac{\log n}{n-j-1}}$$

for all $n \ge N_0$. Define $e_{0,n} = \eta$, and

$$e_{j+1,n} = e_{j,n} + L_{j,n}e_{j,n} + C_3\sqrt{\log n}(n-j-1)^{-\frac{3}{2}} \quad 0 \le j \le n-2$$

Then, by Proposition 2(d),

$$\begin{aligned} e_{n-\iota-1,n} =& \eta \prod_{j=0}^{n-\iota-2}(1+L_{j,n}) + C_3\sqrt{\log n}\sum_{j=1}^{n-\iota-1}\left(\prod_{q=j}^{n-\iota-2}(1+L_{q,n})(n-j)^{-\frac{3}{2}}\right) \\ \leq& \eta e^{\frac{\mu\bar{L}\alpha^{m+2}}{\underline{\sigma}}} + C_3 e^{\frac{\mu\bar{L}\alpha^{m+2}}{\underline{\sigma}}}\sqrt{\log n}\sum_{j=\iota+1}^{n-1} j^{-\frac{3}{2}} \end{aligned}$$

In addition,

$$\sum_{j=\iota+1}^{T-1} j^{-\frac{3}{2}} \leq \int_{\iota}^{+\infty} s^{-\frac{3}{2}} ds = \frac{2}{\sqrt{\iota}}$$

Thus, there exists $\eta > 0$ and constants C_4 hold for all $P \in \mathcal{P}$ such that, when $n \geq N_0$, $\iota_n \equiv \lceil C_4 \log n \rceil \geq \left\lceil \frac{\alpha}{\underline{d}} \right\rceil$ and $e_{n-\iota_n-1,n}(\eta) < \epsilon_d$. In addition, $\epsilon_{0,n}(\eta) = e_{0,n}$. Suppose that $\epsilon_{j,n} \leq e_{j,n}$, then

$$\max_{d \in B_{\epsilon_{j,n}}(\delta_{j,n}(d_0))} \left\|Z(j,d)\right\|_2 \leq L_{j,n}e_{j,n} + C_3(n-j-1)^{-\frac{3}{2}}$$

which implies that $\epsilon_{j+1,n} \leq e_{j+1,n}$. By induction, $\epsilon_{n-\iota_n-1,n} < \epsilon_d$. Thus, the hypothesis in Lemma 11 is satisfied by η and ι_n's, and τ_n defined by η and ι_n satisfies (17). Therefore, $\tau_n \leq \bar{\tau}_n$ almost surely, and

$$\mathbb{E}[n-\bar{\tau}_n] \leq \mathbb{E}[n-\tau_n] \leq 2m\left(\exp\left(\frac{\eta^2}{(16\alpha+4\epsilon_d)^2}\right)-1\right)^{-1} + \lceil C_4\log n\rceil + 1 \equiv C_{Res}\log n$$

when n is large enough. The proof is complete.

D.2 Proof of Lemma 9

We first provide a Lemma that shows a uniform Lipschitz continuous property of the population price.

Lemma 12. *Under Assumptions 1, 3, there exists L_d, $\epsilon_d > 0$ such that for all $P \in \mathcal{P}$, when n is large enough,*

$$\left\|p^*_{j,n}(d) - p^*_{j,n}(\delta_{j,n}(d_0))\right\|_2 \leq L_d\left\|d - \delta_{j,n}(d_0)\right\|_2 \quad \forall d \in B_{\epsilon_d}(\delta_{j,n}(d_0))\ \forall 0 \leq j < n$$

