Wait-Less Offline Tuning and Re-solving for Online Decision Making

Jingruo Sun [1] **Wenzhi Gao** [2] **Ellen Vitercik** [1 3] **Yinyu Ye** [1 4 5]

Abstract

Online linear programming (OLP) has found broad applications in revenue management and resource allocation. State-of-the-art OLP algorithms achieve low regret by repeatedly solving linear programming (LP) subproblems that incorporate updated resource information. However, LP-based methods are computationally expensive and often inefficient for large-scale applications. By contrast, recent first-order OLP algorithms are more computationally efficient but typically suffer from weaker regret guarantees. To address these shortcomings, we propose a new algorithm that combines the strengths of LP-based and first-order OLP algorithms. Our algorithm re-solves the LP subproblems periodically at a predefined frequency f and uses the latest dual prices to guide online decision-making. In parallel, a first-order method runs during each interval between LP re-solves and smooths resource consumption. Our algorithm achieves $\mathcal{O}(\log(T/f) + \sqrt{f})$ regret and delivers a "wait-less" online decision-making process that balances computational efficiency and regret guarantees. Extensive experiments demonstrate at least 10-fold improvements in regret over first-order methods and 100-fold improvements in runtime over LP-based methods.

1. Introduction

Sequential decision-making has garnered significant attention for its utility in guiding optimal strategies in dynamic environments. The goal is to identify effective decisions and policies in environments where knowledge of the system continuously accumulates and evolves. Online Linear Programming (OLP) (Agrawal et al., 2014) offers a powerful framework that encapsulates the core principles of sequential decision-making and has been extensively applied to different domains, including resource allocation (Balseiro et al., 2022a), online advertising (Mehta et al., 2007), and inventory management (Talluri et al., 2004).

We study an OLP problem where customers arrive sequentially, each requesting a combination of resources and offering a bidding price. The objective is to determine which resource requests to fulfill to maximize revenue while respecting resource constraints. The challenge is that decisions must be made immediately and irrevocably, relying solely on historical data without knowledge of future arrivals. The goal is to minimize regret with respect to the optimal hindsight linear programming (LP) solution.

To guide real-time decision-making, state-of-the-art OLP algorithms estimate optimal *dual prices* and use them to make decisions. These algorithms fall into two main categories: LP-based and first-order methods. Specifically, LP-based methods update dual prices by repeatedly solving linear programs at each time step with all available information so far. However, the substantial computational demands limit their application in time-sensitive settings. First-order methods offer quick, incremental updates to dual prices using gradient information, but generally fall short of achieving the strong regret guarantee of LP-based methods. These trade-offs motivate an open question at the intersection of online learning and decision-making:

Can we simultaneously achieve

low regret and computational efficiency?

Our contributions. We answer this question in the affirmative. We summarize our contributions as follows:

- *Parallel Multi-Stage Framework:* We separate online learning and decision-making into distinct processes that interact via a feedback loop. We re-solve the LP subproblems periodically at a fixed frequency and feed the updated dual price to guide decision-making until the next time that the LP is re-solved. To further enhance efficiency, we apply a first-order method during the initial and final stages, restarting with the most recent learn-

[1]Department of Management Science & Engineering, Stanford University [2]Institute for Computational and Mathematical Engineering, Stanford University [3]Department of Computer Science, Stanford University [4]Chinese University of Hong Kong (Shen Zhen) [5]Hong Kong University of Science and Technology. Correspondence to: Jingruo Sun <jingruo@stanford.edu>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

ing outcomes. By integrating the LP-based and first-order techniques, our approach leverages their respective strengths and achieves a balance between regret performance and computational costs.

- *Regret Analysis:* We unify the analysis of LP-based and first-order methods by deriving a new performance metric to account for their inter-dependency. Our analysis yields a "spectrum theorem" that bounds regret for any feasible choice of the re-solving frequency:

  Theorem 1.1 (Informal version of Theorem 3.2)**.** *If we re-solve the LPs every f time steps within horizon T and apply a first-order method in the initial and final f steps, we achieve a worst-case regret of $\mathcal{O}\big(\log(T/f) + \sqrt{f}\big)$.*

  In particular, $f = 1$ yields a pure LP-based method with $\mathcal{O}(\log T)$ regret, while $f = T$ reduces to a pure first-order method with $\mathcal{O}(\sqrt{T})$ regret. By choosing an "optimal" f, one can achieve the best possible regret based on available computational resources and enable a "wait-less" decision-making system across all time steps.

- *Experiments:* Through experiments across diverse distributions, we demonstrate that our algorithm achieves over a 10-fold improvement in regret compared to first-order methods and over a 100-fold improvement in runtime with comparable regret to LP-based methods.

Key Challenges. OLP methods face a fundamental challenge in balancing efficient real-time decision-making and accurate dual-price learning. Specifically, LP-based methods ensure high-quality decisions with $\mathcal{O}(\log T)$ regret, but their computational costs increase quadratically with problem size. A natural alternative—batching customers and solving the LP once every f arrivals—reduces this cost, but leaves customers waiting until the batch concludes. In contrast, first-order methods update dual prices via gradient information, allowing for faster computation. However, these methods require small step sizes to maintain decision quality, which slows their adaptation to new data. Even using different step sizes for learning and decision-making only achieves $\mathcal{O}(T^{1/3})$ regret when the distribution of customers has continuous support. Our parallel framework addresses these issues: an LP-based method periodically refines dual prices, while a first-order method immediately processes arriving customers using the most recent dual updates, thereby eliminating delays.

A second challenge emerges when analyzing the regret of this hybrid approach. LP-based methods rely on a stopping-time analysis (halting when resources are depleted), whereas first-order methods track constraint violations as part of the regret. Since these two formulations are fundamentally different, existing regret bounds cannot be naively combined. We overcome this challenge by introducing a unified performance metric that decomposes regret into three components: dual convergence, constraint violation, and leftover resources. This approach yields the first integrated analysis of LP-based and first-order OLP, culminating in our $\mathcal{O}(\log(T/f) + \sqrt{f})$ regret bound.

1.1. Related Literature

Online resource allocation and OLP problems have been widely studied under two predominant models: the stochastic input (Goel & Mehta, 2008; Devanur et al., 2019) and stochastic permutation models (Agrawal et al., 2014; Gupta & Molinaro, 2016). We study the former, where each customer's resource request and bidding price are drawn i.i.d. from an unknown distribution. We use expected regret and constraint violation as the performance metric. We summarize some recent works in Table 1.

LP-based Methods. These methods solve the OLP dual repeatedly to update dual prices and make decisions. Early approaches enforced a fixed average resource constraint (Agrawal et al., 2014), whereas recent work dynamically tracks remaining resources and enables LP-based methods to achieve $\mathcal{O}(\log T)$ regret under continuous support and non-degeneracy assumptions (Li & Ye, 2022). Variants include multi-secretary problems (Bray, 2022), regularized resource constraints (Ma et al., 2024), and finite-support distributions yielding constant regret (Chen et al., 2024).

First-order Methods. These methods generate decisions using gradient updates without solving LPs, enabling efficient computation. They achieve $\mathcal{O}(\sqrt{T})$ regret with mirror descent (Li et al., 2020; Balseiro et al., 2022a) and $\mathcal{O}(T^{3/8})$ under finite support and non-degeneracy assumptions (Sun et al., 2020). Variants include proximal updates (Gao et al., 2023), momentum-based mirror descent (Balseiro et al., 2022b), resource adjustments (Ma et al., 2024), and restart strategies yielding $\mathcal{O}(T^{1/3})$ regret (Gao et al., 2024).

Delay in Decision-making. Delays arise from the time-consuming process of solving the large-scale, up-to-date LP subproblems for each customer. Golrezaei & Yao (2021) study a mix of impatient and partially patient customers, while Xie et al. (2023) show that batching requests can reduce regret. Concurrent work (Xu et al., 2024) reduces delay by solving LPs in batches but assumes a lower bound on resource requests and still requires waiting in initial and final batches. To achieve delay-free decisions, we shift the re-solving process offline and only fine-tune the solution online. Compared with previous works, our framework imposes standard assumptions, achieves $\mathcal{O}(\log(T/f)+\sqrt{f})$ regret, and ensures "wait-less" decision-making throughout the entire horizon.

Table 1: Performances of Dual Algorithms in Recent OLP Literature

Paper	Setting	Algorithm	Regret	Decision-Making
(Li & Ye, 2022)	Bounded, continuous support, non-degeneracy	LP-based	$\mathcal{O}(\log T \log\log T)$	Delay
(Bray, 2022)	Bounded, continuous support, non-degeneracy	LP-based	$\mathcal{O}(\log T)$	Delay
(Chen et al., 2024)	Bounded, finite support, non-degeneracy	LP-based	$\mathcal{O}(1)$	Delay
(Li et al., 2024)	Bounded, finite support, non-degeneracy	LP-based	$\mathcal{O}(1)$	Delay
(Xu et al., 2024)	Bounded, continuous support, non-degeneracy	LP-based	$\mathcal{O}(\log(T/f))$	Delay
This paper	**Bounded, continuous support, non-degeneracy**	**LP-based & First-order**	$\mathcal{O}(\log(T/f)+\sqrt{f})$	**No Delay**
(Li et al., 2020)	Bounded	First-order	$\mathcal{O}(\sqrt{T})$	No Delay
(Balseiro et al., 2022a)	Bounded	First-order	$\mathcal{O}(\sqrt{T})$	No Delay
(Gao et al., 2023)	Bounded	First-order	$\mathcal{O}(\sqrt{T})$	No Delay
(Sun et al., 2020)	Bounded, finite support, non-degeneracy	First-order	$\mathcal{O}(T^{3/8})$	No Delay
(Gao et al., 2024)	Bounded, continuous support, non-degeneracy	First-order	$\mathcal{O}(T^{1/3})$	No Delay

Paper organization. The rest of the paper is organized as follows. Section 2 introduces the problem formulation and assumptions. Section 3 proposes our algorithms and main theoretical guarantee: a $\mathcal{O}(\log(T/f)+\sqrt{f})$ regret bound. Section 4 presents experiments to validate our theory.

2. Problem Setup

We use $\|\cdot\|$ to denote the Euclidean norm and $\langle\cdot\rangle$ to denote Euclidean inner product. Bold letters $\boldsymbol{A}$ and $\boldsymbol{a}$ denote matrices and vectors, respectively. Subscript $(\cdot)_{it}$ denotes the index for resource type i at time t. The notation $(\cdot)^+ = \max\{\cdot, 0\}$ denotes the element-wise positive part function, and $\mathbb{I}(\cdot)$ denotes the 0-1 indicator function.

2.1. OLP Formulation

We study an online resource allocation problem over the time horizon T under a *stochastic input model*. Initially, we have an inventory vector $\boldsymbol{b} \in \mathbb{R}^m$, representing m resource types. At each time step t, a customer arrives with a request sampled i.i.d. as $(r_t, \boldsymbol{a}_t) \sim \mathcal{P}$, where $\boldsymbol{r} = (r_1, \ldots, r_T)^\top \in \mathbb{R}^T$ is the offered payment (bid), $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_T) \in \mathbb{R}^{m\times T}$ is the matrix of customers' resource demands, and $\mathcal{P}$ is a fixed, unknown distribution. We must decide whether to accept or reject each request, represented by the decision variables $\boldsymbol{x} = (x_1, \ldots, x_T) \in \{0,1\}^T$. The goal is to maximize the cumulative reward. This problem can be formulated as the following OLP, referred to as the *primal linear program (PLP)*:

$$\max_{\boldsymbol{0}\le\boldsymbol{x}\le\boldsymbol{1}} \langle \boldsymbol{r}, \boldsymbol{x}\rangle \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}. \tag{PLP}$$

The dual problem of (PLP) is given by

$$\min_{(\boldsymbol{p},\boldsymbol{y})\ge\boldsymbol{0}} \langle \boldsymbol{b}, \boldsymbol{p}\rangle + \langle \boldsymbol{1}, \boldsymbol{y}\rangle \quad \text{s.t.} \quad \boldsymbol{A}^\top\boldsymbol{p} + \boldsymbol{y} \ge \boldsymbol{r} \tag{DLP}$$

where $\boldsymbol{p}$ is the vector of *dual prices*. Let $\boldsymbol{d} = \boldsymbol{b}/T \in \mathbb{R}^m$ denote the initial average resource. As demonstrated by (Li et al., 2020), (DLP) can be written as

$$\min_{\boldsymbol{p}\ge\boldsymbol{0}} f_T(\boldsymbol{p}) := \tfrac{1}{T}\textstyle\sum_{t=1}^T [\langle \boldsymbol{d}, \boldsymbol{p}\rangle + (r_t - \langle \boldsymbol{a}_t, \boldsymbol{p}\rangle)^+] \tag{1}$$

This formulation can be written as a T-sample approximation of the following stochastic program:

$$\min_{\boldsymbol{p}\ge\boldsymbol{0}} f(\boldsymbol{p}) := \mathbb{E}[f_T(\boldsymbol{p})] = \boldsymbol{d}^\top\boldsymbol{p} + \mathbb{E}[(r - \boldsymbol{a}^\top\boldsymbol{p})^+] \tag{2}$$

where the expectation is taken with respect to $(r, \boldsymbol{a}) \sim \mathcal{P}$.

We define optimal solutions to the T-sample approximation problem (1) and stochastic program (2) respectively as

$$\boldsymbol{p}_T^* = \arg\min_{\boldsymbol{p}\ge\boldsymbol{0}} f_T(\boldsymbol{p}) \quad \text{and} \quad \boldsymbol{p}^* = \arg\min_{\boldsymbol{p}\ge\boldsymbol{0}} f(\boldsymbol{p}).$$

The decision variable x_t^* can be established from the complementary slackness condition as

$$x_t^* = \begin{cases} 0, & r_t < \boldsymbol{a}_t^\top\boldsymbol{p}_T^*, \\ 1, & r_t > \boldsymbol{a}_t^\top\boldsymbol{p}_T^* \end{cases} \tag{3}$$

and $x_t^* \in [0,1]$ if $r_t = \boldsymbol{a}_t^\top\boldsymbol{p}_T^*$.

This connection between primal and dual solutions inspires OLP algorithms that make decisions based on dual prices:

$$x_t = \mathbb{I}(r_t \ge \boldsymbol{a}_t^\top\boldsymbol{p}_t). \tag{4}$$

2.2. Algorithms for OLP

We summarize two main dual-based OLP algorithms.

LP-based Method. The LP-based method (Li & Ye, 2022) calculates dual prices by re-solving the online LP problem at each time step. Specifically, define $d_{it} = b_{it}/(T-t)$ as the average remaining resource for type i at time t. The resulting optimization problem can be viewed as a t-sample approximation to the stochastic program specified by $\boldsymbol{d}_t = (d_{it}, ..., d_{mt})^\top$ as

$$\min_{\boldsymbol{p}\ge\boldsymbol{0}} \quad f_{\boldsymbol{d}_t}(\boldsymbol{p}) := \boldsymbol{d}_t^\top\boldsymbol{p} + \mathbb{E}[(r - \boldsymbol{a}^\top\boldsymbol{p})^+] \tag{5}$$

By updating $\boldsymbol{d}_t$, this method incorporates past decisions when computing $\boldsymbol{p}_t$. If resources are over-utilized in earlier periods, the supply decreases, prompting us to raise the dual price and be more selective with future orders. We outline this method in Algorithm 3 of Appendix A.2.

First-order Method. The first-order method (Li et al., 2020) calculates dual prices via online subgradient updates. We maintain a static average resource $\boldsymbol{d} = \boldsymbol{b}/T$, compute x_t as per (4), and update the dual price as

$$\boldsymbol{p}_{t+1} = (\boldsymbol{p}_t - \alpha_t(\boldsymbol{d} - \boldsymbol{a}_t x_t))^+$$

where the subgradient term evaluated at $\boldsymbol{p}_t$ is

$$\boldsymbol{d} - \boldsymbol{a}_t x_t \in \partial_{\boldsymbol{p}=\boldsymbol{p}_t}\big(\boldsymbol{d}^\top \boldsymbol{p} + (r_t - \boldsymbol{a}_t^\top \boldsymbol{p})^+\big).$$

This process can be interpreted as a projected stochastic subgradient method for solving (1). It reduces computational cost by requiring only a single pass through the data and eliminates solving LPs explicitly. A restart strategy improves the first-order method to have $\mathcal{O}(T^{1/3})$ regret (Gao et al., 2024). Algorithm 4 and Algorithm 5 in Appendix A.3 summarizes this method.

2.3. Performance Metrics

We aim to design algorithms that optimize a bi-objective performance measure involving regret and constraint violation. The regret measures the difference between the objective value of the algorithm's output and that of the true optimal solution, while the constraint violation measures the degree to which the algorithm's output fails to meet the given constraints. We denote the offline optimal solution to (PLP) by $\boldsymbol{x}^* = (x_1^*, \ldots, x_T^*)$, and the online algorithm output by $\boldsymbol{x} = (x_1, \ldots, x_T)$. Then, we define the regret $r(\boldsymbol{x})$ and resource violation $v(\boldsymbol{x})$ as

$$r(\boldsymbol{x}) \coloneqq \langle \boldsymbol{r}, \boldsymbol{x}^* \rangle - \langle \boldsymbol{r}, \boldsymbol{x} \rangle, \tag{6}$$

$$v(\boldsymbol{x}) \coloneqq \|(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^+\|. \tag{7}$$

Therefore, we define the following bi-objectives for evaluating an algorithm's worst-case performance:

$$\Delta_T = \sup_{\mathcal{P} \in \Xi} \mathbb{E}_{\mathcal{P}}[r(\boldsymbol{x}) + v(\boldsymbol{x})] \tag{8}$$

where Ξ denotes a family of distributions satisfying regularity assumptions specified later.

