Optimal Regularized Online Convex Allocation by Adaptive Re-Solving

Wanteng Ma¹, Ying Cao², Danny H.K. Tsang², Dong Xia¹∗

¹Department of Mathematics, HKUST
²Department of Electronic and Computer Engineering, HKUST

(September 2, 2022)

Abstract

This paper introduces a dual-based algorithm framework for solving the regularized online resource allocation problems, which have cumulative convex rewards, hard resource constraints, and a non-separable regularizer. Under a strategy of adaptively updating the resource constraints, the proposed framework only requests an approximate solution to the empirical dual problem up to a certain accuracy, and yet delivers an optimal logarithmic regret under a locally strongly convex assumption. Surprisingly, a delicate analysis of dual objective function enables us to eliminate the notorious loglog factor in regret bound. The flexible framework renders renowned and computationally fast algorithms immediately applicable, e.g., dual gradient descent and stochastic gradient descent. A worst-case square-root regret lower bound is established if the resource constraints are not adaptively updated during dual optimization, which underscores the critical role of adaptive dual variable update. Comprehensive numerical experiments and real data application demonstrate the merits of proposed algorithm framework.

1 Introduction

Online resource allocation seeks to maximize the total rewards in an online service system that is subject to resource constraints. As an exemplary model for sequential decision making, online allocation has drawn considerable attentions in recent decades. Meanwhile, it is strongly connected to other online problems such as revenue management (Talluri et al., 2004), online linear programming (Agrawal et al., 2014) and ads bidding problems (Lee et al., 2013), to name but a few. Online allocation finds applications in diverse fields, e.g., computer science and operation research. Oftentimes, online allocation problems feature resource constraints that are either hard (Mehta et al., 2007) or soft (Mahdavi et al., 2012), with different constraint capacities. The goal of a decision maker is to maximize the total rewards (revenue, utility) function by a real-time decision policy that enforces each of the resource constraints.

So far, existing literature on online allocation mostly focused on additively separable objectives, i.e., the objective function only involves the total rewards that can be simply described as the cumulative rewards by time (e.g., Mehta et al. (2007); Devanur and Hayes (2009); Balseiro and Gur (2019)). While a separable objective is favorable for tracking additive total rewards, it falls short of describing globally non-separable quantities such as total resource consumption or average actions. For instance,
the average action (Agrawal and Devanur, 2014) in online advertising measures the amount of under-delivery of impressions. Unfortunately, non-separable objectives are considerably under-explored in the literature, and particularly, there is a paucity of work investigating the impact of non-separable regularization on separable cumulative reward functions. Here we are interested in regularized online allocation problems, which add a non-separable regularizer to the objective function as a penalty for various purposes such as resource-saving, load balancing, diversity, and fairness (Ghosh et al., 2009; Balseiro et al., 2021). Compared with non-regularized online resource allocation that maximizes an additively separable objective, the non-separable regularization poses new challenges to algorithm design and regret analysis.

In this paper, we study regularized online allocation problem with a concave reward function and linear resource constraints under the so-called random input model (Goel and Mehta, 2008) where i.i.d. requests arrive sequentially and follow an unknown distribution. Decisions must be made sequentially, that is, once a request is received with a known reward function, the decision maker shall instantly make a decision based on current request, previous history and remaining resources. Throughout the paper, we impose hard constraints on the total resource consumption, which shall never be violated so that the decision maker must wisely control the resource consumption at any time. Clearly, the challenges of online allocation problems mainly stem from the dilemma of fulfilling the current request or reserving the resources for, possibly more rewardable, future ones. The task for a decision maker is to design a strategy that maximizes the regularized total rewards subject to resource constraints. The regularizer is a non-separable function of total resource consumption. A typical application of the problem under study is online advertising (Mehta et al., 2007; Agrawal et al., 2018) where a publisher needs to assign each impression to some advertiser and maximize the click-through rate with budget constraints on each advertiser. Oftentimes, other aspects of resource consumptions, including fairness of advertisers or load balancing, are put into consideration. Towards that end, a regularizer on total click-through rates can be added, in which case the objective function turns out to be the regularized cumulative total click-through rates.

Our main goal is to design computationally efficient algorithms for the aforementioned regularized online allocation problems, which, simultaneously, achieve theoretically optimal regrets. In the absence of non-separable regularizer, it has been well recognized that the lower bound of regret of online allocation problems grows at a logarithmic rate (Bray, 2019; Li and Ye, 2021). The forgoing works also proposed adaptive policies that achieve the logarithmic-order regrets up to an additional loglog factor. Moreover, Arlotto and Gurvich (2019) shows that adaptive policies are, generally, necessary to make a low regret possible. In sharp contrast, to our best knowledge, regrets achieved by prior algorithms (Balseiro et al., 2021) on regularized online allocation problems are of a square-root order. A first natural question is: can a logarithmic-order regret be achieved in the existence of a non-separable regularizer? Actually, we seek an even more ambitious goal: can we achieve a regret of exactly order $O(\log T)$ without the loglog factor so that the lower bound is sharply met? The next question is more crucial: is there any computationally efficient algorithm that attains the desired regret? Surprisingly, we give affirmative answers to both questions by designing an adaptive algorithm framework that is flexible, computationally fast, and theoretically guaranteed to achieve the sharply optimal regret. Extensive numerical simulations and real data experiments are presented to corroborate the effectiveness of our algorithms.

1.1 Contributions

To summarize, we make the following contributions in this paper.