Proof. Fix $P \in \mathcal{P}$. Let N_0 be given by Proposition 2(c) and fix $n \geq N_0$. Proposition 2(c) and Lemma 4(a) together show that $\delta_{j,n}(d_0) > \frac{1-\kappa}{2}\epsilon_p \cdot e$ for all j. Select $\epsilon_d \in (0, \min\{\frac{1-\kappa}{4}\epsilon_p, \frac{(1-\kappa)\underline{\sigma}\epsilon_p\epsilon^2}{64\bar{u}}\})$. Fix d such that $\left\|d - \delta_{j,n}(d_0)\right\|_2 \leq \epsilon_d$

$$\begin{aligned} 0 \leq& g^*_{j,n}(p^*_{j,n}(d), \delta_{j,n}(d_0)) - g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), \delta_{j,n}(d_0)) \\ =& \left(g^*_{j,n}(p^*_{j,n}(d), \delta_{j,n}(d_0)) - g^*_{j,n}(p^*_{j,n}(d), d)\right) + \left(g^*_{j,n}(p^*_{j,n}(d), d) - g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), d)\right) \\ &+ \left(g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), d) - g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), \delta_{j,n}(d_0))\right) \\ \leq& \left|g^*_{j,n}(p^*_{j,n}(d), \delta_{j,n}(d_0)) - g^*_{j,n}(p^*_{j,n}(d), d)\right| \\ &+ \left|g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), d) - g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), \delta_{j,n}(d_0))\right| \\ \leq& \frac{8\bar{u}}{(1-\kappa)\epsilon_p}\left\|d - \delta_{j,n}(d_0)\right\| \leq \frac{\underline{\sigma}\epsilon^2}{8} \end{aligned} \tag{22}$$

where the last line is by Lemma 2. By Proposition 2(b), if $\left\|p - p^*_{j,n}(\delta_{j,n}(d_0))\right\|_2 = \left\|p - p^*_n(d_0)\right\|_2 \in (\frac{\epsilon}{2}, \epsilon)$, then for all $0 \leq j < n$,

$$g^*_{j,n}(p, \delta_{j,n}(d_0)) - g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), \delta_{j,n}(d_0)) \geq \frac{\underline{\sigma}}{2}\left\|p - p^*_{j,n}(\delta_{j,n}(d_0))\right\|_2^2 > \frac{\underline{\sigma}\epsilon^2}{8}$$

Together with the level set of a convex function is convex,

$$\left\{p : g^*_{j,n}(p, \delta_{j,n}(d_0)) \le g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), \delta_{j,n}(d_0)) + \frac{\underline{\sigma}\epsilon^2}{8}\right\} \subseteq \Omega_p$$

Then, (22) shows that $p^*_{j,n}(d) \in \Omega_p$. By strong convexity again,

$$g^*_{j,n}(p^*_{j,n}(d), \delta_{j,n}(d_0)) - g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), \delta_{j,n}(d_0)) \ge \frac{\underline{\sigma}}{2}\left\|p^*_{j,n}(d) - p^*_{j,n}(\delta_{j,n}(d_0))\right\|_2^2$$
$$g^*_{j,n}(p^*_{j,n}(\delta_{j,n}(d_0)), d) - g^*_{j,n}(p^*_{j,n}(d), d) \ge \frac{\underline{\sigma}}{2}\left\|p^*_{j,n}(d) - p^*_{j,n}(\delta_{j,n}(d_0))\right\|_2^2$$

Then, adding up the two inequalities

$$\begin{aligned}&\underline{\sigma}\left\|p^*_{j,n}(d) - p^*_{j,n}(\delta_{j,n}(d_0))\right\|_2^2\\ \le&(\delta_{j,n}(d_0) - d)^T(p^*_{j,n}(d) - p^*_{j,n}(\delta_{j,n}(d_0))) \le \left\|d - \delta_{j,n}(d_0)\right\|_2\left\|p^*_{j,n}(d) - p^*_{j,n}(\delta_{j,n}(d_0))\right\|_2\end{aligned}$$

Thus,

$$\left\|p^*_{j,n}(d) - p^*_{j,n}(\delta_{j,n}(d_0))\right\|_2 \le \frac{1}{\underline{\sigma}}\left\|d - \delta_{j,n}(d_0)\right\|_2$$

□

To prove Lemma 9, Fix $P \in \mathcal{P}$. Let N_0 be given by Proposition 2(c) and fix $n \ge N_0$. We can assume ϵ_d given in Lemma 12 is small enough such that $L_d \cdot \epsilon_d \le \frac{\epsilon}{2}$. Then, Lemma 12 implies that, for any $d \in B_{\epsilon_d}(\delta_{j,n}(d_0))$ and $0 \le j < n$, $p^*_{j,n}(d) \in B_{\frac{\epsilon}{2}}(p^*_n(d_0)) \subseteq \Omega_p$. Then, by Proposition 2(a) and (b), $\nabla g^*_{j,n}(p^*_{j,n}(d), d) = 0$. Thus, $d = \bar{\delta}_{j,n}(p^*_{j,n}(d))$. Finally, by definition of $\Omega_{j,n}$, $\bar{\delta}_{j,n}(p^*_{j,n}(d)) \in \Omega_{j,n}$. Thus, $d \in \Omega_{j,n}$. The proof is complete.