This metric is commonly used for first-order OLP algorithms (Gao et al., 2023; Li et al., 2020) and is also aligned with the literature on online convex optimization with constraints (Mahdavi et al., 2012; Yu et al., 2017). Integrating $r(\boldsymbol{x})$ and $v(\boldsymbol{x})$ into a single performance measure promotes balanced resource consumption over the decision horizon. In Section 3, we derive a decomposition of (8) that unifies the regret analysis for all OLP methods.

2.4. Assumptions and Auxiliary Results

We adopt the following assumptions regarding the stochastic inputs. These assumptions are standard in the online learning literature (Li & Ye, 2022; Jiang et al., 2022; Xu et al., 2024). In particular, we require the input data to be bounded, follow a linear growth, and be non-degenerate.

Assumption 2.1 (Boundedness)**.** We assume

(a) The order inputs $\{(r_t, \boldsymbol{a}_t)\}_{t=1}^T$ are generated i.i.d from an unknown distribution $\mathcal{P}$.

(b) There exist constants $\bar{r}, \bar{a} > 0$ such that $|r_t| \le \bar{r}$ and $\|\boldsymbol{a}_t\|_\infty \le \bar{a}$ almost surely for $t = 1, ..., T$.

(c) The average resource capacity $\boldsymbol{d} = \boldsymbol{b}/T$ satisfies $d_i \in [\underline{d}, \bar{d}]$ for some $\bar{d} > \underline{d} > 0$ for any $i = 1, ..., m$.

In this assumption, (a) states that $\{(r_t, \boldsymbol{a}_t)\}_{t=1}^T$ are independent of each other, but we allow dependencies between their components. Part (b) introduces the bounds $\bar{r}, \bar{a}$ solely for analytical purposes. This is a minimal requirement on $(r_t, \boldsymbol{a}_t)$ compared to previous work (Agrawal et al., 2014; Li & Ye, 2022; Xu et al., 2024). Part (c) requires the average resource to grow linearly with T, ensuring that a constant fraction of x_t values can be set to 1. Consequently, the number of fulfillable orders is proportional to T, facilitating a stable service level over time.

Assumption 2.2 (Uniform Non-degeneracy)**.** We assume

(a) The second-moment matrix $\mathbb{E}[\boldsymbol{a}\boldsymbol{a}^\top]$ is positive definite with minimum eigenvalue λ.

(b) There are constants μ, ν such that for any $(r, \boldsymbol{a}) \sim \mathcal{P}$,

$$\begin{aligned} &\nu|\boldsymbol{a}^\top(\boldsymbol{p} - \boldsymbol{p}^*)| \\ \le\ &\big|\mathbb{P}(r \ge \boldsymbol{a}^\top \boldsymbol{p} \mid \boldsymbol{a}) - \mathbb{P}(r \ge \boldsymbol{a}^\top \boldsymbol{p}^* \mid \boldsymbol{a})\big| \\ \le\ &\mu|\boldsymbol{a}^\top(\boldsymbol{p} - \boldsymbol{p}^*)| \end{aligned}$$

holds for all

$$\boldsymbol{p} \in \mathcal{V}_{\boldsymbol{p}} \coloneqq \left\{\boldsymbol{p} \in \mathbb{R}^m : \boldsymbol{p} \ge \boldsymbol{0}, \|\boldsymbol{p}\| \le \frac{\bar{r}}{\underline{d}}\right\}$$

and $\boldsymbol{d} \in \mathcal{V}_{\boldsymbol{d}} \coloneqq [\underline{d}, \bar{d}]^m$.

(c) The optimal $\boldsymbol{p}^*$ satisfies $p_i^* = 0$ if and only if $d_i - \mathbb{E}[a_i \mathbb{I}(r > \boldsymbol{a}^\top \boldsymbol{p}^*)] > 0$ for all $\boldsymbol{d} \in \mathcal{V}_{\boldsymbol{d}}$ and $i \in [m]$.

In this assumption, part (b) ensures that the cumulative distribution of the reward given the resource consumption request $r|\boldsymbol{a}$ is continuous and exhibits a stable growth rate. Part (c) requires strict complementarity for the optimal solutions of the stochastic program in (2), which is a non-degeneracy condition for both the primal and dual LPs.

According to stochastic program (2), we define the binding and non-binding index sets as:

$$I_B = \{i : d_i - \mathbb{E}[a_i \mathbb{I}(r > \boldsymbol{a}^\top \boldsymbol{p}^*)] = 0\},$$
$$I_N = \{i : d_i - \mathbb{E}[a_i \mathbb{I}(r > \boldsymbol{a}^\top \boldsymbol{p}^*)] > 0\}. \quad (9)$$

By Assumption 2.2(c), these sets are complements, as $I_B \cap I_N = \emptyset$ and $I_B \cup I_N = \{1, ..., m\}$.

3. Parallel Multi-Phase OLP Algorithm

We now present our algorithm for parallel multi-phase online learning and decision-making. Our approach is motivated by the challenges of existing methods. Specifically, the LP-based method has a strong worst-case regret bound of $\mathcal{O}(\log T)$, but its high computational costs lead to decision-making delays. Meanwhile, the first-order method updates decisions efficiently but suffers a high regret of $\mathcal{O}(\sqrt{T})$. By combining the strengths of these two methods, our new framework balances decision-making quality and computational efficiency.

3.1. Algorithm Design

We establish our framework with the following design:

1. We maintain two parallel paths of online learning and online decision-making. The online learning results are periodically sent to the decision-making path as a re-start point to guide subsequent updates.

2. In the online learning path, we employ a streamlined LP-based method, which only re-solves the updated OLP problem according to a predefined frequency, reducing the computational overhead.

3. In the online decision-making path, we apply the first-order method during the initial and final batches. The learning rate is optimally tuned for these two intervals.

Figure 1 illustrates our framework, which generates two parallel sequences of dual prices: $\{\boldsymbol{p}_t^D\}_{t=1}^T$ from the first-order method and $\{\boldsymbol{p}_t^L\}_{t=1}^T$ from the LP-based method. The algorithm proceeds in batches of length f. In the first batch, it uses the first-order method to iteratively update the dual prices, thereby guiding decision-making. At the end of this batch, it applies the LP-based method to obtain a refined dual price and passes it to the decision-making path. The LP re-solving occurs only once per batch, making the approach computationally efficient. Formally, we re-solve the OLP at every time t satisfying $t \le kf$ and $t \bmod f = 0$, where $k = \lfloor T/f \rfloor$ is the number of batches.

During intermediate batches, decisions are made based on the most recent dual prices computed from the LP at the end of the previous batch. The algorithm restarts the subgradient updates only in the final batch to guide the remaining decisions. Algorithm 1 summarizes this approach.

Algorithm 1 Parallel Multi-Phase OLP Algorithm

Input: total resource $\boldsymbol{b}$, time horizon T, average resource $\boldsymbol{d} = \boldsymbol{b}/T$, initial dual price $\boldsymbol{p}_1 = \boldsymbol{0}$, re-solving frequency f, and number of batches $k = \lfloor T/f \rfloor$

for $t = 1$ to T **do**
  Observe $(r_t, \boldsymbol{a}_t)$ and make decision x_t as rule (4)
  Update constraint for $i = 1, ..., m$:
    remaining resource $b_{it} = b_{i,t-1} - a_{it}x_t$
    average remaining resource $d_{it} = \frac{b_{it}}{T-t}$
  if $t \le f$ or $t \ge kf$ **then**
    Update learning rate $\alpha_t = \begin{cases} \mathcal{O}(1/f^{1/2}) & t \le f \\ \mathcal{O}(1/f^{2/3}) & t \ge kf \end{cases}$
    Compute subgradient and update dual price $\boldsymbol{p}_{t+1}$:
    $$\boldsymbol{p}_{t+1} = (\boldsymbol{p}_t - \alpha_t(\boldsymbol{d} - \boldsymbol{a}_t x_t))^+$$
  end
  if $t \bmod f = 0$ and constraints are not violated **then**
    Solve OLP and update dual price $\boldsymbol{p}_{t+1}$:
    $$\boldsymbol{p}_{t+1} = \operatorname*{argmin}_{\boldsymbol{p} \ge \boldsymbol{0}} \ \boldsymbol{d}_t^\top \boldsymbol{p} + \tfrac{1}{t}\textstyle\sum_{j=1}^t (r_j - \boldsymbol{a}_j^\top \boldsymbol{p})^+$$
  end
  else
    Update dual price $\boldsymbol{p}_{t+1}$ to be the most recent solution $\boldsymbol{p}_{t+1} = \boldsymbol{p}_t$.
  end
end

Algorithm 1 balances online learning and efficient decision-making. It is responsive to dynamic environments, adapts to the latest information, and reduces computational cost by periodic re-solving. We thus establish a "wait-less" online decision-making framework where each customer's order is processed immediately without delays from earlier requests or large-scale LP computations.

3.2. Algorithm Analysis

We decompose the performance metric of Algorithm 1 into three key components. All proofs (including essential properties of the dual price $\boldsymbol{p}_t$) are in Appendix B and C.

Theorem 3.1 (Performance Metric)**.** *Under Assumptions 2.1 and 2.2, the performance Δ_T of Algorithm 1 satisfies*

$$\Delta_T \ \le \ \mu\bar{a}^2 \textstyle\sum_{t=1}^T \mathbb{E}\left[\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2\right] + \mathbb{E}\left[\|(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^+\|\right] + \|\boldsymbol{p}^*\| \cdot \mathbb{E}\left[\|(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})^{B+}\|\right]. \quad (10)$$

where $(\cdot)^{B+}$ indicates the projection of binding terms onto the positive orthant.

Based on the definition of Δ_T (8), we derive the perfor-

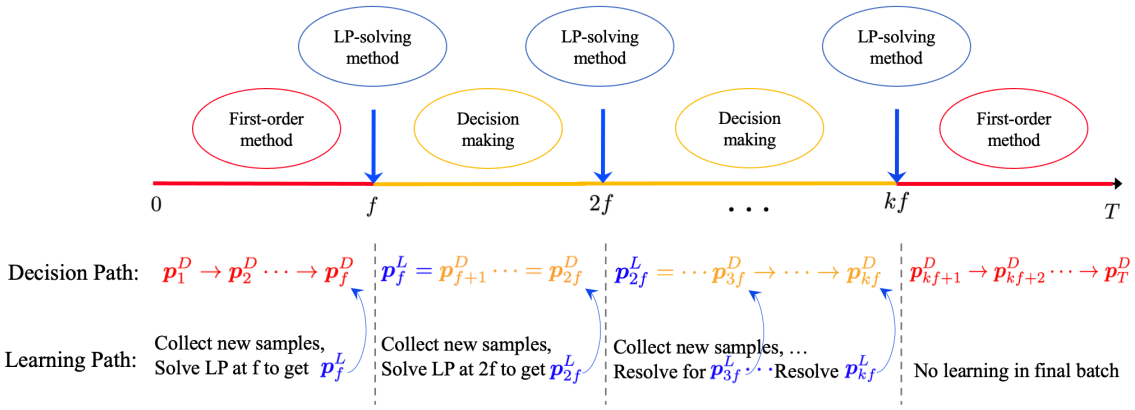


Figure 1: Algorithm 1 illustration of parallel paths and the interactions between online learning and decision-making. Decisions are generated based on 1) the LP-based method (blue) with frequency f, 2) the first-order method (red) for the initial and final phases (with a warm start), and 3) employing the latest dual price (yellow) during intermediate phases.

mance metric as the sum of three key components: dual convergence, remaining resources, and binding constraint violation. This bound demonstrates that our algorithm encourages 1) smooth and balanced resource utilization, and 2) full resource consumption by the end of the time horizon. Building on Theorem 3.1, we provide a "spectrum theorem" for the algorithm's performance.

Theorem 3.2. *Under Assumptions 2.1 and 2.2, the worst-case performance Δ_T of Algorithm 1 is bounded by*

$$\Delta_T \in \mathcal{O}\Big(\log\Big(\frac{T}{f}\Big) + \sqrt{f}\Big) \tag{11}$$

where T/f represents the number of re-solving batches and f is the length of each batch.

Remark 3.3 (Spectrum Theorem)**.** We elucidate the trade-offs in total regret induced by using LP-based methods with $\mathcal{O}(\log T)$ regret and first-order methods with $\mathcal{O}(\sqrt{T})$ regret. Our algorithm introduces a re-solving frequency $f \in [1, T]$ and achieves $\mathcal{O}(\log(T/f) + \sqrt{f})$ regret, which recovers previous results with extreme cases of $f = 1$ and $f = T$. Specifically, the first-order method gives us the regret of $\sqrt{f}$ for the first batch and $f^{1/3}$ for the last batch, and the LP-based method contributes the regret of $\log(T/f)$.

Remark 3.4 (Warm Start)**.** We can obtain a tighter regret bound of $\mathcal{O}(\log(T/f) + f^{1/3})$ given a warm start of the initial dual price satisfying $\|\boldsymbol{p}_0 - \boldsymbol{p}^*\| \le f^{-1/3}$ or if we use the LP-based method at each time step for the first batch.

Remark 3.5 (Learning Rate Selection)**.** We select the best learning rate to minimize the regret upper bound consisting of (6) and (7). For the first batch, the regret grows linearly with α_t while the constraint violation is inversely proportional to α_t. To achieve the tightest bound, we balance this trade-off by selecting the learning rate that minimizes the overall expression, which yields the optimal choice $\alpha_t = 1/\sqrt{f}$ as derived in Theorem B.7. A similar analysis for the final batch leads to the choice $\alpha_t = 1/f^{2/3}$ as shown in Theorem B.9.

Technical Intuitions. Theorem 3.1 decomposes the performance metrics to make the regret analysis compatible between the LP-based and first-order methods.

- **LP-based Method:** As Theorem 3.1 suggests, achieving small regret requires $\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$ to be close to zero. To enforce this, we impose a stronger condition to track the average remaining resource—if $\boldsymbol{d}_t$ exceeds the allowed limit, we manually set the dual prices to zero to accept all subsequent orders. This approach enables us to eliminate the "stopping time" argument and make the analysis compatible with our new framework.
- **First-order Method:** The regret improvement comes from the final batch since it starts with the LP-derived learning result $\boldsymbol{p}_{kf}$, which lies within a $\mathcal{O}(1/\sqrt{kf})$-sized neighborhood around $\boldsymbol{p}^*$. Using the subgradient method, we bound the three components in Theorem 3.1 in terms of $\|\boldsymbol{p}_t - \boldsymbol{p}^*\|$ and express them in terms of f, k, T, and α_t. Optimizing the learning rate to $\alpha_t = f^{-2/3}$ yields an improved regret of $\mathcal{O}(f^{1/3})$ in the final batch.

3.3. Algorithm Extension

We propose an enhanced version of Algorithm 1 that employs the first-order method for decision-making between two consecutive LP resolves. Rather than making decisions solely with the most recently solved dual price from the learning path, Algorithm 2 treats it as a new starting point for each re-solving interval and adopts a smaller step size for subgradient updates to avoid deviating too far from the LP-guided solutions. Figure 2 illustrates this framework. Specifically, for each interval $[jf, (j+1)f]$, we attain the new start $\boldsymbol{p}_{jf}$ from the learning path at time jf and con-

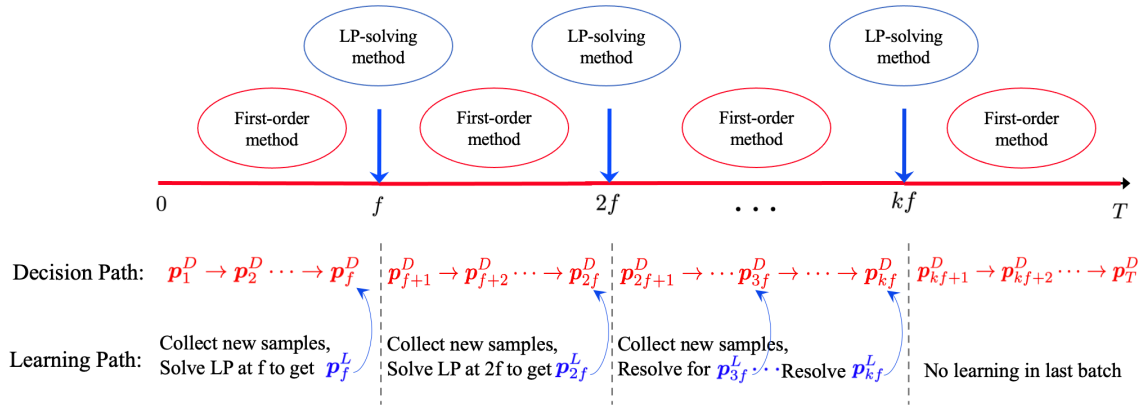


Figure 2: Algorithm 2 illustration of parallel paths with multi-time restart. Decisions are generated based on 1) the LP-based method (blue) with frequency f and 2) the first-order method (red) during re-solving intervals with a warm start.