**Sharp dual convergence in non-linear and regularized cases.** We derive the convergence rate of sample-version dual solution to its population counterpart in the case of additive non-linear rewards function and in the existence of a non-separable regularizer. The convergence rate is at $O(T^{-1})$, which
improves the known rate $O(T^{-1}\log \log T)$ that was established only for non-regularized linear reward functions \cite{Li and Ye, 2021}. The improvement is made possible by, jointly, a local strongly convex assumption on reward functions and a delicate analysis of the local behavior of sample-version stochastic dual program near the population optimal solution. The observed local behavior and derived convergence performance are also valid for linear or non-regularized cases. This dual convergence crucially motivates our approach to treat a non-separable objective, which converts the non-separable primal problem into a dual one that consists of separable functions. Our analysis establishes a connection between the approximation errors measured by function values and the deviations of approximate solutions, which are determined by both intrinsic randomness and approximation of solutions. It suggests that any approximate solution, up to a certain accuracy, to the dual optimization suffices to guarantee the overall convergence of a primal-dual algorithm, which lays the theoretical foundation for our history-dependent algorithm design. It is noteworthy that, as a stochastic optimization problem, the derived dual convergence sheds new light on the open Sample Average Approximation (SAA) problems and may be of independent interest.

**Adaptive algorithm framework.** We propose a flexible dual-based and history-dependent, i.e., reliant on past data and actions, algorithm framework for solving the regularized online allocation problem. As a primal-dual algorithm framework, each iteration mainly consists of two routines: primal decision making and dual optimization. At a high level, our adaptive algorithm framework generalizes the history-dependent policy in online linear programming \cite{Li and Ye, 2021}, which evolves from the budget-ration policy \cite{Arlotto and Gurvich, 2019, Balseiro and Gur, 2019} and the re-solving heuristic in network revenue management \cite{Jasin and Kumar, 2012, Wu et al., 2015}. There are two key ingredients in dual optimization of our algorithm framework. First, at each iteration, we adaptively update the average remaining resources in the dual problem. Besides fulfilling the resource constraints, this adaptive resource control plays a critical role in achieving a $O(\log T)$ regret rather than the $O(T^{1/2})$ one attained by \cite{Balseiro et al., 2021}. Secondly, at each iteration, our algorithm framework only requires an approximate solution, up to a certain accuracy, to the dual optimization. This allows a flexible choice of computationally efficient algorithms for dual optimization, be them deterministic or stochastic. Paired with first-order methods, our algorithm enjoys an acceptable polynomial-time cost comparable to prior algorithms. More specifically, for strongly convex objectives, it requires computing gradients for $O(t)$ times at time $t$; for more general convex objectives, it requires $O(t^3)$ times of gradient computation. Note that our algorithm framework is also applicable to linear reward functions or non-regularized online allocation problems.

**Regret analysis.** With its offline optimum as the benchmark, we investigate the regret attained by the adaptive algorithm framework for regularized online allocation problems. Since the regret is characterized by dual convergence, the aforementioned new result of dual convergence allows us to derive a sharp regret bound. More exactly, we show that our adaptive algorithm achieves an $O(\log T)$ regret, which matches the best results in constraint-free and non-regularized online convex optimization \cite{Hazan et al., 2007} and multi-secretary problem \cite{Bray, 2019}. A matching lower bound is established under our assumptions demonstrating the optimality of our adaptive algorithm framework. To our best knowledge, this is the first theoretical guarantee of an exact $O(\log T)$ regret bound for online non-linear allocation with hard constraints and a non-separable regularizer. The best known regret even for online learning programming \cite{Li and Ye, 2021} contains an additional $\log \log T$ factor. By comparing with existing algorithms, we clarify the critical role played by the adaptive resource control in controlling the stopping time and achieving a logarithmic-order regret. In particular, we establish a worst-case $O(T^{1/2})$ lower bound for dual-based algorithms if the resource constraints are not adaptively updated. Basically, without updating the resource constraints, dual-based algorithms suffer from early-stopping.

We then elaborate the applications of our method and theory to online linear programming, online convex optimization, online welfare maximization and online convex packing. Simulation results are also presented.
1.2 Related Work

1.2.1 Online Linear Allocation

Many online problems with resource constraints can be formulated into online allocation problems. A large proportion of early work mainly focused on linear models. Vazirani et al. (2005); Mehta et al. (2007); Buchbinder et al. (2007) studied the AdWords problem, where a search engine tries to assign some keywords to a set of competing bidders, each with a spending limit (i.e., constraint), and the goal is to maximize the revenue generated by these keyword sales. The rewards in AdWords problem are proportional to consumed resources and, thus, is a special case of online linear allocation. By viewing AdWords as a generalization of online bipartite matching problem, Mehta et al. (2007) achieved an optimal $(1 - e^{-1})$-competitive ratio, which is defined as the ratio of the revenue of an online algorithm to the revenue of the best offline algorithm. Under a so-called random permutation model, Devanur and Hayes (2009) proposed a two-phase dual training algorithm for AdWords problem and achieved the regret $O(T^{2/3})$. The random permutation model, which assumes that the arrivals are in random order and the order itself is uniformly distributed over all permutations, is more general than the random input model, which assumes i.i.d. arrivals. But random input model can be treated as a special case of random permutation model (Mehta, 2013). More discussions on the online allocation problems under random permutation model can be found in Babaioff et al. (2008); Goel and Mehta (2008); Molinaro and Ravi (2014) and references therein.

Apart from AdWords, two major topics related to online linear allocation are online revenue management problem and online multi-secretary problem. In online revenue management, a decision maker aims to find a dynamic pricing policy that maximizes a company’s linear total rewards when the number of supplied products is finite, demands of these products arrive sequentially, and the resources for manufacturing the products are limited. Online revenue management finds diverse applications in industry such as rental services, air travel, hospital services (Talluri et al., 2004), etc. The earliest regret analysis of this problem dates back to Cooper (2002), which proposed a static LP-based algorithm and achieved an $O(T^{1/2})$ regret. Later works show that better regret is achievable by a re-solving strategy, i.e., repeatedly solving an optimization program but with updated information. By combining the re-solving strategy and a trigger-and-threshold mechanism, Reiman and Wang (2008) reduced the regret significantly to $O(T^{1/4})$. Equipped with sufficiently frequent re-solving’s, Jasin (2015) proposed to re-estimate the parametric distribution of arrivals and proved that an $O(log^2 T)$ regret is attained. Jasin and Kumar (2012); Wu et al. (2015) and Bumpensanti and Wang (2020) investigated the special case when the i.i.d. arrivals obey a known finite-support discrete distribution, and established $O(1)$ regrets for re-solving style algorithms when the resource constraints are constants. Online multi-secretary problem (Kleinberg, 2005; Babaioff et al., 2007) is one of the simplest online allocation problems as it has only one integer constraint. Assuming the arrivals obey a known finite-support discrete distribution, Arlotto and Gurvich (2019) proposed an online budget-ratio (BR) policy where decisions to fulfill or ignore requests are made by comparing the remaining average budget with some fixed thresholds. Their BR policy is adaptive and achieved an $O(1)$ regret but is inapplicable to the case of multiple resource constraints. They also established a regret lower bound $\Omega(T^{1/2})$ for all non-adaptive policies. Conversely, if the arrival distribution is continuous, e.g. a simple uniform distribution over $[0, 1]$, Bray (2019) developed a regret lower bound $\Omega(log T)$ even when the distribution is known to a decision maker.