E Proof of Proposition 1

Define

$$R^{off}_{j,n}(d) = \max\left\{\sum_{k=j+1}^{n} u_{k,n}x_{k,n} : \sum_{k=j+1}^{n} a_{k,n}x_{k,n} \le (n-j)d,\ x_{k,n} \in [0,1]\ \forall j < k \le n\right\}, \tag{23}$$

$R^{off}_{j,n}(d)$ is the optimal total reward earned if we allocate $(n-j)d$ resources to the last $n-j$ orders in an offline fashion. $p_{j,n}(d)$ is then an optimal dual solution corresponding to the resource constraint of (23). In addition, define

$$R^{\pi}_{j,n} = \sum_{k=j+1}^{n} u_{k,n}x^{\pi}_{k,n},$$

which is the total reward earned by Algorithm 1 on the last $n-j$ orders. Fix $P \in \mathcal{P}$, recall $d_{j,n} = \frac{b_{j,n}}{n-j}$ which is the average remaining resource under Algorithm 1.

$$\begin{aligned}&Regret_n(\pi, P)\\ =&\mathbb{E}\left[R^{off}_{0,n}(d_0) - R^{\pi}_{0,n}\right]\\ =&\mathbb{E}\left[\max_{x_{1,n}\in[0,1]\ a_{1,n}x_{1,n}\le nd_0} u_{1,n}x_{1,n} + R^{off}_{1,n}\left(\frac{nd_0 - a_{1,n}x_{1,n}}{n-1}\right) - u_{1,n}x^{\pi}_{1,n} - R^{\pi}_{1,n}\right]\\ =&\mathbb{E}\left[\max_{x_{1,n}\in[0,1]\ a_{1,n}x_{1,n}\le nd_0} u_{1,n}x_{1,n} + R^{off}_{1,n}\left(\frac{nd_0 - a_{1,n}x_{1,n}}{n-1}\right) - R^{off}_{1,n}(d_{1,n}) - u_{1,n}x^{\pi}_{1,n}\right]\\ &+\mathbb{E}\left[R^{off}_{1,n}(d_{1,n}) - R^{\pi}_{1,n}\right]\end{aligned}$$

By LP Duality Theory, $R^{off}_{j,n}(d) = \min_{p\geq 0}\ (n-j)\cdot g_{j,n}(p,d)$. Since $p_{1,n}(d_{1,n})$ is a feasible solution for evaluating $R^{off}_{1,n}\left(\frac{nd_0 - a_{1,n}x_{1,n}}{n-1}\right)$ for any $x_{1,n}$,

$$\begin{aligned} R^{off}_{1,n}\left(\frac{nd_0 - a_{1,n}x_{1,n}}{n-1}\right) - R^{off}_{1:n}(d_{1,n}) &\leq \left(nd_0 - a_{1,n}x_{1,n} - (n-1)d_{1,n}\right)^T p_{1,n}(d_{1,n}) \\ &= (x^{\pi}_{1,n} - x_{1,n}){a_{1,n}}^T p_{1,n}(d_{1,n}) \end{aligned}$$