Algorithm 2 Enhanced Multi-Start OLP Algorithm

Input: total resource $\boldsymbol{b}$, horizon T, average resource $\boldsymbol{d} = \boldsymbol{b}/T$, initial dual $\boldsymbol{p}_1 = \boldsymbol{0}$, re-solving frequency f

for $t = 1$ to T **do**

  Observe $(r_t, \boldsymbol{a}_t)$ and make decision x_t as rule (4)

  Update constraint for $i = 1, ..., m$:

    remaining resource constraint $b_{it} = b_{i,t-1} - a_{it}x_t$

    average resource capacity $d_{it} = \frac{b_{it}}{T-t}$

  if $t \bmod f \neq 0$ **then**

    Update learning rate $\alpha_t = \mathcal{O}(1/t)$

    Compute subgradient and update dual price $\boldsymbol{p}_{t+1}$:

$$\boldsymbol{p}_{t+1} = (\boldsymbol{p}_t - \alpha_t(\boldsymbol{d} - \boldsymbol{a}_t x_t))^+$$

  end

  if $t \bmod f = 0$ and constraints are not violated **then**

    Solve OLP and update dual price $\boldsymbol{p}_{t+1}$:

$$\boldsymbol{p}_{t+1} = \underset{\boldsymbol{p} \geq \boldsymbol{0}}{\operatorname{argmin}} \ \boldsymbol{d}_t^\top \boldsymbol{p} + \frac{1}{t}\sum_{j=1}^{t}(r_j - \boldsymbol{a}_j^\top \boldsymbol{p})^+$$

  end

end

tinue to fine-tune the dual price $\{\boldsymbol{p}_t\}_{t=jf}^{(j+1)f}$ using the first-order method with a learning rate $\alpha_t = \mathcal{O}(1/t)$. Algorithm 2 summarizes this approach.

As we illustrate in our experiments (Section 4), the incorporation of an intermediate first-order method improves the algorithm's stability and ensures smooth resource consumption. Therefore, the multi-restart mechanism results in better performance during the final batch and improves the algorithm's total performance.

3.4. Algorithm Application

Our motivation in designing Algorithm 1 is to effectively balance computational efficiency and decision optimality. Building on the Spectrum Theorem 3.2, this section aims to translate our theoretical results into practical applications.

We formulate a new optimization problem to find the optimal re-solving frequency that minimizes the regret in Theorem 3.2, taking into account computational resource capacities. Denote $c_1(\cdot), c_2(\cdot)$ as computational cost functions for LP-based and first-order methods, and R as the total computational resource capacity. The optimal value of f is determined by solving the following optimization problem:

$$\begin{aligned} \min_{f \in \{1,\ldots,T\}} \quad & \log\left(\tfrac{T}{f}\right) + \sqrt{f} \\ \text{s.t.} \quad & c_1(k) + 2c_2(f) \leq R. \end{aligned} \tag{12}$$

Specifically, if we use the interior-point method or the simplex method as the LP solver, the computational cost is $m^2(m+t)$ for any time t. The first-order method updates gradients in constant time. The following proposition provides a concrete example in practice.

Proposition 3.6 (Optimal Re-solving Frequency)**.** *Given a fixed computation resource capacity R, if we use the interior-point or simplex method as the LP solver in Algorithm 1, we can instantiate Constraint (12) as*

$$\textstyle\sum_{j=1}^{k}(m^2(m+jf)) + 2mf \leq R.$$

This proposition enables users to determine the optimal re-solving frequency that balances regret and computational cost based on available computational resources.

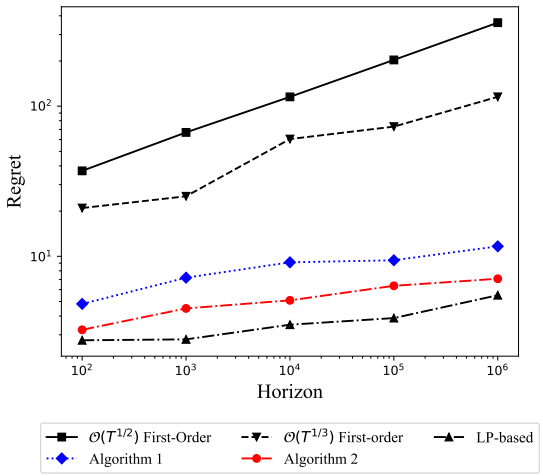
4. Numerical Experiments

We conduct extensive experiments to evaluate our algorithm's performance and validate our theoretical results. This section is divided into two parts. In the first part (Section 4.1), we evaluate Algorithms 1 and 2 across different choices of re-solving frequency. In the second part (Section 4.2), we compare our algorithm with LP-based and

Table 3: Algorithms comparison.

T	Regret	Algorithm	Compute Time (s)
10^4	115.32	$\mathcal{O}(T^{1/2})$ First-Order	0.008
	60.39	$\mathcal{O}(T^{1/3})$ First-Order	0.013
	3.50	LP-based	123.497
	9.12	Algorithm 1	**1.8**
	5.09	Algorithm 2	**1.8**
10^5	203.20	$\mathcal{O}(T^{1/2})$ First-Order	0.118
	73.07	$\mathcal{O}(T^{1/3})$ First-Order	0.108
	3.88	LP-based	> 3600
	9.41	Algorithm 1	**56.9**
	6.36	Algorithm 2	**56.8**
10^6	351.91	$\mathcal{O}(T^{1/2})$ First-Order	1.211
	115.39	$\mathcal{O}(T^{1/3})$ First-Order	1.577
	5.50	LP-based	> 100000
	11.65	Algorithm 1	**2155.9**
	7.09	Algorithm 2	**2242.1**

Figure 4: Regret for various algorithms.



first-order methods in terms of regret and running time. All implementations can be found at GitHub Link.

We consider the following distributions:

$$\text{Input I: } a_{it} \sim \text{Unif}[0,2], r_t \sim \text{Unif}[0,10]$$
$$\text{Input II: } a_{it} \sim \mathcal{N}(0.5,1), r_t \sim \mathcal{N}(0.5m,m)$$

Learning rates are selected as specified in Section 3:

$$\text{Algorithm 1: } \alpha_t = \begin{cases} \mathcal{O}(1/f^{1/2}) & t \le f \\ \mathcal{O}(1/f^{2/3}) & t \ge kf \end{cases}$$
$$\text{Algorithm 2: } \alpha_t = \mathcal{O}(1/t)$$

4.1. Regret under Varying Re-solving Frequencies

We choose $m = 1$ and generate the sequence $\{(r_t, \boldsymbol{a}_t)\}_{t=1}^T$ randomly from the uniform (Input I) and normal (Input II) distributions, and include more complex distributions in Appendix E.1. The time horizon T spans evenly over $[10^2, 10^6]$, the initial average resource is sampled as $d_i \sim$ Uniform$[1/3, 2/3]$, and $f \in \{T^{1/3}, T^{1/2}, T^{2/3}\}$ representing high, medium, and low re-solving frequencies. We report the average result over 100 trials for each experiment. We use the classic first-order method with $\mathcal{O}(T^{1/2})$ regret (Li et al., 2020) as a baseline. Results are summarized in Table 2 and Figure 3 on a logarithmic scale.

We analyze the absolute value of regret and its growth rate over time. We observe that regret decreases as the re-solving frequency increases. This trend holds consistently across both algorithms and input types. In addition, while regret accumulates over longer time horizons, the increasing rate remains stable for algorithms employing higher re-solving frequencies. These findings are consistent with the guarantees of Theorem 3.2, as more frequent updates enable better adaptation to dynamic environments.

4.2. Comparative Analysis with Baseline Methods

We next compare our algorithm's regret and computation time with a classic LP-based (Li & Ye, 2022) and two first-order methods with $\mathcal{O}(T^{1/2})$ (Li et al., 2020) and $\mathcal{O}(T^{1/3})$ regrets (Gao et al., 2024). We generate $\{(r_t, \boldsymbol{a}_t)\}_{t=1}^T$ from a uniform distribution (Input I). We set the resource types to $m = 5$, time T to range evenly over $[10^2, 10^6]$, average resource $d_i \sim$ Uniform$[1/3, 2/3]$, and re-solving frequency to $f = T^{1/3}$ from Section 4.1. Each result is averaged over 100 trial runs. We summarize the findings in Table 3 and Figure 4, and we include more comparisons with an infrequent re-solving method (Li et al., 2024) in Appendix E.2. We observe the effectiveness of Algorithm 1 and 2:

1. Our algorithms exhibit strong performance in terms of decision optimality. They achieve over a 20-fold and 10-fold improvement in regret compared to the $\mathcal{O}(T^{1/2})$ first-order method and the $\mathcal{O}(T^{1/3})$ first-order method respectively. These numerical results also corroborate the theoretical bounds in Theorem 3.2.

2. Our algorithms are computationally efficient. Their runtimes exhibit over 100-fold improvement compared to the LP-based method with only a minimal increase (less than 2-fold) in regret.

Therefore, we achieve a balance between effective decision-making and efficient computation. Our algorithms demonstrate better regret than the first-order method and obtain substantial computational speed-ups compared to the LP-based method. They exhibit strong scalability and adaptability across various re-solving frequencies and stochastic input models, consistently delivering superior performance as the problem size grows.

Table 2: Regret of algorithms with various re-solving frequencies.

	T	First-Order	Low freq	Mid freq	High freq
Input I	10^3	12.13	7.76	5.77	**4.86**
	10^4	38.50	10.96	7.55	**5.67**
	10^5	122.44	23.90	9.92	**8.36**
	10^6	404.59	56.70	21.90	**8.99**
Input II	10^3	11.44	6.28	4.86	**3.95**
	10^4	36.50	10.21	7.34	**3.81**
	10^5	115.57	14.61	11.78	**4.66**
	10^6	365.99	35.20	15.68	**6.26**

(a) Algorithm 1

	T	First-Order	Low freq	Mid freq	High freq
Input I	10^3	12.13	6.78	5.20	**4.50**
	10^4	38.50	10.37	8.03	**5.99**
	10^5	122.44	22.33	11.57	**6.36**
	10^6	404.59	48.21	22.44	**7.09**
Input II	10^3	11.44	3.20	2.56	**1.75**
	10^4	36.50	5.48	4.30	**2.52**
	10^5	115.57	12.35	4.48	**3.86**
	10^6	365.99	30.48	13.20	**4.77**

(b) Algorithm 2

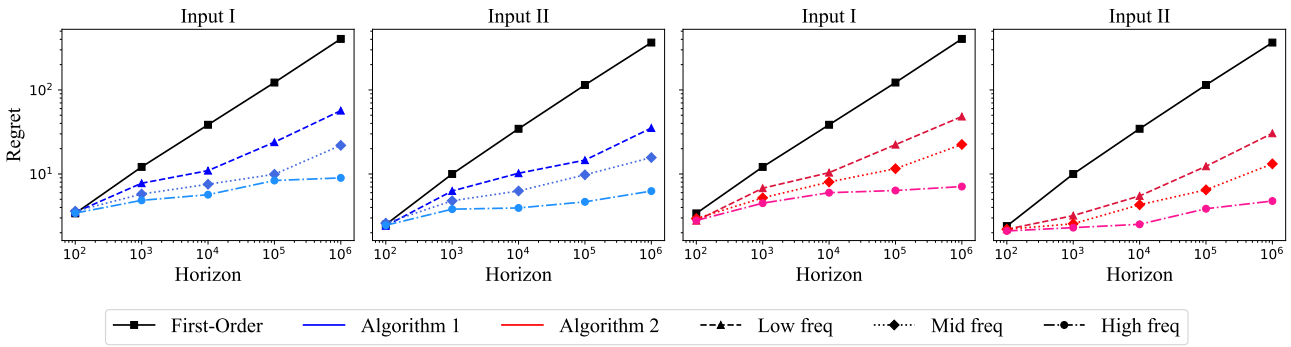


Figure 3: Evaluations of Algorithm 1 and 2 across various horizons, re-solving frequencies, and stochastic inputs, validating the positive relationship between regret and frequency stated in Theorem 3.2.

5. Conclusion and Discussion

This paper presents a new approach to online linear programming for dynamic resource allocation, addressing the inherent trade-offs between decision-making optimality and computational efficiency. Recognizing the limitations of existing methods, we propose a parallel framework that decouples online learning and decision-making into independent yet complementary processes. By integrating LP-based and first-order methods, our framework effectively balances total regret and computational cost.

We establish rigorous theoretical guarantees, proving that our algorithm achieves a worst-case regret bound of $\mathcal{O}(\log(T/f) + \sqrt{f})$ under continuous support. This result highlights our method's ability to interpolate between LP-based and first-order methods based on computational capability. Furthermore, extensive experiments validate the effectiveness of our approach, demonstrating improvements in both regret minimization and runtime efficiency over competitive baselines.

Beyond these contributions, our work opens avenues for further research in adaptive online decision-making. Future directions include refining the re-solving frequency based on real-time computational constraints and extending our framework to broader classes of online optimization problems. Overall, our results underscore the potential of hybrid algorithms in high-dimensional, large-scale environments, offering practical insights for applications in operations research, machine learning, and beyond.

Discussion of Unknown Horizon

In this paper, we consider decision-making under a finite horizon, since the average resource Assumption 2.1(c) widely used in OLP literature relies on a fixed horizon to be well-defined. When the horizon is unknown, this assumption becomes ill-posed, and prior work (Balseiro et al., 2023) shows that it may not be possible to achieve sublinear regret. Addressing these challenges would likely require first adapting LP-based and first-order OLP methods to uncertain horizons before tackling a hybrid approach.

In practice, it is often possible to make a prior prediction of the horizon using data-driven approaches or based on the resources available. For example, in online advertising, the number of customers can be predicted from historical statistics. Besides, in retail or event-based sales, the selling horizon may be externally decided by upper-level decision-makers. These practical applications suggest a reasonable modeling choice for finite time horizons.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

Acknowledgments

This work was supported in part by NSF grant CCF-2338226.

References

Agrawal, S., Wang, Z., and Ye, Y. A dynamic near-optimal algorithm for online linear programming. *Operations Research*, 62(4):876–890, 2014.

Balseiro, S., Kroer, C., and Kumar, R. Online resource allocation under horizon uncertainty. In *Abstract Proceedings of the 2023 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pp. 63–64, 2023.

Balseiro, S. R., Lu, H., and Mirrokni, V. The best of many worlds: Dual mirror descent for online allocation problems. *Operations Research*, 2022a.

Balseiro, S. R., Lu, H., Mirrokni, V., and Sivan, B. From online optimization to PID controllers: Mirror descent with momentum. *arXiv preprint arXiv:2202.06152*, 2022b.

Bray, R. L. Logarithmic regret in multisecretary and online linear programming problems with continuous valuations. *arXiv preprint arXiv:1912.08917*, 2022.

Chen, G., Li, X., and Ye, Y. An improved analysis of lp-based control for revenue management. *Operations Research*, 72(3):1124–1138, 2024.

Devanur, N. R., Jain, K., Sivan, B., and Wilkens, C. A. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. *Journal of the ACM (JACM)*, 66(1):1–41, 2019.

Gao, W., Ge, D., Sun, C., and Ye, Y. Solving linear programs with fast online learning algorithms. In *International Conference on Machine Learning*, pp. 10649–10675. PMLR, 2023.

Gao, W., Sun, C., Xue, C., and Ye, Y. Decoupling learning and decision-making: Breaking the $\mathcal{O}(\sqrt{T})$ barrier in online resource allocation with first-order methods. In *International Conference on Machine Learning*, pp. 14859–14883. PMLR, 2024.

Goel, G. and Mehta, A. Online budgeted matching in random input models with applications to adwords. In *SODA*, volume 8, pp. 982–991, 2008.

Golrezaei, N. and Yao, E. Upfront commitment in online resource allocation with patient customers. *arXiv preprint arXiv:2108.03517*, 2021.

Gupta, A. and Molinaro, M. How the experts algorithm can help solve lps online. *Mathematics of Operations Research*, 41(4):1404–1431, 2016.

Jiang, J., Ma, W., and Zhang, J. Degeneracy is ok: Logarithmic regret for network revenue management with indiscrete distributions. *arXiv preprint arXiv:2210.07996*, 2022.

Li, G., Wang, Z., and Zhang, J. Infrequent resolving algorithm for online linear programming. *arXiv preprint arXiv:2408.00465*, 2024.

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

Li, X., Sun, C., and Ye, Y. Simple and fast algorithm for binary integer and online linear programming. *Advances in Neural Information Processing Systems*, 33:9412–9421, 2020.

Ma, W., Cao, Y., Tsang, D. H., and Xia, D. Optimal regularized online allocation by adaptive re-solving. *Operations Research*, 2024.

Mahdavi, M., Jin, R., and Yang, T. Trading regret for efficiency: Online convex optimization with long term constraints. *Journal of Machine Learning Research*, 13(81): 2503–2528, 2012.

Mehta, A., Saberi, A., Vazirani, U., and Vazirani, V. Adwords and generalized online matching. *Journal of the ACM (JACM)*, 54(5):22–es, 2007.

Sun, R., Wang, X., and Zhou, Z. Near-optimal primal-dual algorithms for quantity-based network revenue management. *arXiv preprint arXiv:2011.06327*, 2020.

Talluri, K. T., Van Ryzin, G., and Van Ryzin, G. *The theory and practice of revenue management*, volume 1. Springer, 2004.

Xie, Y., Ma, W., and Xin, L. The benefits of delay to online decision-making. *Available at SSRN 4248326*, 2023.

Xu, H., Glynn, P. W., and Ye, Y. Online linear programming with batching. *arXiv preprint arXiv:2408.00310*, 2024.

Yu, H., Neely, M., and Wei, X. Online convex optimization with stochastic constraints. *Advances in Neural Information Processing Systems*, 30, 2017.