Other independent works of online linear programming also contribute greatly to the understanding of online allocation problems. Agrawal et al. (2014) proposed a history-dependent dual-based algorithm that dynamically update dual variables and periodically solve linear programs. Their algorithm achieved an $O(T^{1/2})$ regret under the random permutation model. When the arrivals satisfy the random input model, Devanur et al. (2019) proved that a dual-based algorithm that attained an $O(T^{1/2})$ regret. But their algorithm relies on the knowledge of the optimal allocation, which is unrealistic for
most applications. Otherwise, their algorithm requires periodically computing the optimal solution to an offline linear programming. More recently, Li and Ye (2021) introduced a history-dependent algorithm that adaptively updates the resource constraints, which achieved a regret $O(\log T \log \log T)$ that is almost optimal except the log log factor. But their strategy also requires exact solutions to an offline linear programs of growing sizes, which may be computationally intractable for large $T$. An $\Omega(\log T)$ regret lower bound was established, which is consistent with Bray (2019).

1.2.2 Online Convex Allocation

Linear objective functions only find limited applications in practice. Online convex allocation moves one step further by allowing convex objective functions. In Agrawal and Devanur (2014), the authors investigated online convex programming that is equipped with a fixed and convex reward function. The imposed stochastic constraints are soft meaning that a certain degree of constraint violations is allowed. They proposed a flexible algorithm framework based on online convex optimization, which, for general convex objectives, achieved an $O(T^{1/2})$ regret with $O(T^{1/2})$ constraint violations. Furthermore, if the objective function is smooth, their algorithm achieves an $O(\log T)$ regret with $O(\log T)$ constraint violations. The computational cost of their algorithm can be linear provided that the offline optimum is partially known. Otherwise, their algorithm requires solving convex programs for logarithmic times to estimate the benchmark information periodically, which can be computationally expensive.

Recently, partly due to its computational efficiency, dual mirror descent is extensively studied for online convex allocation problems. Balseiro et al. (2022, 2020) focused on a class of online allocation problems with separable reward functions and resource constraints that is proportional to time horizon $T$. They proposed a dual-based mirror descent algorithm that achieves $O(\sqrt{T})$ regret and was said to be unimprovable under their assumptions. Their algorithm updates dual variables by mirror descent and makes primal decisions by the conjugate functions. Their approach of controlling regret put less emphasis on stopping time but focused more on the complementary slackness of dual variables within updates. The rationale behind dual mirror descent is that it presents a self-correcting mechanism that naturally prevents resources from depleting too fast. This self-correcting mechanism relies on dual updates; that is, when a request consumes more resources, the corresponding dual variables will move against the excessive consumptions, and thus leading to a more conservative subsequent action. The problem we study in this paper is closer to Balseiro et al. (2021), which is the first to study online convex allocation problems with a non-separable regularizer and hard resource constraints. Their approach is similar to the non-regularized cases (Balseiro et al., 2022, 2020), except that they define a new separable dual problem and update dual variables using regularized subgradients since they allow non-smooth regularizers. They showed that, for regularized online convex allocation, dual mirror descent can still perform well and attain an $O(T^{1/2})$ regret. While this regret is optimal for general convex reward functions under both stochastic and adversarial input model, it is sub-optimal when the reward functions possess more favourable conditions like strong convexity. More recently, Lobos et al. (2021) extended dual mirror descent to an even more challenging online allocation problem where the separable objective and non-linear constraints are not necessarily convex. They proposed a novel benchmark to measure the regret and concluded that an $O(T^{1/2})$ regret is achievable by dual mirror (sub-gradient) descent.

Besides deterministic and hard constraints, a large body of literature on online convex programming focus on stochastic constraints and allow constraint violations. For instance, Yu et al. (2017) investigated online convex optimization with stochastic constraints and adversarial rewards. An $O(T^{1/2})$ bound is achieved for both the regret and constraint violations. A closely related problem is the long-term constraint problem, which aims to solve an online convex optimization problem by permitting a small number of cumulative constraint violations. Mahdavi et al. (2012); Jenatton et al. (2016) designed algorithms achieving $O(T^{1/2})$ regrets and $O(T^{3/4})$ constraint violations. When the objective function
is strongly convex, Yuan and Lamperski (2018) proposed an algorithm that achieves an $O(\log T)$ regret at the cost of $O((T \log T)^{1/2})$ constraint violations.

It is worth briefly mentioning the literature on general online convex optimization, which laid the early foundations of online convex allocation problems. For strongly convex objectives, classical literature on online convex optimization have revealed an optimal logarithmic regret. See Zinkevich (2003); Hazan et al. (2007) and references therein. It is reasonable to expect a logarithmic-order regret for other online problems in the existence of strong convexity. Nevertheless, achieving a logarithmic-order regret is challenging if an additional non-separable regularizer is posed. In literature, regularized online convex programming is commonly solved by the follow-the-regularized-leader style algorithms. For instance, Xiao (2010) introduced regularized dual averaging (RDA), which is an extension of the simple dual averaging algorithm originally proposed by Nesterov (2009), showing that an $O(T^{1/2})$ regret is achieved for general convex regularizer and $O(\log T)$ regret for strongly convex regularizer. However, their regularizer is separable, and hence their RDA scheme is inapplicable for our problem. A generalized follow-the-regularized-leader framework is summarized in McMahan (2011, 2017), which includes many online-mirror-descent style algorithms as special cases. Our dual-based adaptive algorithm differs from the follow-the-regularized-leader algorithms as it exploits more historical information rather than just the gradients and past actions, and it does not follow the leader. More introduction for general online convex optimization can be found in Hazan et al. (2016).