Then,

$$\begin{aligned} Regret_n(\pi, P) \leq& \mathbb{E}\left[\max_{x_{1,n}\in[0,1]}\left(u_{1,n} - {a_{1,n}}^T p_{1,n}(d_{1,n})\right)\left(x_{1,n} - x^{\pi}_{1,n}\right)\right] \\ &+ \mathbb{E}\left[R^{off}_{1,n}(d_{1,n}) - R_{1,n}(\pi)\right] \\ =& \mathbb{E}\left[\left(u_{1,n} - {a_{1,n}}^T p_{1,n}(d_{1,n})\right)\left(\mathbb{1}\left\{u_{1,n} > {a_{1,n}}^T p_{1,n}(d_{1,n})\right\} - x^{\pi}_{1,n}\right)\right] \\ &+ \mathbb{E}\left[R^{off}_{1,n}(d_{1,n}) - R^{\pi}_{1,n}\right] \end{aligned}$$

Repeat the above argument for each j, we have the following general regret upper bound. Let $\bar{\tau}_n$ be defined in Section D.

$$\begin{aligned} Regret_n(\pi, P) \leq& \sum_{j=1}^{n-1}\mathbb{E}\left[\left(u_{j,n} - {a_{j,n}}^T p_{j,n}(d_{j,n})\right)\left(\mathbb{1}\left\{u_{j,n} > {a_{j,n}}^T p_{j,n}(d_{j,n})\right\} - x^{\pi}_{j,n}\right)\right] \\ &+ \mathbb{E}\left[R^{off}_{n-1,n}(d_{n-1,n}) - R^{\pi}_{n-1,n}\right] \\ \leq& \sum_{j=1}^{n-1}\mathbb{E}\left[\left(u_{j,n} - {a_{j,n}}^T p_{j,n}(d_{j,n})\right)\left(\mathbb{1}\left\{u_{j,n} > {a_{j,n}}^T p_{j,n}(d_{j,n})\right\}\right.\right. \\ &\qquad\left.\left. - \mathbb{1}\{u_{j,n} > {a_{j,n}}^T p^{\pi}_{j-1,n}\}\right)\mathbb{1}\{j < \bar{\tau}_n\}\right] \\ &+ \left(\bar{u} + \frac{m\alpha\bar{u}}{\underline{d}}\right)\mathbb{E}[n - \bar{\tau}_n + 1] + 2\bar{u} \end{aligned}$$

where the last line is by Lemma 5.

F Proof of Lemma 1

We start from a Lemma that characterize the difference between $p_{j,n}(d_{j,n})$ and $p^*_{j-1,n}(d_{j-1,n})$.

Lemma 13. *Under Assumptions 1, 3, for any $P \in \mathcal{P}$, when n is large enough,*

$$\begin{aligned} &\left\|p_{j,n}(d_{j,n}) - p^*_{j-1,n}(d_{j-1,n})\right\|_2 \mathbb{1}\{j < \bar{\tau}_n\} \\ \leq& \left(\sup_{d\in\Omega_{j,n}}\left\|p_{j,n}(d) - p^*_{j,n}(d)\right\|_2 + \frac{2\alpha}{\underline{\sigma}(n-j)}\right)\mathbb{1}\{j < \bar{\tau}_n\} \end{aligned}$$

Proof. Fix $P \in \mathcal{P}$. Let N_0 be given in Proposition 2(c), and fix $n \geq N_0$ and $1 \leq j \leq n-1$. Then, define $EA_{j,n}(p) = \mathbb{E}\left[a_{j,n}\mathbb{1}\{u_{j,n} > {a_{j,n}}^T p\}\right]$, and by Lemma 4,

$$\begin{aligned} &p^*_{j-1,n}(d_{j-1,n})\mathbb{1}\{j < \bar{\tau}_n\} \\ =& p^*_{j,n}\left(\frac{(n-j+1)d_{j-1,n} - EA_{j,n}(p^*_{j-1,n}(d_{j-1,n}))}{n-j} - d_{j,n} + d_{j,n}\right)\mathbb{1}\{j < \bar{\tau}_n\} \end{aligned}$$

Since on the event $\{j < \bar{\tau}_n\}$

$$\begin{aligned}
&\left\| \frac{(n-j+1)d_{j-1,n} - EA_{j,n}(p^*_{j-1,n}(d_{j-1,n}))}{n-j} - d_{j,n} \right\|_2 \\
\leq &\left\| \frac{a_{j,n}\mathbb{1}\{u_{j,n} > {a_{j,n}}^T p^\pi_{j-1,n}\} - A_{j,n}(p^*_{j-1,n}(d_{j-1,n}))}{n-j} \right\|_2 \leq \frac{2\alpha}{n-j}
\end{aligned}$$