Appendix

Table of Contents

Structure of the Appendix We organize the appendix as follows. In **Section A**, we present some primary results for the OLP problem and the foundation analysis for LP-based and first-order methods; **Section B** characterizes the unique properties from our parallel multi-phase structure, and presents the cohesive mathematical framework which decomposes the performance metrics to unify all the OLP methods; **Section C** demonstrates the outline of Algorithm 2, combines all the previous results and proves the main results in our paper; **Section D** provides auxiliary results and technical support for the previous three sections; **Section E** includes supplementary experiments of more general input distributions and comparisons with recent algorithms.

A. Primary Results and Properties

In this section, we present some primary results for the problem and OLP dual algorithms which will help present our main results. We also include basic properties and convergence of the LP-based and first-order methods. All the analyses are stated under Assumption 2.1 and 2.2.

A.1. Preliminary results

We propose the following lemma as directed results from Assumptions. Let Ξ denote the family of distributions satisfying Assumption 2.1 and 2.2. We propose the following lemma.

Lemma A.1 (Dimension Stability)**.** *Under Assumptions 2.1 and 2.2, there exists a constant $\delta > 0$ such that*

$$\forall \boldsymbol{d}_t \in \mathcal{D} := [d_i - \delta, d_i + \delta]^m, \text{ stochastic programs (5) specified by } \boldsymbol{d}_t \text{ share the same } I_B \text{ and } I_N \text{ sets}.$$

This lemma is a consequence of Lemmas 12 and 13 in (Li & Ye, 2022). The existence of δ comes from the continuity of $f_{\boldsymbol{d}_t}(\boldsymbol{p})$. Thus, δ is only associated with stochastic program (5), and it is independent of T. Note that $\mathcal{D} \subset \mathcal{V}_{\boldsymbol{d}}$. We will analyze our feasible dual solutions derived from the online algorithm with $\boldsymbol{d}_t \in \mathcal{D}$.

Lemma A.2 (Quadratic Regularity, Proposition 2 in (Li & Ye, 2022))**.** *Under Assumptions 2.1 and 2.2, for any $\boldsymbol{p} \in \mathcal{V}_{\boldsymbol{p}}$,*

$$f(\boldsymbol{p}) \le f(\boldsymbol{p}^*) + \nabla f(\boldsymbol{p}^*)^\top (\boldsymbol{p} - \boldsymbol{p}^*) + \frac{\mu \bar{a}^2}{2} \|\boldsymbol{p} - \boldsymbol{p}^*\|^2,$$
$$f(\boldsymbol{p}) \ge f(\boldsymbol{p}^*) + \nabla f(\boldsymbol{p}^*)^\top (\boldsymbol{p} - \boldsymbol{p}^*) + \frac{\nu \lambda}{2} \|\boldsymbol{p} - \boldsymbol{p}^*\|^2.$$

Moreover, $\boldsymbol{p}^*$ is the unique optimal solution to (2).

This lemma establishes a local form of semi-strong convexity and smoothness at $\boldsymbol{p}^*$, which is guaranteed by our assumptions on the distribution $\mathcal{P} \in \Xi$. By focusing on these local properties rather than insisting on strong convexity and global smoothness, we relax the classical requirements typically imposed in such settings. In later sections, we will leverage this result to derive our regret bound.

A.2. LP-based analysis

We state the LP-based method in Algorithm 3 and its properties of boundedness and dual convergence results.

Algorithm 3 LP-based method

Input: total resource $\boldsymbol{b}$, time horizon T, average resource $\boldsymbol{d} = \boldsymbol{b}/T$, and initial dual price $\boldsymbol{p}_1 = \boldsymbol{0}$.
for t = 1 to T **do**
  Observe $(r_t, \boldsymbol{a}_t)$ and make decision x_t based on rule (4) if constraints are not violated.
  Update constraint for $i = 1, ..., m$:
    remaining resource constraint $b_{it} = b_{i,t-1} - a_{it} x_t$
    average resource capacity $d_{it} = \frac{b_{it}}{T-t}$
  Solve the updated dual problem and obtain dual price $\boldsymbol{p}_{t+1}$:

$$\boldsymbol{p}_{t+1} = \underset{\boldsymbol{p} \ge \boldsymbol{0}}{\operatorname{argmin}} \ \boldsymbol{d}_t^\top \boldsymbol{p} + \frac{1}{t} \sum_{j=1}^{t} (r_j - \boldsymbol{a}_j^\top \boldsymbol{p})^+$$

end

LP-based method incorporates the past decisions into the optimization of $\boldsymbol{p}_t$. If resources were over-utilized in earlier periods, the remaining supply decreases, prompting the algorithm to raise the dual price and become more selective with future orders. Conversely, if ample resources remain, the future dual price will be lowered, allowing more consumer requests to be accepted. This adaptive mechanism accounts for past actions by adjusting the available resource capacity.

Lemma A.3 (Boundedness of LP result)**.** *The online dual price $\boldsymbol{p}_t$ from Algorithm 3 and the optimal dual price $\boldsymbol{p}^*$ of stochastic program (2) are bounded as*

$$\|\boldsymbol{p}^*\| \le \frac{\bar{r}}{\underline{d}}, \quad \|\boldsymbol{p}_t\| \le \frac{\bar{r}}{\underline{d} - \delta}.$$

Lemma A.4 (Dual Convergence of LP-based algorithm)**.** *Under Assumptions 2.1 and 2.2, $\boldsymbol{p}_t$ represents the online solution from LP-based method, there exists a constant $C_{lp} > 0$ depending on $\bar{r}, \bar{a}, \underline{d}, m, \nu$, and λ such that*

$$\mathbb{E}\left[\|\boldsymbol{p}_t - \boldsymbol{p}_t^*\|^2\right] \leq \frac{C_{lp}}{t}. \tag{13}$$

In addition, the difference between $\boldsymbol{p}_t^*$ and $\boldsymbol{p}^*$ satisfies

$$\|\boldsymbol{p}_t^* - \boldsymbol{p}^*\|^2 \leq \frac{1}{\nu^2\lambda^2}\|\boldsymbol{d}_t - \boldsymbol{d}\|^2. \tag{14}$$

This lemma establishes that the LP-based online AHDL algorithm (Jiang et al., 2022) produces dual solutions $\boldsymbol{p}_t$ that converge to $\boldsymbol{p}_t^*$. This convergence highlights the stability of the online dual variables during the intermediate stages of decision-making and ensures a warm start for the first-order method in the last re-solving batch. Moreover, we bound the distance between $\boldsymbol{p}_t^*$ and $\boldsymbol{p}^*$ by relating their respective average resource capacities, $\boldsymbol{d}_t$ and $\boldsymbol{d}$, in the associated stochastic programs (Li & Ye, 2022). These results provide the foundation for analyzing $\|\boldsymbol{p}_t - \boldsymbol{p}^*\|$ in our final result.

Proof of Lemma A.3. By the optimality of $\boldsymbol{p}^*$ and boundedness of $\boldsymbol{d}$, we have

$$\underline{d}\|\boldsymbol{p}^*\|_1 \leq \boldsymbol{d}^\top \boldsymbol{p}^* \leq \mathbb{E}[r] \leq \bar{r}.$$

This holds because if otherwise, $\boldsymbol{p}^*$ can not be the optimal solution due to $f(\boldsymbol{p}^*) > f(\boldsymbol{0})$. Given the non-negativeness of $\boldsymbol{p}^*$ and $\boldsymbol{p}^* \in \mathcal{V}_{\boldsymbol{d}}$, we obtain $\|\boldsymbol{p}^*\| \leq \|\boldsymbol{p}^*\|_1$, and thus $\|\boldsymbol{p}^*\| \leq \frac{\bar{r}}{\underline{d}}$.

Similarly, by the optimality of $\boldsymbol{p}_t^*$ and its associated $\boldsymbol{d}_t \in \mathcal{D}$ in Lemma A.1, we know

$$(\underline{d} - \delta)\|\boldsymbol{p}_t\|_1 \leq \boldsymbol{d}_t^\top \boldsymbol{p}_t \leq \mathbb{E}[r] \leq \bar{r},$$

so we get the bound as $\|\boldsymbol{p}_t\| \leq \frac{\bar{r}}{\underline{d}-\delta}$.

Proof of Lemma A.4. According to the latest results in (Jiang et al., 2022), the online dual price $\boldsymbol{p}_t$ achieves a sublinear convergence $\mathcal{O}(\frac{1}{\sqrt{t}})$ to $\boldsymbol{p}_t^*$. Since

$$\begin{aligned}
\boldsymbol{p}^* &\in \arg\min_{\boldsymbol{p}\geq \boldsymbol{0}} f(\boldsymbol{p}) := \boldsymbol{d}^\top \boldsymbol{p} + \mathbb{E}\left[(r - \boldsymbol{a}^\top \boldsymbol{p})^+\right], \\
\boldsymbol{p}_t^* &\in \arg\min_{\boldsymbol{p}_t\geq \boldsymbol{0}} f_{\boldsymbol{d}_t}(\boldsymbol{p}_t) := \boldsymbol{d}_t^\top \boldsymbol{p}_t + \mathbb{E}\left[(r - \boldsymbol{a}^\top \boldsymbol{p}_t)^+\right],
\end{aligned}$$

By *Lemma 12* in (Li & Ye, 2022), we have $\|\boldsymbol{p}_t^* - \boldsymbol{p}^*\|_2^2 \leq \frac{1}{\nu^2\lambda^2}\|\boldsymbol{d}_t - \boldsymbol{d}\|_2^2$.

A.3. First-order Analysis

We present the classic first-order method (Li et al., 2020) in Algorithm 4 and the first-order method with restart strategy (Gao et al., 2024) in Algorithm 5. We also show their properties of boundedness and dual convergence results.

Algorithm 4 First-Order Online algorithm

Input: total resource $\boldsymbol{b}$, time horizon T, average resource $\boldsymbol{d} = \boldsymbol{b}/T$, and initial dual price $\boldsymbol{p}_1 = \boldsymbol{0}$.
for t = 1 to T **do**
  Observe $(r_t, \boldsymbol{a}_t)$ and make decision x_t based on rule (3).
  Update learning rate $\alpha_t = \mathcal{O}(1/\sqrt{T})$.
  Compute subgradient and obtain dual price $\boldsymbol{p}_{t+1}$:

$$\boldsymbol{p}_{t+1} = \operatorname*{argmin}_{\boldsymbol{p}\geq \boldsymbol{0}} (\boldsymbol{d} - \boldsymbol{a}_t x_t)^\top \boldsymbol{p} + \frac{1}{2\alpha_t}\|\boldsymbol{p} - \boldsymbol{p}_t\|^2$$

end

Algorithm 5 First-Order Restart algorithm

Input: total resource $\boldsymbol{b}$, time horizon T, average resource $\boldsymbol{d} = \boldsymbol{b}/T$, and initial dual price $\boldsymbol{p}_1 = \mathbf{0}, \boldsymbol{p}_1^L = \mathbf{0}$
for t = 1 to T **do**
Observe $(r_t, \boldsymbol{a}_t)$ and make decision x_t based on rule (4)
Update learning rate $\alpha_t = \begin{cases} \mathcal{O}(1/T^{1/3}) & t \le T^{2/3} \\ \mathcal{O}(1/T^{2/3}) & t > T^{2/3} \end{cases}$
Compute subgradient and obtain dual price $\boldsymbol{p}_{t+1}$:

$$\boldsymbol{p}_{t+1} = \underset{\boldsymbol{p} \ge \mathbf{0}}{\operatorname{argmin}} \ (\boldsymbol{d} - \boldsymbol{a}_t x_t)^\top \boldsymbol{p} + \frac{1}{2\alpha_t} \|\boldsymbol{p} - \boldsymbol{p}_t\|^2$$

Run subgradient method with stepsize $\alpha_t = \mathcal{O}(1/t)$ and update $\{\boldsymbol{p}_t^L\}$. At $t = T^{2/3}$, restart $\boldsymbol{p}_{T^{2/3}} = \boldsymbol{p}_{T^{2/3}}^L$.
end

Lemma A.5 (Boundedness of first-order result)**.** *The online dual price $\mathbf{p}_t$ from Algorithm 4 is bounded as*

$$\|\boldsymbol{p}_t\| \le \frac{2\bar{r} + m(\bar{a} + \bar{d})^2}{\underline{d}} + m(\bar{a} + \bar{d}).$$

Lemma A.6 (Dual convergence of First-order algorithm)**.** *Under Assumptions 2.1 and 2.2, $\mathbf{p}_t$ represents the online solution from first-order method, if $\alpha_t < \nu\lambda$, the subgradient updates satisfy the following recursion rule:*

$$\mathbb{E}\left[\|\boldsymbol{p}_{t+1} - \boldsymbol{p}^*\|^2\right] \le (1 - \alpha_t \nu\lambda)\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 + \alpha_t^2 m(\bar{a} + \bar{d})^2. \tag{15}$$

Case 1. *if $\alpha_t \equiv \alpha < \frac{1}{\nu\lambda}$, then there exists a constant $C_{fo} = \frac{\bar{p}^2 + m(\bar{a}+\bar{d})^2}{\nu\lambda}$ such that*

$$\mathbb{E}\left[\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2\right] \le C_{fo}\left(\frac{1}{\alpha t} + \alpha\right). \tag{16}$$

Case 2. *if $\alpha_t = \frac{2}{\nu\lambda(t+1)}$, then there exists a constant $C_{fo} = \frac{4m(\bar{a}+\bar{d})^2}{\nu^2\lambda^2}$ such that*

$$\mathbb{E}\left[\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2\right] \le \frac{C_{fo}}{t}. \tag{17}$$

This lemma shows that the first-order method (Gao et al., 2024) guarantees the convergence of $\boldsymbol{p}_t$ to $\boldsymbol{p}^*$. It also highlights a key trade-off in choosing the learning rate α_t, an aspect that will be central to our later optimality analysis. It also highlights a key trade-off in choosing the learning rate α_t, an aspect that will be central to our later optimality analysis. With this groundwork in place, we now move on to bounding the total regret of our algorithm.

Proof of Lemma A.5. As our initial choice $\boldsymbol{p}_1 = \mathbf{0}$, according to *Lemma 1* in (Li et al., 2020), we get the above result. In addition, by *Lemma B.1* in (Gao et al., 2024), we have

$$\|\boldsymbol{p}_t\| \le \frac{\bar{r}}{\underline{d}} + \frac{m(\bar{a} + \bar{d})^2 \alpha_t}{2\underline{d}} + \alpha_t\sqrt{m}(\bar{a} + \bar{d}).$$

Proof of Lemma A.6. Based on the updated rule in Algorithm 4, we derive:

$$\begin{aligned} \|\boldsymbol{p}_{t+1} - \boldsymbol{p}^*\|^2 &\le \|\boldsymbol{p}_t - \alpha_t(\boldsymbol{d} - \boldsymbol{a}_t x_t) - \boldsymbol{p}^*\|^2 \\ &= \|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 - 2\alpha_t\langle \boldsymbol{d} - \boldsymbol{a}_t x_t, \boldsymbol{p}_t - \boldsymbol{p}^*\rangle + \alpha_t^2\|\boldsymbol{d} - \boldsymbol{a}_t x_t\|^2 \\ &\le \|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 - 2\alpha_t\langle \boldsymbol{d} - \boldsymbol{a}_t x_t, \boldsymbol{p}_t - \boldsymbol{p}^*\rangle + \alpha_t^2 m(\bar{a} + \bar{d})^2. \end{aligned}$$

1) With convexity of f and $\mathbb{E}[\boldsymbol{d} - \boldsymbol{a}_t x_t] \in \partial f(\boldsymbol{p}_t)$, we have:

$$f(\boldsymbol{p}^*) \ge f(\boldsymbol{p}_t) + \langle \boldsymbol{d} - \boldsymbol{a}_t x_t, \boldsymbol{p}_t - \boldsymbol{p}^*\rangle.$$

2) With quadratic regularity of f as in Lemma A.2, we have:

$$f(\boldsymbol{p}_t) \geq f(\boldsymbol{p}^*) + \nabla f(\boldsymbol{p}^*)^\top(\boldsymbol{p}_t - \boldsymbol{p}^*) + \frac{\nu\lambda}{2}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2,$$
$$\text{which indicates} \quad f(\boldsymbol{p}_t) - f(\boldsymbol{p}^*) \geq \frac{\nu\lambda}{2}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2.$$

Combine the above results and take expectation conditioned on history information $\{(r_j, \boldsymbol{a}_j), j \leq t\}$, we obtain:

$$\begin{aligned}
\mathbb{E}\|\boldsymbol{p}_{t+1} - \boldsymbol{p}^*\|^2 &\leq \|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 - 2\alpha_t\mathbb{E}[\langle \boldsymbol{d} - \boldsymbol{a}_t x_t, \boldsymbol{p}_t - \boldsymbol{p}^*\rangle] + \alpha_t^2 m(\bar{a} + \bar{d})^2 \\
&\leq \|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 - 2\alpha_t(f(\boldsymbol{p}_t) - f(\boldsymbol{p}^*)) + \alpha_t^2 m(\bar{a} + \bar{d})^2 \\
&\leq \|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 - \alpha_t\nu\lambda\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 + \alpha_t^2 m(\bar{a} + \bar{d})^2 \\
&= (1 - \alpha_t\nu\lambda)\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 + \alpha_t^2 m(\bar{a} + \bar{d})^2.
\end{aligned}$$

This proves (15) with the general case of learning rate α_t.