1.3 Notations

Some notations will be used throughout the paper. Define $a \land b := \min\{a, b\}$ and $a \lor b := \max\{a, b\}$. Write $[n]$ as the shorthand of $\{1, \ldots, n\}$. Define the non-negative region $\mathbb{R}_+ := \{x | x \geq 0\}$. We will always use $i$ to denote dimensions and use $d_i$ for the $i$-th dimension of vector $d$, and for vector sequence $\{d_i\}_{i=1}^T$, i.e., $d_{it}$ stands for the $i$-th entry of vector $d_t$. Denote $(x)^+ := \max\{x, 0\}$, $\| \cdot \|_2$ and $\| \cdot \|_\infty$ for the vector $\ell_2$-norm and $\ell_\infty$-norm, respectively.

2 Regularized Online Allocation Problem

We describe the convex regularized online allocation problem with finite time period $T$ as following:

$$\max_{\{x_t, t \in [T]\}} \sum_{t=1}^T f_t(x_t) + T \cdot r(\frac{\sum_{t=1}^T b_t x_t}{T})$$

$$\text{s.t.} \quad \sum_{t=1}^T b_t x_t \leq dT, \quad d \in \mathbb{R}_+^m$$

$$x_t \in \mathcal{X}, \forall t \in [T].$$

(2.1)

where $f_t : \mathbb{R}^n \to \mathbb{R}$ is the concave reward function, $r : \mathbb{R}^m \to \mathbb{R}$ is a concave regularizer to penalize the average resource consumption, $b_t \in \mathbb{R}^{m \times n}$ is the cost matrix and its entry could be both positive or negative (i.e., we can replenish the resource). We assume our inputs are stochastic, meaning that the i.i.d. requests $\{(f_t, b_t)\}_{t=1}^T$ are sampled from an unknown distribution $\mathcal{P}$: $(f_t, b_t) \sim \mathcal{P}$. The decision region $\mathcal{X} \subseteq \mathbb{R}_+^n$ is closed and convex with void action $0 \in \mathcal{X}$.

Following the online sequential learning setting, we assume that at each time $1 \leq t \leq T$, we first receive a request with known reward function and cost $(f_t, b_t)$ and then make the decision $x_t$ based on the observation of $t$-th request and history $\mathcal{H}_{t-1} := \{f_j, b_j, x_j\}_{j=1}^{t-1}$.

$$x_t := A(f_t, b_t, \mathcal{H}_{t-1}),$$

where $A$ is a decision rule.
by taking the total resource constraints \( \sum_{j=1}^{t} b_j x_j \leq dT \) into consideration. Here \( A \) denotes a history-dependent algorithm. Our goal is to design such an online algorithm \( A \) that can maximize the regularized total reward \( \sum_{t=1}^{T} f_t(x_t) + T \cdot r(T^{-1} \cdot \sum_{t=1}^{T} b_t x_t) \). Define the algorithm expected reward over a given distribution \( P \) as

\[
R(A|P) := \mathbb{E}_{A,P} \left[ \sum_{t=1}^{T} f_t(x_t) + T \cdot r(T^{-1} \cdot \sum_{t=1}^{T} b_t x_t) \right].
\] (2.2)

Here we take expectation with respect to both the inputs and the algorithm \( A \) if \( A \) is a stochastic algorithm. To measure the performance of an online algorithm, we compare the algorithm reward with the expected offline optimum (or hindsight optimum) defined by

\[
R^*(P) := \mathbb{E}_P \left[ \max_{x_t \in X} \sum_{t=1}^{T} f_t(x_t) + T \cdot r(T^{-1} \cdot \sum_{t=1}^{T} b_t x_t) \right],
\] (2.3)

which serves as the benchmark performance. For a given \( P \), define the regret as

\[
\text{Regret}(A) := \sup_{P \in \Xi} \{ R^*(P) - R(A|P) \}.
\] (2.4)

Compared with unconstrained online optimization, the key obstacle to designing algorithms for the online allocation problem is to enforce the total resource constraints, which shall not be violated at any time. However, we can transform the primal problem into a dual one with fewer constraints by the duality theory. This motivates us to investigate the problem (2.1) from the dual perspective.

### 2.1 The dual problem

We consider the dual problem of online allocation (2.1). The Lagrangian of this problem is

\[
L(x, a, \lambda, \mu) := \sum_{t=1}^{T} f_t(x_t) + T \cdot r(a) + \mu^\top (aT - \sum_{t=1}^{T} b_t x_t) + \lambda^\top (dT - \sum_{t=1}^{T} b_t x_t).
\] (2.5)

Here we introduce the equality constraint \( a = (\sum_{t=1}^{T} b_t x_t)/T \) in order to separate \( r(T^{-1} \cdot \sum_{i=1}^{T} b_t x_t) \) into additive terms. Denote the domain of \( r(a) \) as \( Z \) with \( b \circ X := \text{span} \{ b \cdot x \mid \text{for all possible } b \text{ and } x \in X \} \subseteq Z \). Define the conjugate function

\[
f_t^*(\lambda) := \max_{x \in X} \left\{ f_t(x) - x^\top \lambda \right\},
\]

\[
r^*(\mu) := \max_{a \in Z} \left\{ r(a) - a^\top \mu \right\}.
\] (2.6)

Then, the dual problem of (2.1) can be written as

\[
\min_{\mu, \lambda} \quad D_T(\lambda, \mu, d) := \frac{1}{T} \sum_{t=1}^{T} f_t^*(b_t^\top (\mu + \lambda)) + r^*(-\mu) + d^\top \lambda
\]

\[
s.t. \quad \lambda \succeq 0.
\] (2.7)