Lemma 4 and Mean Value Theorem implies that

$$\left\| p^*_{j,n}(d_{j,n}) - p^*_{j-1,n}(d_{j-1,n}) \right\|_2 \mathbb{1}\{j < \bar{\tau}_n\} \leq \frac{2\alpha}{\underline{\sigma}(n-j)} \mathbb{1}\{j < \bar{\tau}_n\}$$

Then, by triangular inequality,

$$\begin{aligned}
&\left\| p_{j,n}(d_{j,n}) - p^*_{j-1,n}(d_{j-1,n}) \right\|_2 \mathbb{1}\{j < \bar{\tau}_n\} \\
&\leq \left(\sup_{d \in \Omega_{j,n}} \left\| p_{j,n}(d) - p^*_{j,n}(d) \right\|_2 + \frac{2\alpha}{\underline{\sigma}(n-j)} \right) \mathbb{1}\{j < \bar{\tau}_n\}
\end{aligned}$$

□

We can now give the complete proof of Lemma 1. Fix $P \in \mathcal{P}$. Let N_0 be given in Proposition 2(c), and fix $n \geq N_0$ and $1 \leq j \leq n-1$. Define

$$\begin{aligned}
\xi_{j,n,1} &= \left(\sup_{d \in \Omega_{j,n}} \left\| p_{j,n}(d) - p^*_{j,n}(d) \right\|_2 + \frac{2\alpha}{\underline{\sigma}(n-j)} \right) \cdot e \\
\xi_{j-1,n,2} &= \left(\sup_{d \in \Omega_{j-1,n}} \left\| \tilde{p}_{j-1,n}(d) - p^*_{j-1,n}(d) \right\|_2 \right) \cdot e
\end{aligned}$$

where e is the vector with all entries equal to 1 and $\tilde{p}_{j,n}(d)$ is defined the same as in the proof of Theorem 4. Then, by Lemma 13, on the event $\{j < \bar{\tau}_n\}$

$$\begin{aligned}
{a_{j,n}}^T p_{j,n}(d_{j,n}) =& {a_{j,n}}^T p^*_{j-1,n}(d_{j-1,n}) + {a_{j,n}}^T \left(p_{j,n}(d_{j,n}) - {a_{j,n}}^T p^*_{j-1,n}(d_{j-1,n}) \right) \\
\geq& {a_{j,n}}^T p^*_{j-1,n}(d_{j-1,n}) - {a_{j,n}}^T \xi_{j,n,1} = {a_{j,n}}^T \left(p^*_{j-1,n}(d_{j-1,n}) - \xi_{j,n,1} \right)
\end{aligned}$$

In addition, on the event $\{j < \bar{\tau}_n\}$

$$\begin{aligned}
{a_{j,n}}^T p^\pi_{j-1,n} =& {a_{j,n}}^T p^*_{j-1,n}(d_{j-1,n}) + {a_{j,n}}^T \left(p^\pi_{j-1,n} - {a_{j,n}}^T p^*_{j-1,n}(d_{j-1,n}) \right) \\
\leq& {a_{j,n}}^T p^*_{j-1,n}(d_{j-1,n}) + {a_{j,n}}^T \xi_{j,n,2} = {a_{j,n}}^T \left(p^*_{j-1,n}(d_{j-1,n}) + \xi_{j-1,n,2} \right)
\end{aligned}$$

For simplicity of the notation, define $\Delta_{j,n}(p) = u_{j,n} - {a_{j,n}}^T p$. Then,