Case 1. When $\alpha_t = \alpha < \frac{1}{\nu\lambda}$ is a constant, we take recursion of (15). Note that $(1 - \nu\lambda\alpha)^t < 1/\nu\lambda\alpha t$, we get:

$$\begin{aligned}
\mathbb{E}\|\boldsymbol{p}_{t+1} - \boldsymbol{p}^*\|^2 &\leq (1 - \alpha\nu\lambda)^t\|\boldsymbol{p}_1 - \boldsymbol{p}^*\|^2 + \sum_{j=0}^{t-1}\alpha^2 m(\bar{a} + \bar{d})^2(1 - \alpha\nu\lambda)^j \\
&\leq \frac{\|\boldsymbol{p}_1 - \boldsymbol{p}^*\|^2}{\nu\lambda\alpha t} + \frac{m(\bar{a} + \bar{d})^2}{\nu\lambda}\alpha.
\end{aligned}$$

Since all feasible $\boldsymbol{p} \in \mathcal{V}_{\boldsymbol{p}}$ is bounded, let $C_{fo} = \frac{\bar{p}^2 + m(\bar{a}+\bar{d})^2}{\nu\lambda}$, we are able to obtain (16).

Case 2. When $\alpha_t = \frac{2}{\nu\lambda(t+1)}$, for any $j \leq t$, Lemma A.6 gives us

$$\mathbb{E}\|\boldsymbol{p}_{j+1} - \boldsymbol{p}^*\|^2 \leq \frac{j-1}{j+1}\|\boldsymbol{p}_j - \boldsymbol{p}^*\|^2 + \frac{4m(\bar{a} + \bar{d})^2}{\nu^2\lambda^2(j+1)^2}.$$

Re-arranging the above inequality, we have

$$(j+1)^2\mathbb{E}\|\boldsymbol{p}_{j+1} - \boldsymbol{p}^*\|^2 \leq (j^2 - 1)\|\boldsymbol{p}_j - \boldsymbol{p}^*\|^2 + \frac{4m(\bar{a} + \bar{d})^2}{\nu^2\lambda^2},$$
$$\text{then} \quad (j+1)^2\mathbb{E}\|\boldsymbol{p}_{j+1} - \boldsymbol{p}^*\|^2 - j^2\|\boldsymbol{p}_j - \boldsymbol{p}^*\|^2 \leq \frac{4m(\bar{a} + \bar{d})^2}{\nu^2\lambda^2}.$$

Then by telescoping from $j = 2$ to t, we have

$$\sum_{j=1}^{t}(j+1)^2\mathbb{E}\|\boldsymbol{p}_{j+1} - \boldsymbol{p}^*\|^2 - j^2\|\boldsymbol{p}_j - \boldsymbol{p}^*\|^2 = (t+1)^2\mathbb{E}\|\boldsymbol{p}_{t+1} - \boldsymbol{p}^*\|^2 - \|\boldsymbol{p}_1 - \boldsymbol{p}^*\|^2 \leq \frac{4m(\bar{a} + \bar{d})^2 t}{\nu^2\lambda^2}$$

which then gives us

$$\mathbb{E}\|\boldsymbol{p}_{t+1} - \boldsymbol{p}^*\|^2 \leq \frac{\|\boldsymbol{p}_1 - \boldsymbol{p}^*\|^2}{(t+1)^2} + \frac{4m(\bar{a} + \bar{d})^2 t}{\nu^2\lambda^2(t+1)^2},$$
$$\text{thus} \quad \mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 \leq \frac{\|\boldsymbol{p}_1 - \boldsymbol{p}^*\|^2}{t^2} + \frac{4m(\bar{a} + \bar{d})^2}{\nu^2\lambda^2 t}.$$

As $\boldsymbol{p}_1 = \boldsymbol{0}$ and $\boldsymbol{p}^*$ is bounded, taking $C_{fo} = \frac{4m(\bar{a}+\bar{d})^2}{\nu^2\lambda^2}$ completes the proof for (17).

B. Algorithm Design and Analysis

In this section, we present a cohesive mathematical framework to integrate the LP-based method and first-order method by developing a new performance metric. We also provide some unique properties for our algorithm design of parallel multi-phase structure.

As stated in Algorithm 1, we construct independent paths for online learning and online decision-making conducted simultaneously. Specifically, for the online learning path, we use the LP-based method to resolve updated linear programs at a fixed frequency and send this result to the decision-making path; For the online decision-making path, we apply the first-order method in the initial and final batches, and use the latest dual price for decision-making in the intermediate resolving intervals. This process is illustrated in Figure 1.

B.1. Regret Analysis

First, we construct an upper bound for the offline optimal objective value. The challenge comes from the intractable dependency of constraints on objective value under the online setting. We tackle this issue by introducing a Lagrangian function to integrate constraints into the objective and balance revenue maximization with constraint satisfaction. The formulation is stated as follows.

Lemma B.1 (Lagrangian Upper Bound)**.** *Under Assumptions (2.1) and (2.2), define the deterministic Lagrangian dual function as*

$$\ell(\boldsymbol{p}) \coloneqq \mathbb{E}\left[rI(r > \boldsymbol{a}^\top \boldsymbol{p}) + (\boldsymbol{d} - \boldsymbol{a}I(r > \boldsymbol{a}^\top \boldsymbol{p}))^\top \boldsymbol{p}^*\right].$$

Then for any feasible $\boldsymbol{p} \in \mathcal{V}_{\boldsymbol{p}}$, we have:

$$\begin{aligned} &(a) \quad \mathbb{E}\left[\sum_{t=1}^{T} r_t x_t^*\right] \le T\ell(\boldsymbol{p}^*), \\ &(b) \quad \ell(\boldsymbol{p}^*) - \ell(\boldsymbol{p}) \le \mu\bar{a}^2 \|\boldsymbol{p} - \boldsymbol{p}^*\|^2. \end{aligned} \tag{18}$$

Lemma B.2 (Dual Price Boundedness)**.** *Under Assumption 2.1, the online and offline optimal dual prices are bounded respectively by $\|\boldsymbol{p}^*\| \le \frac{\bar{r}}{\underline{d}}$ and*

$$\|\boldsymbol{p}_t\| \le \bar{p} \coloneqq \max\left(\frac{\bar{r}}{\underline{d} - \delta}, \frac{2\bar{r} + m(\bar{a} + \bar{d})^2}{\underline{d}} + m(\bar{a} + \bar{d})\right).$$

This lemma establishes that the optimal dual prices remain bounded. Our algorithm maintains these bounds because if $\boldsymbol{p}_t$ grows large, the algorithm responds by accepting more orders, which in turn reduces $\boldsymbol{p}_{t+1}$. This self-correcting mechanism keeps the dual prices within these limits. With this groundwork in place, we now move on to derive the upper bound for the total performance of Algorithm 1.

Theorem B.3 (Decomposition of Regret)**.** *Under Assumptions (2.1) and (2.2), let $C_r = \max\{\frac{\bar{r}}{\underline{d}}, \frac{\nu\lambda\bar{a}^2}{2}\}$, we derive an upper bound for the regret $r(\boldsymbol{x})$ as:*

$$r(\boldsymbol{x}) \le C_r\left[\mathbb{E}\left\|\left(\boldsymbol{b} - \sum_{t=1}^{T} \boldsymbol{a}_t x_t\right)^{B^+}\right\| + \sum_{t=1}^{T} \mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2\right]. \tag{19}$$

Proof of Lemma B.1. This lemma is proved as *Lemma 3* in (Li & Ye, 2022) by strong duality and optimality of $\boldsymbol{p}_T^*$.

Proof of Lemma B.2. By the boundedness results in Lemma A.3 and Lemma A.5, define

$$\bar{p} \coloneqq \max\left(\frac{\bar{r}}{\underline{d} - \delta}, \frac{2\bar{r} + m(\bar{a} + \bar{d})^2}{\underline{d}} + m(\bar{a} + \bar{d})\right),$$

then we complete the proof.

Proof of Theorem B.3. By the definition of $\ell(\boldsymbol{p})$, we derive the online objective as

$$\begin{aligned}\mathbb{E}\Big[\sum_{t=1}^{T} r_t x_t\Big] &= \sum_{t=1}^{T} \mathbb{E}[\mathbb{E}[r_t x_t | \boldsymbol{p}_t]] \\ &= \sum_{t=1}^{T} \mathbb{E}[\ell(\boldsymbol{p}_t) - (\boldsymbol{d} - \boldsymbol{a}_t x_t)^\top \boldsymbol{p}^*].\end{aligned}$$

Then by Lemma B.1, we derive the upper bound for the regret $r(\boldsymbol{x})$ as defined in (6) to be:

$$\begin{aligned} r(\boldsymbol{x}) &= \mathbb{E}\Big[\sum_{t=1}^{T} r_t x_t^* - r_t x_t\Big] \\ &\le T\ell(\boldsymbol{p}^*) - \sum_{t=1}^{T} \mathbb{E}[\ell(\boldsymbol{p}_t) - (\boldsymbol{d} - \boldsymbol{a}_t x_t)^\top \boldsymbol{p}^*] \\ &= \sum_{t=1}^{T} \mathbb{E}[(\boldsymbol{d} - \boldsymbol{a}_t x_t)^\top \boldsymbol{p}^*] + \sum_{t=1}^{T} \mathbb{E}[\ell(\boldsymbol{p}^*) - \ell(\boldsymbol{p}_t)] \\ &\le \mathbb{E}\Big[\Big(\boldsymbol{b} - \sum_{t=1}^{T} \boldsymbol{a}_t x_t\Big)^\top \boldsymbol{p}^*\Big] + \frac{\mu \bar{a}^2}{2} \sum_{t=1}^{T} \mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 \\ &\le \|\boldsymbol{p}^*\| \cdot \mathbb{E}\Big\|\Big(\boldsymbol{b} - \sum_{t=1}^{T} \boldsymbol{a}_t x_t\Big)^{B^+}\Big\| + \frac{\mu \bar{a}^2}{2} \sum_{t=1}^{T} \mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 \end{aligned}$$

where $(\cdot)^{B+}$ denotes the positive part only for binding constraints.

As $\|\boldsymbol{p}^*\| \le \frac{\bar{r}}{\underline{d}}$ is bounded, taking the constant $C_r = \max\{\frac{\bar{r}}{\underline{d}}, \frac{\mu \bar{a}^2}{2}\}$ completes the proof.

B.2. Total Performance Analysis

We consider the constraint violation as $v(\boldsymbol{x}) = \|(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^+\|$ in (7). Then the total performance is the combination of regret and constraint violation. Since Theorem B.3 holds for any unknown distribution $\mathcal{P} \in \Xi$, we have:

$$\begin{aligned} \Delta_T &= \sup_{\mathcal{P} \in \Xi} \mathbb{E}_{\mathcal{P}}[r(\boldsymbol{x}) + v(\boldsymbol{x})] \\ &\le C_r\Big[\mathbb{E}\Big\|\Big(\boldsymbol{b} - \sum_{t=1}^{T} \boldsymbol{a}_t x_t\Big)^{B^+}\Big\| + \sum_{t=1}^{T} \mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2\Big] + \mathbb{E}\Big\|\Big(\sum_{t=1}^{T} \boldsymbol{a}_t x_t - \boldsymbol{b}\Big)^+\Big\| \\ &\le C_r\Big[\sum_{t=1}^{T} \mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 + \mathbb{E}\big\|(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})^{B^+}\big\| + \mathbb{E}\big\|(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^+\big\|\Big]. \end{aligned} \tag{20}$$

This new framework is adaptive to all OLP methods. We proceed with this structure to analyze various OLP methods.

Theorem B.4 (Horizon Division)**.** *Based on the change of methods, we separate the horizon into three intervals T_1, T_2, and T_3, and reorganize the total performance for analysis to be*

$$\Delta_T \le C_r\big[\Delta_{T_1} + \Delta_{T_2} + \Delta_{T_3}\big]. \tag{21}$$

where $\Delta_{T_1}, \Delta_{T_2}$, and Δ_{T_3} are the performances for each interval.

Proof of Theorem B.4. We split the total horizon into three intervals of initial batch $T_1 = [0, f]$, intermediate process $T_2 = [f, kf]$, and final batch $T_3 = [kf, T]$. We use the first-order method for the first and final batches and the LP-based method for intermediate processes. Notice that for the initial resolving batch, we maintain the classical analysis for the

first-order method since it has not reached the mixture with the LP-based method. Then based on (20), we define

$$\begin{aligned}
\Delta_{T_1} &:= \sum_{t=1}^{f} \mathbb{E}[r_t x_t^* - r_t x_t] + \mathbb{E}\Big\| \Big(\sum_{t=1}^{f} \boldsymbol{a}_t x_t - f\boldsymbol{d} \Big)^+ \Big\| \\
\Delta_{T_2} &:= \sum_{t=f}^{kf} \mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 + \mathbb{E}\Big\| \Big((k-1)f\boldsymbol{d} - \sum_{t=f}^{kf} \boldsymbol{a}_t x_t \Big)^{B^+} \Big\| + \mathbb{E}\Big\| \Big(\sum_{t=f}^{kf} \boldsymbol{a}_t x_t - (k-1)f\boldsymbol{d} \Big)^+ \Big\| \\
\Delta_{T_3} &:= \sum_{t=kf}^{T} \mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 + \mathbb{E}\Big\| \Big((T-kf)\boldsymbol{d} - \sum_{t=kf}^{T} \boldsymbol{a}_t x_t \Big)^{B^+} \Big\| + \mathbb{E}\Big\| \Big(\sum_{t=kf}^{T} \boldsymbol{a}_t x_t - (T-kf)\boldsymbol{d} \Big)^+ \Big\|
\end{aligned} \tag{22}$$

Then we have

$$\Delta_T \le C_r \big[\Delta_{T_1} + \Delta_{T_2} + \Delta_{T_3} \big]$$

since $\|(a+b)^+\| \le \|a^+\| + \|b^+\|$ for any a, b.

B.3. Regret for LP-based Method

In this section, we are going to bound Δ_{T_2} for the regret from LP-based method. To provide a tighter analysis, we examine the real-time average resource capacity $\boldsymbol{d}_t$ instead of the original process $\boldsymbol{b}_t$. Specifically, by Lemma A.1, we constraint $\boldsymbol{d}_t$ within a feasible range $\mathcal{D}$. If $\boldsymbol{d}_t \notin \mathcal{D}$, we set the dual price to zero to accept all subsequent orders. We analyze the remaining resources and constraint violation due to this strategy. Note that this is required solely for rigor analysis purposes and is not necessary when running the algorithm.

Lemma B.5 (Dynamics of Resource Usage)**.** *Under Assumption 2.1 and 2.2, there exists a constant $C > 0$ depending on $\bar{d}, \bar{a}, m, \nu, \lambda_1, \mu$, and C_{lp} such that*

$$\sum_{t=f}^{kf} \mathbb{E}[(d_{i,t} - d_i)^2] \le C \log(k). \tag{23}$$

Theorem B.6 (Regret of Intermediate Process)**.** *Under Assumption 2.1 and 2.2, following the result in (22), we prove the regret satisfies*

$$\Delta_{T_2} \le \log\left(\frac{T}{f}\right) = \mathcal{O}(\log(k)). \tag{24}$$

Proof of Lemma B.5. During the re-solving process, we denote each interval to be $[jf, (j+1)f]$ where $j = 1, 2, ..., k$. As the resource usage follows $\boldsymbol{b}_{t+1} = \boldsymbol{b}_t - \boldsymbol{a}_{t+1} I(r_{t+1} > \boldsymbol{a}_{t+1}^T \boldsymbol{p}_{t+1})$, normalizing both sides, we derive the update of average resource consumption to be:

$$d_{i,(j+1)f} = d_{i,jf} + \frac{\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf} - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(k+1)f})}{T - (j+1)f}.$$

Subtracting d_i on both sides, it becomes

$$\begin{aligned}
d_{i,(j+1)f} - d_i &= d_{i,jf} - d_i + \frac{\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf} - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})}{T - (j+1)f} \\
&= d_{i,jf} - d_i + \frac{\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf} - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{jf}^*)}{T - (j+1)f} \\
&\quad + \frac{\sum_{\ell=jf+1}^{(j+1)f} a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{jf}^*) - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})}{T - (j+1)f}.
\end{aligned}$$

Taking expectations of squares, we have

$$\begin{aligned}\mathbb{E}(d_{i,(j+1)f}-d_i)^2 =&\mathbb{E}(d_{i,jf}-d_i)^2+\mathbb{E}\left[\frac{(\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf}-a_{i,\ell}I(r_\ell>\boldsymbol{a}_\ell^\top \boldsymbol{p}_{jf}^*))^2}{(T-(j+1)f)^2}\right]\\ &+\mathbb{E}\left[\frac{(\sum_{\ell=jf+1}^{(j+1)f} a_{i,\ell}I(r_\ell>\boldsymbol{a}_\ell^\top \boldsymbol{p}_{jf}^*)-a_{i,\ell}I(r_\ell>\boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f}))^2}{(T-(j+1)f)^2}\right]\\ &+2\mathbb{E}\left[(d_{i,jf}-d_i)\left(\frac{\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf}-a_{i,\ell}I(r_\ell>\boldsymbol{a}_\ell^\top \boldsymbol{p}_{jf}^*)}{T-(j+1)f}\right)\right]\\ &+2\mathbb{E}\left[(d_{i,jf}-d_i)\left(\frac{\sum_{\ell=jf+1}^{(j+1)f} a_{i,\ell}I(r_\ell>\boldsymbol{a}_\ell^\top \boldsymbol{p}_{jf}^*)-a_{i,\ell}I(r_\ell>\boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})}{T-(j+1)f}\right)\right]\\ &+2\mathbb{E}\left[\frac{\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf}-a_{i,\ell}I(r_\ell>\boldsymbol{a}_\ell^\top \boldsymbol{p}_{jf}^*)}{T-(j+1)f}\cdot\frac{\sum_{\ell=jf+1}^{(j+1)f} a_{i,\ell}I(r_\ell>\boldsymbol{a}_\ell^\top \boldsymbol{p}_{jf}^*)-a_{i,\ell}I(r_\ell>\boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})}{T-(j+1)f}\right].\end{aligned} \tag{25}$$

By Lemma D.1, we obtain the recursion relation as

$$\mathbb{E}(d_{i,(j+1)f}-d_i)^2\le\mathbb{E}(d_{i,jf}-d_i)^2+\frac{C_{rec}}{(k-j-1)^2 f}+\frac{4\mu\bar{a}^2\sqrt{C_{lp}}}{(k-j-1)\sqrt{(j+1)f}}\sqrt{\mathbb{E}[(d_{i,jf}-d_i)^2]} \tag{26}$$

where $C_{rec}>0$ is a constant defined in Lemma D.1.