Under our stochastic input assumption, (2.7) can be viewed as a sample average approximation (SAA) \( \text{Shapiro et al. 2009} \) of the following stochastic program:
\[
\min_{\mu, \lambda} \quad D(\lambda, \mu, d) := \mathbb{E} f_t^* (b_t^\top (\mu + \lambda)) + r^* (-\mu) + d^\top \lambda \\
\text{s.t.} \quad \lambda \succeq 0.
\] (2.8)

In the following discussion, we will sometimes write the dual variable uniformly as \( \lambda := [\lambda^\top, \mu^\top]^\top \) in shorthand. If we have known the exact offline solution to (2.7), denoted by \( \lambda^*_T \), then by choosing the corresponding primal variables we can optimize the primal problem (2.7). However, in online setting it is impossible to find such exact dual solution before time \( T \). Thus at time \( t \) we turn to solve the \( t \)-sample average approximation of \( D(\lambda, \mu, d) \), i.e.,

\[
\min_{\mu, \lambda} \quad \bar{D}_t(\lambda, \mu, d) := \frac{1}{t} \sum_{j=1}^{t} f_j^* (b_j^\top (\mu + \lambda)) + r^* (-\mu) + d^\top \lambda \\
\text{s.t.} \quad \lambda \succeq 0.
\] (2.9)

and then use the dual approximate solution \( \lambda_t \) to decide the primal solution \( x_t \). Such a re-solving idea can also be found in other contexts (Jasin, 2015; Ferreira et al., 2018; Li and Ye, 2021) and has shown its merit in controlling the regret both in theory and in practice. Hence we expect that this idea also works in convex online allocation problems. Nevertheless, to discuss how practical this re-solving idea is in our setting, we still have three crucial questions to answer:

1. What is the behavior of \( \lambda^*_T \) when \( T \) is large? We know that \( \lambda^*_T \) varies depending on the data we collected. But from the stochastic programming perspective, as \( T \) goes large, the optimal solution to the SAA (2.7), \( \lambda^*_T \), will converge to the solution to its stochastic program (2.8), denoted by \( \lambda^* \). If we want to establish the theory of dual-based algorithms that rely on the approximation of \( \lambda^*_T \), we need to explore the convergence behavior of \( \lambda^*_T \) toward \( \lambda^* \) before we proceed with the study of algorithms.

2. How will the dual approximate solutions affect our reward and, consequently, the regret? This question is the key for the algorithm design. For online allocation problems, a good approximation of \( \lambda^* \) or \( \lambda^*_T \) does not necessarily mean a good reward because of the restriction imposed by resource depletion and stopping time. As we will show later, simply solving the convex programming (2.9) is not enough to achieve the optimal regret. We attempt to explain the influence of dual approximate solutions on regret in two phases: before and after stopping time, and show that the adaptive strategy of updating constraints is necessary for optimal regret.

3. How to control the regret as well as make the algorithm computationally efficient? Most of the re-solving techniques require periodically solving potentially large-scale convex programming, which is computationally demanding. Interestingly, we will show that a proper approximation of dual optimal solutions up to certain precisions can significantly reduce the computational costs, while maintaining the optimal order of regret. The influence of our approximation scheme on the regret is, in general, negligible when compared to the exact optimal solutions.

We propose an online adaptive algorithm for solving program (2.1), which achieves logarithmic regret based on the following assumptions.

2.2 Assumptions

**Assumption 1** (Basic assumptions on arrivals). The arrival sequences \( \{(f_t, b_t)\} \) satisfy:

1.1 \( \{(f_t, b_t)\}_{t=1}^{T} \) are generated i.i.d. from distribution \( \mathcal{P} \).
1.2 $f_t$ is strictly concave in the closed convex decision region $X \subseteq \mathbb{R}^n_+$ with $\|x\|_\infty \leq D$ for any $x \in X$.

1.3 There exists $\bar{f} \in R_+$ such that $\forall x \in X$, $|f_t(x)| \leq \bar{f}$.

1.4 There exists $\bar{b} \in R_+$ such that $\|b_t\|_2 \leq \bar{b}$ for any $t$.

1.5 We assume there exists $\bar{d} > 0$, and a large $\bar{d} > 0$ such that for any $i \in [m]$, $d_i \in (\bar{d}, \bar{d})$. Denote $\Omega_d = \bigotimes_{i=1}^n (d_i, \bar{d})$.

The assumptions on the upper bound $\bar{f}$ and $\bar{b}$ are common and practical in online allocation problems. It helps us control the size of the problem and ease our analysis. Assumption 1.5 follows from [Li and Ye (2021)]. We assume that the average resource constraints $\bar{d}$ is of a reasonable size, i.e., $d_i$ is neither too large nor too small. If $d_i$ is too large, then the constraint itself will be of no interest because the restriction it imposed on the primal variables is negligible. This assumption is crucial for the subsequent discussion of regret, especially for bounding the stopping time.

Under Assumption 1, we can define the general feasible region of our regularizer $r(a)$ as $Z := \{a \mid \|a\|_2 \leq \sqrt{n}Db\}$, which satisfies $b \circ X \subseteq Z$. We then describe the necessary assumptions on the regularizer $r$.

In order to study the influence of the average constraint $d$ on adaptive algorithms and how the variation of it affects the solution, we need the following assumptions.

**Assumption 2** (Assumptions on the regularizer). Suppose $(\lambda^*, \mu^*)$ is the optimal solution to the problem (2.8) when $d \in \Omega_d$. Then for any $d \in \Omega_d$, the concave regularizer $r$ is either 0 or satisfies:

1. $r$ is strictly concave and bounded in $Z$: $|r| \leq \bar{r}$ with bounded (sub)gradient $\|\nabla r(a)\|_\infty \leq G$ for any $a \in Z$.

2. The conjugate $r^*$ satisfies: $\langle \nabla r^*(-\mu) - \nabla r^*(-\mu^*), \mu - \mu^* \rangle \geq \tilde{L}_r \|\mu - \mu^*\|_2^2$ for any $\mu$ satisfying $\|\mu\|_\infty \leq G$ and some constant $\tilde{L}_r > 0$.