$$\begin{aligned}
&\mathbb{E}\Big[\Delta_{j,n}(p_{j,n}(d_{j,n})) \left(\mathbb{1}\{\Delta_{j,n}(p_{j,n}(d_{j,n})) > 0\} \right. \\
&\qquad \left. - \mathbb{1}\{\Delta_{j,n}(p^\pi_{j-1,n}) > 0\} \right) \mathbb{1}\{j < \bar{\tau}_n\} \Big] \\
=&\mathbb{E}\Big[\Delta_{j,n}(p_{j,n}(d_{j,n})) \left(\mathbb{1}\{\Delta_{j,n}(p_{j,n}(d_{j,n})) > 0\} \right. \\
&\qquad \left. - \mathbb{1}\{\Delta_{j,n}(p^\pi_{j-1,n}) > 0\} \right) \mathbb{1}\{\Delta_{j,n}(p_{j,n}(d_{j,n})) > 0\} \mathbb{1}\{j < \bar{\tau}_n\} \Big] \\
&+ \mathbb{E}\Big[\Delta_{j,n}(p_{j,n}(d_{j,n})) \left(\mathbb{1}\{\Delta_{j,n}(p_{j,n}(d_{j,n})) > 0\} \right. \\
&\qquad \left. - \mathbb{1}\{\Delta_{j,n}(p^\pi_{j-1,n}) > 0\} \right) \mathbb{1}\{\Delta_{j,n}(p_{j,n}(d_{j,n})) < 0\} \mathbb{1}\{j < \bar{\tau}_n\} \Big]
\end{aligned} \tag{24}$$

For the first term on the right-hand-side of (24),

$$\begin{aligned}
&\mathbb{E}\Big[\Delta_{j,n}(p_{j,n}(d_{j,n}))\left(\mathbb{1}\{\Delta_{j,n}(p_{j,n}(d_{j,n}))>0\}\right.\\
&\qquad\left.-\mathbb{1}\{\Delta_{j,n}(p^{\pi}_{j-1,n})>0\}\right)\mathbb{1}\{\Delta_{j,n}(p_{j,n}(d_{j,n}))>0\}\mathbb{1}\{j<\bar{\tau}_n\}\Big]\\
\leq&\mathbb{E}\left[\Delta_{j,n}(p^*_{j-1,n}(d_{j-1,n})-\xi_{j,n,1})\left(\mathbb{1}\{\Delta_{j,n}(p^*_{j-1,n}(d_{j-1,n})-\xi_{j,n,1})>0\}\right.\right.\\
&\qquad\left.\left.-\mathbb{1}\{\Delta_{j,n}(p^*_{j-1,n}(d_{j-1,n})+\xi_{j-1,n,2})>0\}\right)\mathbb{1}\{j<\bar{\tau}_n\}\right]\\
\leq&\mathbb{E}\left[{a_{j,n}}^T(\xi_{j,n,1}+\xi_{j,n,2})\left(\mathbb{1}\{\Delta_{j,n}(p^*_{j-1,n}(d_{j-1,n})-\xi_{j,n,1})>0\}\right.\right.\\
&\qquad\left.\left.-\mathbb{1}\{\Delta_{j,n}(p^*_{j-1,n}(d_{j-1,n})+\xi_{j-1,n,2})>0\}\right)\mathbb{1}\{j\leq\bar{\tau}_n\}\right]\\
\leq&\mathbb{E}\left[\left\|\xi_{j,n,1}+\xi_{j-1,n,2}\right\|_2\left\|\mathbb{E}\left[DA_{j,n}\middle|d_{j-1,n},\xi_{j,n,1},\xi_{j-1,n,2}\right]\mathbb{1}\{j\leq\bar{\tau}_n\}\right\|_2\right]
\end{aligned}$$

where

$$DA_{j,n}=A_{j,n}(p^*_{j-1,n}(d_{j-1,n})-\xi_{j,n,1})-A_{j,n}(p^*_{j-1,n}(d_{j-1,n})+\xi_{j-1,n,2})$$

and $A_{j,n}(p)$ is defined the same as the proof of Theorem 4. Let ν be given in Proposition 3. We can assume ν is small enough and that there exits $\tilde{\iota}$ not depending on P, n and j is large enough such that $\sup_{d\in\Omega_{j,n}}\left\|p_{j,n}(d)-p^*_{j,n}(d)\right\|_2\leq\nu$ and $\sup_{d\in\Omega_{j-1,n}}\left\|\tilde{p}_{j-1,n}(d)-p^*_{j-1,n}(d)\right\|_2\leq\nu$ imply that both $p^*_{j-1,n}(d_{j-1,n})-\xi_{j,n,1})$ and $p^*_{j-1,n}(d_{j-1,n})+\xi_{j-1,n,2})$ are in Ω_p when $j\leq\bar{\tau}_n\wedge n-\tilde{\iota}$. Let