Then according to Lemma D.2, take $C=12\max\{C_{rec},16\mu^2\bar{a}^4C_{lp}\}$, we solve recursion (26) and obtain the upper bound of total deviation from original d_i in the re-solving process to be:

$$\sum_{j=1}^{k}\mathbb{E}\left[(d_{i,jf}-d_i)^2\right]\le\frac{C}{f}\log(k). \tag{27}$$

Therefore, we sum the whole re-solving process and obtain:

$$\begin{aligned}\sum_{t=f}^{kf}\mathbb{E}[(d_{i,t}-d_i)^2]&=\sum_{j=1}^{k}\sum_{\ell=jf}^{(j+1)f}\mathbb{E}[(d_{i,\ell}-d_i)^2]\\ &=f\cdot\frac{C}{f}\log(k)\quad\text{(by result in (27)}\\ &\le C\log(k).\end{aligned} \tag{28}$$

This completes the proof.

Proof of Theorem B.6. We analyze the three terms in (22) respectively.

1. For dual convergence, by Lemma A.4, we have

$$\begin{aligned}\sum_{t=f}^{kf}\mathbb{E}\|\boldsymbol{p}_t-\boldsymbol{p}^*\|^2&=\mathbb{E}\left[\sum_{j=1}^{k-1}\sum_{t=jf+1}^{(j+1)f}\|\boldsymbol{p}_t-\boldsymbol{p}^*\|_2^2\right]\\ &=\sum_{j=1}^{k-1}f\,\mathbb{E}\left[\|\boldsymbol{p}_{(j+1)f}-\boldsymbol{p}^*\|_2^2\right]\\ &\le\sum_{j=1}^{k-1}f\,\mathbb{E}\left[\|\boldsymbol{p}_{(j+1)f}-\boldsymbol{p}_{jf}^*\|_2^2+\|\boldsymbol{p}_{jf}^*-\boldsymbol{p}^*\|_2^2\right]\end{aligned}$$

$$\leq \sum_{j=1}^{k-1} f\Big[\frac{C_{lp}}{jf} + \frac{1}{\nu^2\lambda^2}\,\mathbb{E}\big[(\boldsymbol{d}_{jf} - \boldsymbol{d})^2\big]\Big] \quad \text{(by Lemma A.4)}$$

$$\leq C_{lp}\log(k) + \frac{mC}{\nu^2\lambda^2}\log(k). \quad \text{(by Lemma B.5).} \tag{29}$$

2. For the constraint violation, by Chebyshev's inequality, we bound the probability as

$$\sum_{t=f}^{kf} \mathbb{P}(|d_{i,t} - d_i| \leq \delta) \leq \sum_{t=f}^{kf} \frac{\mathbb{E}[(d_{i,t} - d_i)^2]}{\delta^2}$$

$$\leq \frac{C}{\delta^2}\log(k). \quad \text{(by Lemma B.5)} \tag{30}$$

Since our strategy is to automatically set $\boldsymbol{p}_t = 0$ once $\boldsymbol{d}_t \notin \mathcal{D}$, we will accept all the subsequent orders according to our decision condition (3). Denote R as the total number of orders processed specifically due to this rule, then we have:

$$\mathbb{E}[R] \leq \sum_{t=f}^{kf} \mathbb{P}(|d_{i,t} - d_i| \leq \delta) \leq \frac{C}{\delta^2}\log(k) \tag{31}$$

which indicates that the resource violation is at most

$$\mathbb{E}\Big\|\Big(\sum_{t=f}^{kf} \boldsymbol{a}_t x_t - (k-1)f\boldsymbol{d}\Big)^+\Big\| \leq \mathbb{E}[\bar{a}R] \leq \frac{C\bar{a}}{\delta^2}\log(k). \tag{32}$$

3. For the remaining resource, we derive its upper bound by considering the opposite extreme case of our strategy. If we reject all those R orders, then no resource is used on them. Thus, by (31), the positive projection of remaining resource at $t = kf$ is at most:

$$\mathbb{E}\Big\|\Big((k-1)f\boldsymbol{d} - \sum_{t=f}^{kf} \boldsymbol{a}_t x_t\Big)^{B^+}\Big\| \leq \mathbb{E}[(d_i + \delta)R] \leq \frac{C(\bar{d} + \delta)}{\delta^2}\log(k). \tag{33}$$

Combining the results of (29), (32), and (33), we derive the final result as

$$\Delta_{T_2} \leq \Big(C_{lp} + \frac{mC}{\nu^2\lambda^2} + (\bar{a} + \bar{d} + \delta)\frac{C}{\delta^2}\Big)\log(k). \tag{34}$$

This completes the proof.

B.4. Regret for First-Order Method

In this section, we are going to bound Δ_{T_1} and Δ_{T_3} for the regret from the first-order method. The key point is to analyze the connection between dual prices based on the gradient-updated rule. Here $\boldsymbol{p}_t = \boldsymbol{p}_t^{fo}$.

Theorem B.7 (Regret of Initial Batch)**.** *Under Assumption 2.1 and 2.2, we have*

$$\Delta_{T_1} \leq \left(\frac{m(\bar{a} + \bar{d})^2}{2} + \frac{2\bar{r} + m(\bar{a} + \bar{d})^2}{\underline{d}} + m(\bar{a} + \bar{d})\right)\sqrt{f}. \tag{35}$$

Lemma B.8 (Constraint Violation)**.** *Under Assumption 2.1 and 2.2, take the learning rate $\alpha_t = \alpha$, we bound the constraint violation for the last batch as*

$$\mathbb{E}\Big\|\Big(\sum_{t=kf}^{T} \boldsymbol{a}_t x_t - (T - kf)\boldsymbol{d}\Big)^+\Big\| \leq \Big(\frac{C_{lp}}{\sqrt{kf}} + \frac{C_b}{T^2}\Big)\cdot\frac{1}{\alpha} + \frac{C_b}{\sqrt{(T-kf)kf}}\cdot\frac{1}{\alpha\sqrt{\alpha}} + \frac{C_b}{\sqrt{\alpha}}. \tag{36}$$

Theorem B.9 (Regret of Final Batch)**.** *Under Assumption 2.1 and 2.2, we have*

$$\Delta_{T_3} \leq \left(\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} + \frac{2C_{lp}}{\sqrt{k}} + \frac{C_{lp}\log(f)}{\nu\lambda k} + \frac{4C_b}{\sqrt{k}} + 2C_b\right)\cdot f^{1/3}. \tag{37}$$

Lemma B.10 (Warm-Start First-Order Regret)**.** *Assume that we have a warm start for the initial dual price $\boldsymbol{p}_0$ in the first batch, which satisfies $\|\boldsymbol{p}_0 - \boldsymbol{p}^*\| \leq f^{-1/3}$. Then the regret of the first batch follows*

$$\Delta_{T_1 warm} \leq \left(\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} + 4C_w\right)\cdot f^{1/3} + \log(f). \tag{38}$$

Proof of Theorem B.7. We analyze Δ_{T_1} in (22) by parts. By *Lemma B.3, B.4* in (Gao et al., 2024), we have

$$\sum_{t=1}^{f}\mathbb{E}[r_t x_t^* - r_t x_t] \leq \frac{m(\bar{a}+\bar{d})^2}{2}\alpha f,$$
$$\mathbb{E}\Big\|\Big(\sum_{t=1}^{f}\boldsymbol{a}_t x_t - f\boldsymbol{d}\Big)^+\Big\| \leq \frac{1}{\alpha}\left[\frac{2\bar{r} + m(\bar{a}+\bar{d})^2}{\underline{d}} + m(\bar{a}+\bar{d})\right].$$

To achieve the tight upper bound, we select the optimal step size $\alpha = \frac{1}{\sqrt{f}}$, which then gives us the desired result.

Proof of Lemma B.8. With the update rule of first-order method in Algorithm 4, we have

$$\boldsymbol{p}_{t+1} = [\boldsymbol{p}_t - \alpha(\boldsymbol{d} - \boldsymbol{a}_t x_t)]^+ \geq \boldsymbol{p}_t - \alpha(\boldsymbol{d} - \boldsymbol{a}_t x_t),$$
$$\text{which gives us} \quad \boldsymbol{a}_t x_t - \boldsymbol{d} \leq \frac{1}{\alpha}(\boldsymbol{p}_{t+1} - \boldsymbol{p}_t).$$

Summarizing on both sides and applying the telescoping, we derive

$$\sum_{t=kf}^{T}(\boldsymbol{a}_t x_t - \boldsymbol{d}) \leq \frac{1}{\alpha}\sum_{t=kf}^{T}(\boldsymbol{p}_{t+1} - \boldsymbol{p}_t) = \frac{1}{\alpha}(\boldsymbol{p}_{T+1} - \boldsymbol{p}_{kf}).$$

By Lemma D.3, take $C_b = \max\{\frac{C_{lp}}{\sqrt{\nu\lambda}}, \frac{m(\bar{a}+\bar{d})}{\sqrt{\nu\lambda}}, \bar{p}\}$, we have

$$\mathbb{E}\|\boldsymbol{p}_{T+1} - \boldsymbol{p}^*\| \leq C_b\Big[\frac{1}{\sqrt{(T-kf)kf}}\cdot\frac{1}{\sqrt{\alpha}} + \sqrt{\alpha} + \frac{1}{T^2}\Big]. \tag{39}$$

Thus, according to Lemma A.4 and (39), the expectation of constraint violation satisfies

$$\begin{aligned}
\mathbb{E}\Big\|\Big(\sum_{t=kf}^{T}(\boldsymbol{a}_t x_t - \boldsymbol{d})\Big)^+\Big\| &\leq \mathbb{E}\Big\|\sum_{t=kf}^{T}(\boldsymbol{a}_t x_t - \boldsymbol{d})\Big\| \\
&\leq \frac{1}{\alpha}\mathbb{E}\|\boldsymbol{p}_{T+1} - \boldsymbol{p}_{kf}\| \\
&\leq \frac{1}{\alpha}\mathbb{E}\Big[\|\boldsymbol{p}_{kf} - \boldsymbol{p}^*\| + \|\boldsymbol{p}_{T+1} - \boldsymbol{p}^*\|\Big] \\
&\leq \frac{1}{\alpha}\Big(\frac{C_{lp}}{\sqrt{kf}} + C_b\Big[\frac{1}{\sqrt{(T-kf)kf}}\cdot\frac{1}{\sqrt{\alpha}} + \sqrt{\alpha} + \frac{1}{T^2}\Big]\Big) \\
&\leq \frac{C_{lp}}{\sqrt{kf}}\cdot\frac{1}{\alpha} + \frac{C_b}{\sqrt{(T-kf)kf}}\cdot\frac{1}{\alpha\sqrt{\alpha}} + C_b\cdot\frac{1}{\sqrt{\alpha}} + \frac{C_b}{T^2}\cdot\frac{1}{\alpha}.
\end{aligned} \tag{40}$$

This completes the proof.

Proof of Theorem B.9. We decompose Δ_{T_3} into three parts according to (22) and analyze each term respectively. Since the first-order method re-starts from $\boldsymbol{p}_{kf}$, by Lemma A.4, we have:

$$\begin{aligned}\sum_{t=kf}^{T}\mathbb{E}\|\boldsymbol{p}_t-\boldsymbol{p}^*\|^2 &\le \sum_{t=kf}^{T}\Big[\frac{\|\boldsymbol{p}_{kf}-\boldsymbol{p}^*\|^2}{\nu\lambda\alpha t}+\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}\alpha\Big]\\ &\le \frac{\|\boldsymbol{p}_{kf}-\boldsymbol{p}^*\|^2}{\nu\lambda\alpha}\log(T-kf)+\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}(T-kf)\alpha\\ &\le \frac{\|\boldsymbol{p}_{kf}-\boldsymbol{p}^*\|^2}{\nu\lambda\alpha}\log(f)+\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}f\alpha\\ &\le \frac{C_{lp}\log(f)}{\nu\lambda kf}\cdot\frac{1}{\alpha}+\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}f\alpha. \end{aligned}\tag{41}$$

Since $t\in[kf,T]$, by *Proposition 3.3* in (Gao et al., 2024), the dual price is far from the origin and thus the positive projection still equals itself. This indicates that:

$$\boldsymbol{p}_{t+1}=[\boldsymbol{p}_t-\alpha(\boldsymbol{d}-\boldsymbol{a}_t x_t)]^+=\boldsymbol{p}_t-\alpha(\boldsymbol{d}-\boldsymbol{a}_t x_t),$$

$$\text{which then gives us}\quad \boldsymbol{d}-\boldsymbol{a}_t x_t=\frac{1}{\alpha}(\boldsymbol{p}_t-\boldsymbol{p}_{t+1}).$$

Summarizing on both sides and applying the telescoping, we derive

$$\sum_{t=kf}^{T}(\boldsymbol{d}-\boldsymbol{a}_t x_t)\le\frac{1}{\alpha}\sum_{t=kf}^{T}(\boldsymbol{p}_t-\boldsymbol{p}_{t+1})=\frac{1}{\alpha}(\boldsymbol{p}_{kf}-\boldsymbol{p}_{T+1}).$$

Then for the positive binding terms, by Lemma A.4 and (39), we have

$$\begin{aligned}\mathbb{E}\Big\|\Big(\sum_{t=kf}^{T}(\boldsymbol{d}-\boldsymbol{a}_t x_t)\Big)^{B^+}\Big\| &\le \mathbb{E}\Big\|\sum_{t=kf}^{T}(\boldsymbol{d}-\boldsymbol{a}_t x_t)\Big\|\\ &\le \frac{1}{\alpha}\mathbb{E}\|\boldsymbol{p}_{kf}-\boldsymbol{p}_{T+1}\|\\ &\le \frac{1}{\alpha}\mathbb{E}\Big[\|\boldsymbol{p}_{kf}-\boldsymbol{p}^*\|+\|\boldsymbol{p}_{T+1}-\boldsymbol{p}^*\|\Big]\\ &\le \frac{C_{lp}}{\sqrt{kf}}\cdot\frac{1}{\alpha}+\frac{C_b}{\sqrt{(T-kf)kf}}\cdot\frac{1}{\alpha\sqrt{\alpha}}+C_b\cdot\frac{1}{\sqrt{\alpha}}+\frac{C_b}{T^2}\cdot\frac{1}{\alpha}. \end{aligned}\tag{42}$$

Combining the results of (41), Lemma B.8, and (42) together, we obtain the final result as:

$$\Delta_{T_3}\le\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}f\alpha+\Big(\frac{2C_{lp}}{\sqrt{kf}}+\frac{2C_b}{T^2}+\frac{C_{lp}\log(f)}{\nu\lambda kf}\Big)\cdot\frac{1}{\alpha}+\frac{2C_b}{\sqrt{(T-kf)kf}}\cdot\frac{2}{\alpha\sqrt{\alpha}}+\frac{2C_b}{\sqrt{\alpha}}.\tag{43}$$

We select the optimal learning rate $\alpha=f^{-2/3}$ to minimize Δ_{T_3} in (43). We prove this result in cases.

1. If $T-kf\le f^{1/3}$, the regret must be smaller than the length of the batch, so $\Delta_{T_3}\le T-kf\le f^{1/3}$.
2. If $T-kf>f^{1/3}$, then we use this property to bound the term $\frac{2C_b}{\sqrt{(T-kf)kf}}\cdot\frac{2}{\alpha\sqrt{\alpha}}$. We derive

$$\begin{aligned}\Delta_{T_3} &\le \frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}f\cdot f^{-2/3}+\Big(\frac{2C_{lp}}{\sqrt{kf}}+\frac{2C_b}{T^2}+\frac{C_{lp}\log(f)}{\nu\lambda kf}\Big)\cdot f^{2/3}+\frac{4C_b}{\sqrt{f^{1/3}\cdot kf}}\cdot f+2C_b\cdot f^{1/3}\\ &= \frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}f^{1/3}+\frac{2C_{lp}}{\sqrt{k}}f^{1/6}+\frac{2C_b}{T}+\frac{C_{lp}\log(f)}{\nu\lambda k}f^{1/3}+\frac{4C_b}{\sqrt{k}}\cdot f^{1/3}+2C_b\cdot f^{1/3}\\ &\le \Big(\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}+\frac{2C_{lp}}{\sqrt{k}}+\frac{C_{lp}\log(f)}{\nu\lambda k}+\frac{4C_b}{\sqrt{k}}+2C_b\Big)\cdot f^{1/3}. \end{aligned}\tag{44}$$

This completes the proof.