3. The conjugate $r^*$ satisfies: $\|\nabla r^*(-\mu) - \nabla r^*(-\mu^*)\|_2 \leq \tilde{L}_r \|\mu - \mu^*\|_2$ for any $\mu$ satisfying $\|\mu\|_\infty \leq G$ and some constant $\tilde{L}_r > 0$.

Together with Assumption 1, we can show that both the population-version and sample-version optimal solutions, $\lambda^*$ and $\lambda^*_T$, respectively, are uniformly bounded.

**Lemma 1.** Under Assumption 1 and the optimal solutions to problem (2.7) and (2.8) are bounded by:

$$\|\lambda^*_T\|_\infty \leq \frac{2(\bar{f} + \bar{r})}{\bar{d}}, \|\lambda^*\|_\infty \leq \frac{2(\bar{f} + \bar{r})}{\bar{d}}, \|\mu^*_T\|_\infty \leq G, \|\mu^*\|_\infty \leq G$$

(2.10)

By Lemma 1, we define the regions that contain all the possible optimal dual variable as $\Omega_\lambda := \{\lambda \mid \lambda \geq 0 \text{ and } \|\lambda\|_\infty \leq \frac{2(\bar{f} + \bar{r})}{\bar{d}}\}$, and $\Omega_\mu := \{\mu \mid \|\mu\|_\infty \leq G\}$. These regions will be the feasible sets of our dual variables since we do not want them to move far from the optimal solution $\lambda^*$. Assumption 2.2 and 2.3 require the conjugate of regularizer to be smooth and have quadratic growth. This can be achieved if the regularizer $r$ is locally strongly convex and smooth (see, Kakade et al. (2009) or Agrawal and Devanur (2014) for the conjugate of strongly convex/smooth functions). But our assumption is a bit weaker than directly assuming strong convexity and smoothness on $r$ itself.

Here are several possible regularizers that satisfy our assumptions:

1. $\ell_2$-loss: $r(a) := -\kappa \|a\|_2^2$. This regularizer serves as a tool to directly penalize resource consumption and achieve the goal of resource saving.
2. **Smooth minima**: \( r(a) := -\kappa \log(\sum_{i=1}^{m} \exp(-a_i) + \exp(-\sqrt{n}D\bar{b})) \). This LogSumExp regularizer is the smooth approximation of max-min fairness regularizer \( r(a) := \kappa \min_i \{a_i\} \land \sqrt{n}D\bar{b} \), which forces us to maximize the minimum resource consumption. Resources after max-min fairness regularization tend to be distributed fairly so that all resources are utilized adequately. See, e.g., Nash (1950); Bertsimas et al. (2011); Balseiro et al. (2021).

3. **Smooth maxima**: \( r(a) := -\kappa \log(\sum_{i=1}^{m} \exp(a_i) + 1) \). This regularizer is the smooth approximation of negative maximum consumption \( r(a) := -\kappa \max_i \{a_i\} \lor 0 \). This represents the load-balancing task: we minimize the maximum resource consumption so that all the resources are evenly distributed and no resource is over-exploited (or balanced load for every computer server in the load-balancing task).

4. **Entropy loss**: \( r(a) := -\kappa \sum_{i=1}^{m} a_i \log(a_i) + (1 - \sum_{i=1}^{m} a_i) \log(1 - \sum_{i=1}^{m} a_i) \) with the corresponding feasible region: \( \mathcal{Z} := \{a \in \mathbb{R}_+^m \mid \sum_{i=1}^{m} a_i \leq 1\} \). We use this entropy loss when our problem is related to random strategies and probabilistic assignment, e.g., in the online advertising, we randomly assign each impression to different advertisers with selected probabilities. This entropy loss regularizer seeks to find online allocation strategies with high entropy, which may share appealing properties like diversity, fairness, or robustness (Agrawal et al., 2018).

5. **Huber loss** [Huber, 1964]: \( r(a) := -\kappa \sum_{i=1}^{m} \frac{1}{2} a_i^2 \mathbb{I}(\vert a_i \vert \leq \delta) + \delta(|a_i| - \frac{\delta}{2}) \mathbb{I}(|a_i| > \delta) \) for some \( \delta > 0 \). Then conjugate of a Huber loss is also in the form \( r^*(\mu) := \sum_{i=1}^{m} \frac{1}{2} a_i^2 \mathbb{I}(\vert a_i \vert \leq \kappa \delta) + (\sqrt{n}D\bar{b}\vert \mu_i \vert - \kappa \delta(\sqrt{n}D\bar{b} - \frac{\delta}{2}) \mathbb{I}(\vert a_i \vert > \kappa \delta) \). Huber loss satisfies our assumption if the optimal solution sits in the center of \( \Omega_\mu := \{\mu \mid \parallel \mu \parallel_\infty < \kappa \delta \} \). This depends on actual problems since \( \mu^* \) is determined by both \( f \) and \( r \). But Huber loss entails that our regularizer may not necessarily be (globally) strongly convex and smooth. Compared with the \( \ell_2 \)-loss, Huber loss penalizes more mildly to extreme resource consumptions.

6. **No regularizer**: \( r(a) := 0 \). In this case, our problem is reduced to the non-regularized online convex allocation problem. Therefore, the theory developed in this paper is immediately applicable to the non-regularized cases.

In addition, we need the following non-degeneracy assumptions.