$$\mathcal{E}_{j,n}=\left\{\sup_{d\in\Omega_{j,n}}\left\|p_{j,n}(d)-p^*_{j,n}(d)\right\|_2\leq\nu\right\}\cap\left\{\sup_{d\in\Omega_{j-1,n}}\left\|\tilde{p}_{j-1,n}(d)-p^*_{j-1,n}(d)\right\|_2\leq\nu\right\}$$

Then, by Proposition 2(b) and Mean Value Theorem,

$$\left\|\mathbb{E}\left[DA_{j,n}\middle|d_{j-1,n},\xi_{j,n,1},\xi_{j-1,n,2}\right]\mathbb{1}_{\mathcal{E}_{j,n}}\mathbb{1}\{j\leq\bar{\tau}_n\wedge n-\tilde{\iota}\}\right\|_2\leq\bar{\sigma}\left\|\xi_{j,n,1}+\xi_{j-1,n,2}\right\|_2$$

Thus, if $j<n-\tilde{\iota}$, by Theorem 1

$$\begin{aligned}
&\mathbb{E}\left[\left\|\xi_{j,n,1}+\xi_{j-1,n,2}\right\|_2\left\|\mathbb{E}\left[DA_{j,n}\middle|d_{j-1,n},\xi_{j,n,1},\xi_{j-1,n,2}\right]\mathbb{1}_{\mathcal{E}_{j,n}}\mathbb{1}\{j\leq\bar{\tau}_n\}\right\|_2\right]\\
\leq&\mathbb{E}\left[\left\|\xi_{j,n,1}+\xi_{j-1,n,2}\right\|_2^2\right]\leq\frac{C_4\log n}{n-j}
\end{aligned}$$

where C_4 holds for all $P\in\mathcal{P}$, n and j. In addition, Proposition 3, Lemma 5 and Assumption 3 show that

$$\begin{aligned}
&\mathbb{E}\left[\left\|\xi_{j,n,1}+\xi_{j-1,n,2}\right\|_2\left\|\mathbb{E}\left[DA_{j,n}\middle|d_{j-1,n},\xi_{j,n,1},\xi_{j-1,n,2}\right]\mathbb{1}_{\mathcal{E}^c_{j,n}}\mathbb{1}\{j\leq\bar{\tau}_n\wedge n-\tilde{\iota}\}\right\|_2\right]\\
=&o\left(\frac{1}{n-j}\right)
\end{aligned}$$

where the little-o term hold uniformly over $P\in\mathcal{P}$ and n. Then, we can assume $\tilde{\iota}$ is large enough such that there exists a constant C_{Myopic} holds for all all $P\in\mathcal{P}$, n and j such that when $j<n-\tilde{\iota}$

$$\mathbb{E}\Big[\Delta_{j,n}(p_{j,n}(d_{j,n}))\left(\mathbb{1}\{\Delta_{j,n}(p_{j,n}(d_{j,n}))>0\}\right.\\
\left.-\mathbb{1}\{\Delta_{j,n}(p^{\pi}_{j-1,n})>0\}\right)\mathbb{1}\{\Delta_{j,n}(p_{j,n}(d_{j,n}))>0\}\mathbb{1}\{j<\bar{\tau}_n\}\Big]\leq\frac{C_{Myopic}\log n}{n-j}$$

Same analysis can be applied to the second term on the right-hand-side of (24). Thus, the proof is complete.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: See Section 1

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See Section 5

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: See Section 2 and Appendix

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: See Section 4

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: See supplementary material for code

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: See Section 4

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: See Section 4

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: See supplementary material.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: After reviewing NeurIPS Code of Ethics, our answer is Yes.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This paper focuses on theoretical analysis of an online optimization problem and does not have direct societal impacts.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This paper does not release high-risk models or datasets that require safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper does not introduce new assets. The supplementary code is provided only to reproduce the numerical experiments.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.