Proof of Lemma B.10. According to Lemma D.3, with $\|\boldsymbol{p}_0 - \boldsymbol{p}^*\| \le f^{-1/3}$ and $\alpha < \frac{1}{\nu\lambda}$, we have

$$\begin{aligned}
&\mathbb{E}\big[\|\boldsymbol{p}_{f+1} - \boldsymbol{p}^*\|^2 \mid \boldsymbol{p}_0\big] \le \frac{\|\boldsymbol{p}_0 - \boldsymbol{p}^*\|^2}{\nu\lambda\alpha f} + \frac{\alpha m(\bar{a}+\bar{d})^2}{\nu\lambda}, \\
&\mathbb{E}\|\boldsymbol{p}_{f+1} - \boldsymbol{p}^*\| \le \frac{1}{f^{1/3}\sqrt{\nu\lambda f}} \cdot \frac{1}{\sqrt{\alpha}} + \frac{m(\bar{a}+\bar{d})}{\sqrt{\nu\lambda}} \cdot \sqrt{\alpha} + \frac{\bar{p}}{T^2}.
\end{aligned} \tag{45}$$

Then according to Lemma B.8, take $C_w = \max\{\frac{1}{\nu\lambda}, \frac{m(\bar{a}+\bar{d})}{\sqrt{\nu\lambda}}, \bar{p}\}$, we derive the constraint violation follows

$$\begin{aligned}
\mathbb{E}\Big\|\Big(\sum_{t=0}^{f}(\boldsymbol{a}_t x_t - \boldsymbol{d})\Big)^+\Big\| &\le \mathbb{E}\Big\|\sum_{t=0}^{f}(\boldsymbol{a}_t x_t - \boldsymbol{d})\Big\| \\
&\le \frac{1}{\alpha}\mathbb{E}\|\boldsymbol{p}_{f+1} - \boldsymbol{p}_0\| \\
&\le \frac{1}{\alpha}\mathbb{E}\Big[\|\boldsymbol{p}_0 - \boldsymbol{p}^*\| + \|\boldsymbol{p}_{f+1} - \boldsymbol{p}^*\|\Big] \\
&\le \frac{1}{\alpha}\Big(\frac{1}{f^{1/3}} + C_w\Big[\frac{1}{f^{1/3}\sqrt{f}} \cdot \frac{1}{\sqrt{\alpha}} + \sqrt{\alpha} + \frac{1}{T^2}\Big]\Big) \\
&\le \frac{1}{f^{1/3}} \cdot \frac{1}{\alpha} + \frac{C_w}{f^{5/6}} \cdot \frac{1}{\alpha\sqrt{\alpha}} + C_w \cdot \frac{1}{\sqrt{\alpha}} + \frac{C_w}{T^2} \cdot \frac{1}{\alpha}.
\end{aligned} \tag{46}$$

Similar to (41) in Theorem B.9, we derive the dual distance as

$$\begin{aligned}
\sum_{t=0}^{f}\mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 &\le \sum_{t=0}^{f}\Big[\frac{\|\boldsymbol{p}_0 - \boldsymbol{p}^*\|^2}{\nu\lambda\alpha t} + \frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}\alpha\Big] \\
&\le \frac{\|\boldsymbol{p}_0 - \boldsymbol{p}^*\|^2}{\nu\lambda\alpha}\log(f) + \frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} f\alpha \\
&\le \frac{\log(f)}{\nu\lambda f^{2/3}} \cdot \frac{1}{\alpha} + \frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} f\alpha.
\end{aligned} \tag{47}$$

Similar to (42) in Theorem B.9, we derive the positive projection for binding terms as

$$\begin{aligned}
\mathbb{E}\Big\|\Big(\sum_{t=0}^{f}(\boldsymbol{d} - \boldsymbol{a}_t x_t)\Big)^{B^+}\Big\| &\le \mathbb{E}\Big\|\sum_{t=0}^{f}(\boldsymbol{d} - \boldsymbol{a}_t x_t)\Big\| \\
&\le \frac{1}{\alpha}\mathbb{E}\|\boldsymbol{p}_0 - \boldsymbol{p}_{f+1}\| \\
&\le \frac{1}{\alpha}\mathbb{E}\Big[\|\boldsymbol{p}_0 - \boldsymbol{p}^*\| + \|\boldsymbol{p}_{f+1} - \boldsymbol{p}^*\|\Big] \\
&\le \frac{1}{f^{1/3}} \cdot \frac{1}{\alpha} + \frac{C_w}{f^{1/3}\sqrt{f}} \cdot \frac{1}{\alpha\sqrt{\alpha}} + C_w \cdot \frac{1}{\sqrt{\alpha}} + \frac{C_w}{T^2} \cdot \frac{1}{\alpha}.
\end{aligned} \tag{48}$$

Therefore, combining the results in (46), (47), and (48) together, we have

$$\begin{aligned}
\Delta_{T_1\text{warm}} &\le \frac{\log(f)}{\nu\lambda f^{2/3}} \cdot \frac{1}{\alpha} + \frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} f\alpha \\
&\quad + \frac{2}{f^{1/3}} \cdot \frac{1}{\alpha} + \frac{2C_w}{f^{1/3}\sqrt{f}} \cdot \frac{1}{\alpha\sqrt{\alpha}} + 2C_w \cdot \frac{1}{\sqrt{\alpha}} + \frac{2C_w}{T^2} \cdot \frac{1}{\alpha} \\
&\le \frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} f\alpha + \Big(\frac{\log(f)}{\nu\lambda f^{2/3}} + \frac{2}{f^{1/3}} + \frac{2C_w}{T^2}\Big) \cdot \frac{1}{\alpha} + \frac{2C_w}{f^{1/3}\sqrt{f}} \cdot \frac{1}{\alpha\sqrt{\alpha}} + 2C_w \cdot \frac{1}{\sqrt{\alpha}}.
\end{aligned} \tag{49}$$

Thus, taking the optimal learning rate $\alpha = f^{-2/3}$, we obtain the regret for the first batch with a warm start as:

$$\Delta_{T_1\text{warm}} \le \frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} f^{1/3} + \Big(\frac{\log(f)}{\nu\lambda f^{2/3}} + \frac{2}{f^{1/3}} + \frac{2C_w}{T^2}\Big) \cdot f^{2/3} + \frac{2C_w}{f^{1/3}\sqrt{f}} \cdot f + 2C_w \cdot f^{1/3}$$

$$\leq \frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} f^{1/3} + \frac{\log(f)}{\nu\lambda} + 2f^{1/3} + \frac{2C_w}{T} + 2C_w f^{1/6} + 2C_w \cdot f^{1/3}$$
$$\leq \Big(\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} + 4C_w\Big) \cdot f^{1/3} + \log(f). \tag{50}$$

This completes the proof.

C. Main Results

In this section, we demonstrate the proof for all theoretical results that we proposed in the main body of the paper. We instate Assumption 2.1 and 2.2 are satisfied.

C.1. Proof of Theorem 3.1.

By Theorem B.3, we obtain

$$\begin{aligned}
\Delta_T &= \mathbb{E}[r(\boldsymbol{x}) + v(\boldsymbol{x})] \\
&\leq \|\boldsymbol{p}^*\| \cdot \mathbb{E}\Big\|\Big(\boldsymbol{b} - \sum_{t=1}^{T} \boldsymbol{a}_t x_t\Big)^{B^+}\Big\| + \frac{\mu\bar{a}^2}{2}\sum_{t=1}^{T}\mathbb{E}\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2 + \mathbb{E}\Big\|\Big(\sum_{t=1}^{T}\boldsymbol{a}_t x_t - \boldsymbol{b}\Big)^{+}\Big\| \\
&\leq \|\boldsymbol{p}^*\| \cdot \mathbb{E}\left[\|(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})^{B+}\|\right] + \mu\bar{a}^2\sum_{t=1}^{T}\mathbb{E}\left[\|\boldsymbol{p}_t - \boldsymbol{p}^*\|^2\right] + \mathbb{E}\left[\|(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^{+}\|\right].
\end{aligned}$$

This completes the proof.

C.2. Proof of Theorem 3.2.

Combining the results of Theorem B.6, Theorem B.7, and Theorem B.9, for some constant $C_{reg} > 0$, we obtain:

$$\begin{aligned}
\Delta_T &= \Delta_{T_1} + \Delta_{T_2} + \Delta_{T_3} \\
&\leq \left(\frac{m(\bar{a}+\bar{d})^2}{2} + \frac{2\bar{r} + m(\bar{a}+\bar{d})^2}{\underline{d}} + m(\bar{a}+\bar{d})\right)\sqrt{f} \\
&\quad + \Big(C_{lp} + \frac{mC}{\nu^2\lambda^2} + (\bar{a}+\bar{d}+\delta)\frac{C}{\delta^2}\Big)\log(k) \\
&\quad + \Big(\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda} + \frac{2C_{lp}}{\sqrt{k}} + \frac{C_{lp}\log(f)}{\nu\lambda k} + \frac{4C_b}{\sqrt{k}} + 2C_b\Big)\cdot f^{1/3}.
\end{aligned} \tag{51}$$

Therefore, we achieve the worst-case regret of:

$$\Delta_T = \mathcal{O}(\log(k) + \sqrt{f} + f^{1/3}). \tag{52}$$

For special cases,

1. If we use LP-based method on the first batch, then we will have $\Delta_{T_1} \leq \log(f)$. Total performance follows

$$\Delta_T \leq \log(\max\{f,k\}) + f^{1/3}) \leq \log(\sqrt{T}) + f^{1/3} \in \mathcal{O}(\log(T) + f^{1/3}). \tag{53}$$

2. If we have a warm start for the first batch, with the initialization $\|\boldsymbol{p}_0 - \boldsymbol{p}^*\| \leq f^{-1/3}$, then the first batch regret achieves $\Delta_{T_1} \in \mathcal{O}(f^{1/3} + \log(f))$ by Lemma B.10. Thus, total performance follows

$$\Delta_T \leq \log(\max\{f,k\}) + 2f^{1/3}) \in \mathcal{O}(\log(T) + f^{1/3}). \tag{54}$$

This completes our proof.

D. Auxiliary Results

In this section, we provide auxiliary results to support the proof in the previous three sections. These lemmas focus on pure mathematical derivations.

D.1. Technical Support for LP-based Analysis

Lemma D.1. *Denote $d_{i,(j+1)f}, d_{i,jf}$ as the average consumption of i-th type resource at time $(j+1)f$ and jf. There exists a constant C_{rec} depending on $\bar{d}, \bar{a}, m, \nu, \lambda, \mu$, and C_{lp} such that:*

$$\mathbb{E}(d_{i,(j+1)f} - d_i)^2 \le \mathbb{E}(d_{i,jf} - d_i)^2 + \frac{C_{rec}}{(k-j-1)^2 f} + \frac{4\mu\bar{a}^2\sqrt{C_{lp}}}{(k-j-1)\sqrt{(j+1)f}}\sqrt{\mathbb{E}[(d_{i,jf} - d_i)^2]}.$$

Lemma D.2. *With the recursion relation in (26), there exists a constant $C > 0$ depending on $\bar{d}, \bar{a}, m, \nu, \lambda, \mu$, and C_{lp} such that the summation of the total deviation of $\boldsymbol{d}_t$ with the original $\boldsymbol{d}$ satisfies:*

$$\sum_{j=1}^{k} \mathbb{E}\left[(d_{i,jf} - d_i)^2\right] \le \frac{C}{f}\log(k).$$

Proof of Lemma D.1. We analyze each term in (25). The key technique we use is to take conditional expectations and simplify the double summations.

(a) Term 1.

$$\begin{aligned}
&\mathbb{E}\left[\frac{(\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf} - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}))^2}{(T-(j+1)f)^2}\right] \\
&= \frac{1}{(T-(j+1)f)^2}\,\mathbb{E}\left[\sum_{\ell=jf+1}^{(j+1)f}\sum_{s=jf+1}^{(j+1)f}\mathbb{E}\Big[(d_{i,jf} - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}))(d_{i,jf} - a_{i,s} I(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}^*_{jf})) \mid \boldsymbol{d}_{jf}\Big]\right] \\
&= \frac{1}{(T-(j+1)f)^2}\,\mathbb{E}\left[\sum_{\ell=jf+1}^{(j+1)f}\mathbb{E}\Big[(d_{i,jf} - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}))^2 \mid \boldsymbol{d}_{jf}\Big]\right] \\
&\quad + \frac{1}{(T-(j+1)f)^2}\,\mathbb{E}\left[\sum_{\ell\neq j,\ell=jf+1}^{(j+1)f}\mathbb{E}\Big[(d_{i,jf} - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}))(d_{i,jf} - a_{i,j} I(r_j > \boldsymbol{a}_j^\top \boldsymbol{p}^*_{jf})) \mid \boldsymbol{d}_{jf}\Big]\right] \\
&\le \frac{f(\bar{a}+\bar{d})^2}{(T-(j+1)f)^2} + \frac{\mathbb{E}\left[\sum_{\ell\neq j,\ell=jf+1}^{(j+1)f}\mathbb{E}\Big[d_{i,jf} - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf})|\boldsymbol{d}_{jf}\Big]\mathbb{E}\Big[d_{i,jf} - a_{i,j} I(r_j > \boldsymbol{a}_j^\top \boldsymbol{p}^*_{jf})|\boldsymbol{d}_{jf}\Big]\right]}{(T-(j+1)f)^2} \\
&= \frac{f(\bar{a}+\bar{d})^2}{(T-(j+1)f)^2} \quad (\text{since } \mathbb{E}\Big[d_{i,jf} - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf})|\boldsymbol{d}_{jf}\Big] = 0 \text{ for binding terms}) \\
&\le \frac{(\bar{a}+\bar{d})^2}{(k-j-1)^2 f}. \quad (\text{since } \lfloor T \rfloor = k \cdot f)
\end{aligned}$$

(b) Term 2.

$$\begin{aligned}
&\mathbb{E}\left[\frac{(\sum_{\ell=jf+1}^{(j+1)f} a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - a_{i,\ell} I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f}))^2}{(T-(j+1)f)^2}\right] \\
&= \frac{\mathbb{E}\left[\sum_{\ell,s=jf+1}^{jk+1)f}\mathbb{E}\Big[(a_{i,\ell}(I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f}))) \cdot (a_{i,s}(I(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}^*_{jf}) - I(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}_{(j+1)f})))\Big|\boldsymbol{d}_{jf}\Big]\right]}{(T-(j+1)f)^2}.
\end{aligned}$$

When $\ell = s$,

$$\mathbb{E}\Big[(a_{i,\ell}(I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})))^2 \mid \boldsymbol{d}_{jf}\Big] \le \bar{a}^2.$$

When $\ell \neq s$, by Assumption 2.2,

$$\begin{aligned}
&\mathbb{E}\Big[(a_{i,\ell}(I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})))(a_{i,s}(I(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}^*_{jf}) - I(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}_{(j+1)f}))) \mid \boldsymbol{d}_{jf}\Big] \\
&= \mathbb{E}\left[\mathbb{E}\Big[(a_{i,\ell}(I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})))(a_{i,s}(I(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}^*_{jf}) - I(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}_{(j+1)f}))) \mid \boldsymbol{a}_i, \boldsymbol{a}_s\Big] \mid \boldsymbol{d}_{jf}\right] \\
&= \mathbb{E}\left[a_{i,\ell}\mathbb{E}\Big[I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f}) \mid \boldsymbol{a}_\ell\Big] \cdot a_{i,s}\mathbb{E}\Big[I(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}^*_{jf}) - I(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}_{(j+1)f}) \mid \boldsymbol{a}_s\Big] \mid \boldsymbol{d}_{jf}\right] \\
&= \mathbb{E}\left[a_{i,\ell}(P(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf} \mid \boldsymbol{a}_\ell) - P(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f} \mid \boldsymbol{a}_\ell)) \cdot a_{i,s}(P(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}^*_{jf} \mid \boldsymbol{a}_s) - P(r_s > \boldsymbol{a}_s^\top \boldsymbol{p}_{(j+1)f} \mid \boldsymbol{a}_s)) \mid \boldsymbol{d}_{jf}\right] \\
&\le \mathbb{E}\left[\mu a_{i,\ell}\boldsymbol{a}_\ell^\top(\boldsymbol{p}_{(j+1)f} - \boldsymbol{p}^*_{jf}) \cdot \mu a_{i,s}\boldsymbol{a}_s^\top(\boldsymbol{p}_{(j+1)f} - \boldsymbol{p}^*_{jf}) \mid \boldsymbol{d}_{jf}\right] \quad \text{(using Assumption here)} \\
&\le \mu^2\bar{a}^4\,\mathbb{E}\big[(\boldsymbol{p}_{(j+1)f} - \boldsymbol{p}^*_{jf})^2 \mid \boldsymbol{d}_{jf}\big].
\end{aligned}$$

Combining these two cases together and by the convergence of LP-based method, we obtain the bound for Term 2 as

$$\begin{aligned}
&\mathbb{E}\left[\frac{(\sum_{\ell=jf+1}^{(j+1)f} a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f}))^2}{(T-(j+1)f)^2}\right] \\
&\le \frac{1}{(T-(j+1)f)^2}\,\mathbb{E}\left[\sum_{\ell=jf+1}^{(j+1)f}\bar{a}^2 + \sum_{\ell\neq s, \ell=jf+1}^{(j+1)f}\mu^2\bar{a}^4\,\mathbb{E}\big[(\boldsymbol{p}_{(j+1)f} - \boldsymbol{p}^*_{jf})^2\big|\boldsymbol{d}_{jf}\big]\Big|\boldsymbol{d}_{jf}\right] \\
&\le \frac{1}{(T-(j+1)f)^2}\left(f\bar{a}^2 + f^2\mu^2\bar{a}^4\mathbb{E}[(\boldsymbol{p}_{(j+1)f} - \boldsymbol{p}^*_{jf})^2\big|\boldsymbol{d}_{jf}]\right) \\
&\le \frac{1}{(T-(j+1)f)^2}\left(f\bar{a}^2 + f^2\mu^2\bar{a}^4\frac{C_{lp}}{jf}\right) \quad \text{(by Lemma A.4)} \\
&\le \frac{\bar{a}^2 + \frac{1}{j}\mu^2\bar{a}^4C_{lp}}{(k-j-1)^2 f}.
\end{aligned}$$