**Assumption 3** (Non-degeneracy). We assume that our problem is non-degenerate: suppose \( (\lambda^*, \mu^*) \) is the optimal solution to the problem \( 2.8 \) when \( d \in \Omega_d \). For ease of notations, we write \( \lambda^* \) and \( \mu^* \) instead of \( \lambda^*(d) \) and \( \mu^*(d) \), respectively. Then for any \( d \in \Omega_d \),

3.1 Let \( \nu := \lambda + \mu \) and \( \nu^* := \lambda^* + \mu^* \). The conjugate function \( f_t^* \) satisfies

\[
\| \mathbb{E} \left[ \nabla f_t^* (b_t^\top \nu) - \nabla f_t^* (b_t^\top \nu^*) \mid b_t \right] \|_2 \leq \mathcal{L}_f \| b_t^\top \nu - b_t^\top \nu^* \|_2,
\]

\[
\mathbb{E} \left[ \langle \nabla f_t^* (b_t^\top \nu) - \nabla f_t^* (b_t^\top \nu^*), b_t^\top \nu - b_t^\top \nu^* \rangle \mid b_t \right] \geq \mathcal{L}_f^t \| b_t^\top \nu - b_t^\top \nu^* \|_2^2
\]

for any \( \lambda \in \Omega_\lambda, \mu \in \Omega_\mu \) and constants \( \mathcal{L}_f, \mathcal{L}_f^t > 0 \), conditioning on \( b_t \).

3.2 The matrix \( M := \mathbb{E} [b_t b_t^\top] \) is positive definite with minimum eigenvalue \( \sigma_{\text{min}} > 0 \).

3.3 Define the primal variable given \( (\lambda^*, \mu^*) \) as \( \tilde{x}_t(\lambda^*) := \arg \max_{x \in X} \{ f_t(x) - (\lambda^* + \mu^*)^\top b_t x \} = -\nabla f_t^* (b_t^\top (\lambda^* + \mu^*)) \). Then the optimal solution \( (\lambda^*, \mu^*) \) satisfies \( \lambda^*_i = 0 \) if and only if \( d_i - \mathbb{E} (b_t \tilde{x}_t(\lambda^*))_i > 0 \).
Assumption 3.1 requires the expected conjugate of reward function to exhibit a local quadratic growth and smoothness, conditioning on any given $b_t$. Combined with Assumption 2.2 and 2.3, Assumption 3.1 ensures that the stochastic program (2.8) is locally smooth and locally strongly convex. Assumption 3.1 controls the growth rate of the reward function (and its conjugate) so that it will neither grow too fast nor degenerate to a line, which plays a critical role in characterizing dual solutions. Assumption 2.2 is easily satisfied since, oftentimes, the constraints are linearly independent. Assumption 4 is not necessary for the theoretical study of adaptive algorithms with frequently updated constraints. Assumption 3 states the non-degeneracy condition of linear programs (Jasin and Kumar, 2012; Jasin, 2015; Wu et al., 2015; Li and Ye, 2021). Note that Assumption 3 only concerns the deterministic problem (2.8), but the empirical problem not necessarily share these local properties.

In this sequel, all the dimensions that satisfy $d_i - \mathbb{E}(b_i x_t(\lambda^*))_i = 0$ with respect to the original $d$ in (2.1) are referred to as binding dimensions. Denote $I_B =\{i | d_i - \mathbb{E}(b_i x_t(\lambda^*))_i = 0\}$ the collection of binding dimensions. Similarly, non-binding dimensions are written as $I_{NB} =\{i | d_i - \mathbb{E}(b_i x_t(\lambda^*))_i > 0\}$. Here for ease of notations, we omit the dependence of $I_B$ and $I_{NB}$ on the resource constraint $d$. Assumption 3.3 ensures that binding and non-binding dimensions can be uniquely determined by the dual solution $\lambda^*$.

Note that $-\nabla f_t^*(b_t^\top \nu)$ represents the primal solution given dual variable $\nu$. Its randomness stems from the stochastic reward function $f_t$. From this perspective, Assumption 3 concerns the affect of dual variables to their corresponding expected primal solutions. It turns out that merely the perturbation behavior of expected primal solutions is not sufficient for our analysis, and we also need the perturbation behavior of the intrinsically random primal solutions, which can be controlled by the second moment. The following assumption serves for this purpose. Equivalently, it depicts the variation behavior of the random award function $f_t$. This second-order moment establishes the connection between dual variables and primal performances.

**Assumption 4 (Smoothness of the second moment).** Let $\nu := \lambda + \mu$ and $\nu^* := \lambda^* + \mu^*$ when we choose $d \in \Omega_d$ in (2.8). The second moment of the gradient $\nabla f_t^*$ satisfies the following smoothness

$$\mathbb{E}\left[\left\| \nabla f_t^*(b_t^\top \nu) - \nabla f_t^*(b_t^\top \nu^*) \right\|^2 \right] \leq L_2 \left\| b_t^\top (\nu - \nu^*) \right\|^2$$

for any $d \in \Omega_d$, $\lambda \in \Omega_\lambda$, $\mu \in \Omega_\mu$ and given $b_t$, where $L_2 > 0$ is a constant.

Assumption 4 requires the variation of reward function given $b_t$: $f_t \sim P|b_t$ to be mild so that the primal solution $x_t(\lambda) := -\nabla f_t^*(b_t^\top (\lambda + \mu))$ has a second order moment smoothness. Note that this doesn’t mean that $\nabla f_t^*$ must be globally smooth. A similar description of smoothness can be found in Gorbunov et al. (2020). Compared with Assumption 3.1, Assumption 4 actually states the smoothness in a different perspective. Assumption 3.1 only requires the smoothness of the expected reward, but here Assumption 4 focuses more on the variation of the random reward function itself. Basically, Assumption 4 claims that no matter how the reward $f_t$ varies, the difference of primal variables can be bounded by the difference of dual variables in expectation. Assumption 4 is not necessary for the study of dual convergence in section 3, but it is indispensable for the theoretical study of adaptive algorithms and regret analysis. We note that Assumptions 2 and 4 assume the corresponding conditions holds for all the $d \in \Omega_d$. 
3 Dual Convergence

For all dual-based online algorithms, the finite-sample convergence rate of dual variables is of great value since it reveals the best performance dual-based algorithms can achieve compared to the deterministic optimum. Recall the optimal solution \( \lambda_T^* \) to the sample average approximation (SSA) in eq. (2.7). The Law of Large Numbers dictates that \( \lambda_T^* \) converges to \( \lambda^* \) in probability as \( T \to \infty \). While the asymptotic behaviors of optimal solutions to SSA have been intensively studied in the literature (Kleywegt et al., 2002; Shapiro et al., 2009; Kim et al., 2015), they are not enough for us to develop the non-asymptotical dual convergence in the case of regularized online convex programming. In this section, we establish the dual convergence bounds under locally strong convexity, i.e., Assumptions 1-3, for regularized online problem (2.1). We emphasize that our assumptions hold uniformly for all \( d' \in \Omega_d \). Consequently, the dual convergence performance we will derive in this section also holds for all \( d' \in \Omega_d \).