(c) Term 3.

$$\begin{aligned}
&2\mathbb{E}\left[(d_{i,jf} - d_i)\left(\frac{\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf} - a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf})}{T-(j+1)f}\right)\right] \\
&= \frac{2}{T-(j+1)f}\,\mathbb{E}\left[\sum_{\ell=jf+1}^{(j+1)f}\mathbb{E}\Big[d_{i,jf} - a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) \mid \boldsymbol{d}_{jf}\Big] \cdot (d_{i,jf} - d_i)\right] \\
&= 0. \quad \text{(by the definition of binding terms)}
\end{aligned}$$

(d) Term 4.

$$\begin{aligned}
&2\mathbb{E}\left[(d_{i,jf} - d_i)\left(\frac{\sum_{\ell=jf+1}^{(j+1)f} a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})}{T-(j+1)f}\right)\right] \\
&= \frac{2}{T-(j+1)f}\,\mathbb{E}\left[\sum_{\ell=jf+1}^{(j+1)f}(d_{i,jf} - d_i)a_{i,\ell}\,\mathbb{E}\Big[I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f}) \mid \boldsymbol{a}_\ell\Big]\right]
\end{aligned}$$

$$\leq \frac{2}{T-(j+1)f}\sum_{\ell=jf+1}^{(j+1)f}\mathbb{E}\left[(d_{i,jf}-d_i)a_{i,\ell}\left[P(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf} \mid \boldsymbol{a}_\ell) - P(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f} \mid \boldsymbol{a}_\ell)\right]\right]$$

$$\leq \frac{2}{T-(j+1)f}\sum_{\ell=jf+1}^{(j+1)f}\mathbb{E}\left[(d_{i,jf}-d_i)\mu a_{i,\ell}\boldsymbol{a}_\ell^\top(\boldsymbol{p}_{(j+1)f}-\boldsymbol{p}^*_{jf})\right] \quad \text{(by Assumption 2.2)}$$

$$\leq \frac{2\mu\bar{a}^2}{T-(j+1)f}\sum_{\ell=jf+1}^{(j+1)f}\sqrt{\mathbb{E}[(d_{i,jf}-d_i)^2]}\cdot\sqrt{\mathbb{E}[(\boldsymbol{p}_{(j+1)f}-\boldsymbol{p}^*_{jf})^2]} \quad \text{(by Cauchy's inequality)}$$

$$\leq \frac{2\mu\bar{a}^2}{T-(j+1)f}\cdot\frac{\sqrt{C_{lp}}f}{\sqrt{jf}}\sqrt{\mathbb{E}[(d_{i,jf}-d_i)^2]}$$

$$= \frac{4\mu\bar{a}^2\sqrt{C_{lp}}}{(k-j-1)\sqrt{(j+1)f}}\sqrt{\mathbb{E}[(d_{i,jf}-d_i)^2]}.$$

(e) Term 5.

$$2\mathbb{E}\left(\frac{\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf} - a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf})}{T-(j+1)f}\cdot\frac{\sum_{\ell=jf+1}^{(j+1)f} a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})}{T-(j+1)f}I(jf<\tau)\right)$$

$$\leq 2\sqrt{\mathbb{E}\left(\frac{\sum_{\ell=jf+1}^{(j+1)f} d_{i,jf} - a_\ell I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf})}{T-(j+1)f}\right)^2}\sqrt{\mathbb{E}\left(\frac{\sum_{\ell=jf+1}^{(j+1)f} a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}^*_{jf}) - a_{i,\ell}I(r_\ell > \boldsymbol{a}_\ell^\top \boldsymbol{p}_{(j+1)f})}{T-(j+1)f}\right)^2}$$

$$\leq 2\sqrt{\frac{(\bar{a}+\bar{d})^2}{(k-j-1)^2 f}}\cdot\sqrt{\frac{\bar{a}^2+\frac{1}{j}\mu^2\bar{a}^4C_{lp}}{(k-j-1)^2 f}} \quad \text{(by results of Term 1 and Term 2)}$$

$$= \frac{2\bar{a}(\bar{a}+\bar{d})\sqrt{1+\frac{1}{j}\mu^2\bar{a}^2C_{lp}}}{(k-j-1)^2 f}.$$

Combining all the terms, we obtain the upper bound as:

$$\mathbb{E}\left[(d_{i,(j+1)f}-d_i)^2\right] \leq \mathbb{E}(d_{i,jf}-d_i)^2 + \frac{(\bar{a}+\bar{d})^2}{(k-j-1)^2 f} + \frac{\bar{a}^2+\frac{1}{j}\mu^2\bar{a}^4C_{lp}}{(k-j-1)^2 f}$$
$$+ \frac{2\bar{a}(\bar{a}+\bar{d})\sqrt{1+\frac{1}{j}\mu^2\bar{a}^2C_{lp}}}{(k-j-1)^2 f} + \frac{4\mu\bar{a}^2\sqrt{C_{lp}}}{(k-j-1)\sqrt{(j+1)f}}\sqrt{\mathbb{E}[(d_{i,jf}-d_i)^2]}.$$

Taking $C_{rec} = (\bar{a}+\bar{d})^2 + \bar{a}^2 + \mu^2\bar{a}^4C_{lp} + 2\bar{a}(\bar{a}+\bar{d})\sqrt{1+\mu^2\bar{a}^2C_{lp}}$ completes the proof.

Proof of Lemma D.2. We consider a general sequence $\{z_j\}_{j=1}^k$ with

$$z_{j+1} \leq z_j + \frac{R}{(k-j-1)^2 f} + \frac{\sqrt{R}\sqrt{z_j}}{(k-j-1)\sqrt{(j+1)f}}$$

where $R>0$ is a constant.

Taking sum on both sides of the inequality and re-arranging, we have

$$\sum_{j=1}^k (k-j+1)(z_{j+1}-z_j) \leq \sum_{j=1}^k \frac{16R}{(k-j+1)f} + \sqrt{\frac{16R}{f}}\sum_{j=1}^k \frac{\sqrt{z_j}}{\sqrt{j+1}}$$
$$\leq \frac{16R}{f}\log(k) + \sqrt{\frac{16R}{f}}\sum_{j=1}^k \frac{\sqrt{z_j}}{\sqrt{j+1}}.$$

Noticing that $\sum_{j=1}^{k}(k-j+1)(z_{j+1}-z_j)=\sum_{j=1}^{k} z_j$, we have

$$\begin{aligned}\frac{16R}{f}\Big(\sum_{j=1}^{k}\frac{1}{j+1}\Big)\cdot\Big(\sum_{j=1}^{k} z_j\Big) &\geq \left(\sqrt{\frac{16R}{f}}\sum_{j=1}^{k}\frac{\sqrt{z_j}}{\sqrt{j+1}}\right)^2\\ &\geq \left(\sum_{j=1}^{k} z_j-\frac{4R}{f}\log(k)\right)^2.\end{aligned}$$

We treat $\sum_{j=1}^{k} z_j$ as the variable, then solve and get

$$\sum_{j=1}^{k} z_j \leq \frac{12R}{f}\log(k).$$

With this result, consider our recursion in (26), take the constant $C=12\max\{C_{rec},16\mu^2\bar{a}^4C_{lp}\}>0$, we obtain:

$$\sum_{j=1}^{k}\mathbb{E}\left[(d_{i,jf}-d_i)^2\right]\leq\frac{C}{f}\log(k).$$

D.2. Technical Support for First-order Analysis

Lemma D.3. *Following the updated rule of first-order method, we derive the last dual price satisfies:*

$$\mathbb{E}\|\boldsymbol{p}_{T+1}-\boldsymbol{p}^*\|\leq\frac{C_{lp}}{\sqrt{\nu\lambda(T-kf)kf}}\cdot\frac{1}{\sqrt{\alpha}}+\frac{m(\bar{a}+\bar{d})}{\sqrt{\nu\lambda}}\cdot\sqrt{\alpha}+\frac{\bar{p}}{T^2}.$$

Proof of Lemma D.3. According to Lemma A.6, take $\alpha_t=\alpha<\frac{1}{\nu\lambda}$, we derive the conditional expectation as

$$\begin{aligned}\mathbb{E}\big[\|\boldsymbol{p}_{T+1}-\boldsymbol{p}^*\|^2\mid\boldsymbol{p}_{kf}\big]&\leq(1-\nu\lambda\alpha)\mathbb{E}\big[\|\boldsymbol{p}_T-\boldsymbol{p}^*\|^2\mid\boldsymbol{p}_{kf}\big]+\alpha^2m(\bar{a}+\bar{d})^2\\ &\leq(1-\nu\lambda\alpha)^{T-kf}\|\boldsymbol{p}_{kf}-\boldsymbol{p}^*\|^2+\sum_{j=0}^{T-kf-1}\alpha^2m(\bar{a}+\bar{d})^2(1-\nu\lambda\alpha)^j\\ &\leq(1-\nu\lambda\alpha)^{T-kf}\|\boldsymbol{p}_{kf}-\boldsymbol{p}^*\|^2+\frac{\alpha^2m(\bar{a}+\bar{d})^2}{\nu\lambda\alpha}\\ &\leq\frac{1}{\nu\lambda\alpha(T-kf)}\|\boldsymbol{p}_{kf}-\boldsymbol{p}^*\|^2+\frac{\alpha m(\bar{a}+\bar{d})^2}{\nu\lambda}\end{aligned}\tag{55}$$

where we use the technique of $\sum_{j=0}^{T-kf-1}\alpha^2m(\bar{a}+\bar{d})^2(1-\nu\lambda\alpha)^j\leq\frac{1-(1-\nu\lambda\alpha)^{T-kf}}{\nu\lambda\alpha}\leq\frac{1}{\nu\lambda\alpha}$ and $(1-\nu\lambda\alpha)^{T-kf}\leq\frac{1}{1+\nu\lambda\alpha(T-kf)}\leq\frac{1}{\nu\lambda\alpha(T-kf)}$.

By LP-convergence result in Lemma A.4, we know $\mathbb{E}\|\boldsymbol{p}_{kf}-\boldsymbol{p}^*\|\leq\frac{C_{lp}}{\sqrt{kf}}$. By *Proposition 3.3* in (Gao et al., 2024), we know the event $E:=\|\boldsymbol{p}_{kf}-\boldsymbol{p}^*\|\leq\frac{C_{lp}}{\sqrt{kf}}$ with probability $\mathbb{P}\geq1-\frac{1}{T^4}$. By Lemma B.2, we know $\|\boldsymbol{p}_t\|\leq\bar{p}$.

Thus, we have

$$\begin{aligned}\mathbb{E}\|\boldsymbol{p}_{T+1}-\boldsymbol{p}^*\|^2&\leq\mathbb{E}\big[\|\boldsymbol{p}_{T+1}-\boldsymbol{p}^*\|^2|E\big]\cdot\mathbb{P}(E)+\mathbb{E}\big[\|\boldsymbol{p}_{T+1}-\boldsymbol{p}^*\|^2|\bar{E}\big]\cdot\mathbb{P}(\bar{E})\\ &\leq\frac{C_{lp}}{\nu\lambda(T-kf)kf}\cdot\frac{1}{\alpha}+\frac{m(\bar{a}+\bar{d})^2}{\nu\lambda}\cdot\alpha+\frac{\bar{p}}{T^4}.\end{aligned}\tag{56}$$

Thus, as $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ for any $a, b > 0$, by (56), we have

$$\mathbb{E}\|\boldsymbol{p}_{T+1} - \boldsymbol{p}^*\| \le \frac{C_{lp}}{\sqrt{\nu\lambda(T-kf)kf}} \cdot \frac{1}{\sqrt{\alpha}} + \frac{m(\bar{a}+\bar{d})}{\sqrt{\nu\lambda}} \cdot \sqrt{\alpha} + \frac{\bar{p}}{T^2}.$$

This completes the proof.

E. Supplementary Experiments

In this section, we provide more experiments to further evaluate the performance of our algorithms. We consider a more general and complex input distribution and include additional comparisons with recent methods.

E.1. New Distribution

As an extension of Section 4.1, our goal is to evaluate the main algorithms with different choices of re-solving frequency $f \in \{T^{1/3}, T^{1/2}, T^{2/3}\}$. We consider a more complex distribution for reward and resource consumption requests and guarantee that Assumptions 2.1 and 2.2 are still satisfied.

Consider

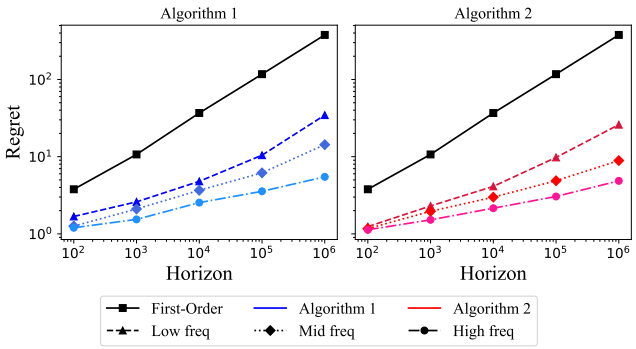
$$\text{Input III: } a_{it} \sim \min(1, \max(0, 1+z)),\ r_t \sim \text{Unif}[0,1]$$
$$\text{where } z \sim t(1) : \text{Student's t-distribution with 1 degree of freedom.}$$

We generate $\{r_t, \boldsymbol{a}_t\}_{t=1}^T$ from Input III and keep other parameters $T \in [10^2, 10^6]$ and $d_i \sim \text{Uniform}[1/3, 2/3]$ the same in Section 4.1. We report the average result over 100 trials for each experiment and use the classic first-order method with $\mathcal{O}(T^{1/2})$ regret (Algorithm 4) as a baseline.

Table 4: Algorithms under New Distribution.

	T	First-Order	Low freq	Mid freq	High freq
Algorithm 1	10^2	3.78	1.68	1.26	**1.20**
	10^3	10.69	2.60	2.10	**1.54**
	10^4	36.79	4.79	3.67	**2.54**
	10^5	117.13	10.45	6.17	**3.56**
	10^6	377.84	34.54	14.30	**5.48**
Algorithm 2	10^2	3.78	1.24	1.17	**1.13**
	10^3	10.69	2.30	1.96	**1.52**
	10^4	36.79	4.12	2.99	**2.15**
	10^5	117.13	9.78	4.88	**3.05**
	10^6	377.84	25.96	8.90	**4.86**

Figure 5: Regret for various re-solving frequencies.



Under the new distribution, our algorithms still exhibit a strong regret performance. As shown in Table 4 and Figure 5, we observe that regret decreases as the re-solving frequency increases. This trend holds consistently across both algorithms and is consistent with the guarantees of Theorem 3.2.

E.2. More Comparison

Building on the comparison of baseline methods in Section 4.2, we evaluate our algorithms against recent works. As a reminder, our algorithms employ frequent LP-solving to learn online dual price under continuous support. Li et al. (2024) considers a similar problem using infrequent LP-solving to update the dual variable but under finite support.

To compare them, we adapt our algorithms to finite support and take the re-solving frequency $f = T^{1/3}$ from Section 4.1; We take the best-performed parameters for the infrequent method where the number of customer types $n = 50$ and $\alpha = \beta = 0.95$ which control the solving times near start and end. We take the horizon over $T \in [10^2, 10^6]$ and report the average result over 100 trials for each experiment. We still use the first-order method in Algorithm 4 as a baseline.

Table 5: Algorithms comparison.

T	Regret	Algorithm	Compute Time (s)
10^3	11.92	First-Order	0.002
	11.12	Infrequent LP-based	0.724
	3.77	Algorithm 1	**0.790**
	3.69	Algorithm 2	**0.782**
10^4	36.37	First-Order	0.01
	14.20	Infrequent LP-based	0.827
	6.69	Algorithm 1	**3.015**
	6.35	Algorithm 2	**3.112**
10^5	110.22	First-Order	0.109
	20.90	Infrequent LP-based	1.028
	16.41	Algorithm 1	**52.95**
	10.84	Algorithm 2	**52.39**
10^6	312.26	First-Order	1.106
	28.80	Infrequent LP-based	4.929
	20.58	Algorithm 1	**1261.0**
	14.30	Algorithm 2	**1305.4**

Figure 6: Regret for various algorithms.

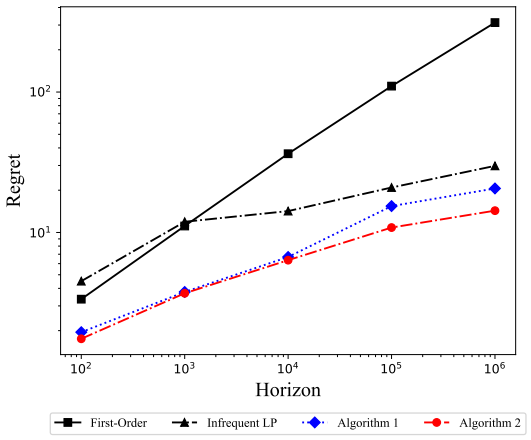


Table 5 and Figure 6 demonstrate the algorithm regret and computation time across different horizons. While the infrequent LP-based method has faster computation, our algorithms show a competitive performance in decision optimality and achieve lower regret under finite support.