Define \( D_t(\lambda, d) := f_t^*(b_t^T (\mu + \lambda)) + r^*(-\mu) + d^T \lambda \), and the corresponding gradient

\[
\phi_t(\lambda, d) := \nabla_\lambda D_t(\lambda, d) = \begin{bmatrix} b_t \nabla f_t^*(b_t^T (\mu + \lambda)) + d \\ b_t \nabla f_t^*(b_t^T (\mu + \lambda) - \nabla r^*(-\mu) \end{bmatrix}
\]

Then we have \( \nabla D(\lambda, d) := \nabla_\lambda D(\lambda, d) = \mathbb{E}\phi_t(\lambda, d) \). Denote \( \tilde{\phi}_T(\lambda, d) := T^{-1} \sum_{t=1}^T \phi_t(\lambda, d) \). One crucial idea to bounding the dual convergence is that the confined growth rate of \( D(\lambda, d) \) indicates that, with high probability, its sample version \( \tilde{D}_T(\lambda, d) \) is lower bounded by a quadratic function (Li and Ye, 2021). The confined growth speed of \( D(\lambda, d) \) is guaranteed by the following proposition.

**Proposition 1.** Under Assumptions 1-3, the objective function \( D(\lambda, d) \) in stochastic program (2.8) satisfies the following growth condition:

\[
\mathcal{L}_D \|\lambda - \lambda^*\|^2 \leq D(\lambda, d) - D(\lambda^*, d) - \nabla D(\lambda^*, d)^\top (\lambda - \lambda^*) \leq \tilde{\mathcal{L}}_D \|\lambda - \lambda^*\|^2,
\]

where the constant \( \mathcal{L}_D := \frac{\mathcal{L}_c}{4} \wedge \frac{1}{2} \frac{\mathcal{L}_c}{\mathcal{L}_c + 2\mathcal{L}_f \sigma_{\min}}; \), \( \tilde{\mathcal{L}}_D := \tilde{b}^2 \tilde{\mathcal{L}}_f + \tilde{\mathcal{L}}_r / 2 \).

By Proposition 1, we now derive an upper bound for dual convergence \( \|\lambda_T^* - \lambda^*\|_2 \) by capturing the shape of dual objective \( \tilde{D}_T(\lambda, d) \). While this idea is typical in literature (Li and Ye, 2021), we seek a more delicate analysis, which enables us to reach a sharper result. Basically, we focus on the local behavior of \( \tilde{D}_T(\lambda, d) \) around \( \lambda^* \). The rationale is obvious. Since \( \tilde{D}_T(\lambda, d) \) is always convex and converges to a deterministic convex function, the shape of \( \tilde{D}_T(\lambda, d) \) in a small neighborhood of \( \lambda^* \) will mimic that of \( D(\lambda, d) \) as long as \( T \) is large enough. Consequently, its optimal solution \( \lambda_T^* \) will lie in a small neighborhood of \( \lambda^* \).

Consider the first order and second order term of \( \tilde{D}_T(\lambda, d) \) separately. Decompose the convex function \( \tilde{D}_T(\lambda, d) = \tilde{D}_T(\lambda^*, d) \) into two parts:

\[
\tilde{D}_T(\lambda, d) = \tilde{D}_T(\lambda^*, d) = (\tilde{\phi}_T(\lambda^*, d), \lambda - \lambda^*) + \tilde{D}_T(\lambda, d) - \tilde{D}_T(\lambda^*, d) - (\tilde{\phi}_T(\lambda^*, d), \lambda - \lambda^*)
\]

It suffices to show that, with high probability, the first order term is lower bounded by a linear function and the second order term is lower bounded by a quadratic function, within a small neighborhood of \( \lambda^* \). For the first order term, we need the concentration of gradients.

**Lemma 2.** Under Assumptions 1-3, the concentration of the gradient in the first order term \( \tilde{\phi}_T(\lambda^*, d) \) satisfies

\[
\mathbb{P}(\|\tilde{\phi}_T(\lambda^*, d) - \nabla D(\lambda^*, d)\|_2 > \varepsilon) \leq 4m \exp(-\frac{T\varepsilon^2}{4mc_1}),
\]

for any \( \varepsilon > 0 \), where the constant \( c_1 := \sqrt{nhD} + \hat{d} \vee G \). 

12
By Lemma 2, we conclude that, with high probability, the first order term in (3.2) is lower bounded by $\langle \phi_T(\lambda^*, d), \lambda - \lambda^* \rangle \geq \langle \nabla D(\lambda^*, d), \lambda - \lambda^* \rangle - \varepsilon \| \lambda - \lambda^* \|_2 \geq -\varepsilon \| \lambda - \lambda^* \|_2$. For the second order term, we focus on a small neighborhood of $\lambda^*$. For a constant $H > 0$ (to be clarified soon), define $\Omega_\lambda(\varepsilon)$ as

$$
\Omega_\lambda(\varepsilon) := \{ (\lambda^*, \mu) | \| \lambda - \lambda^* \|_\infty \leq 4H\varepsilon, \| \mu - \mu^* \|_\infty \leq 4H\varepsilon \}. 
$$

Actually, it suffices to control the second order term for all dual variables in $\Omega_\lambda(\varepsilon)$ since we shall show that $\lambda_T^*$ belong to $\Omega_\lambda(\varepsilon)$ with a high probability depending on the value of $\varepsilon$. In order to control the shape of second order term for all dual variables in $\Omega_\lambda(\varepsilon)$, we systematically split the region $\Omega_\lambda(\varepsilon)$ to derive a uniform concentration of the second order term. The motivation for choosing this small neighborhood rather than a fixed region is that, for a convex function, the local behavior near the deterministic optimal solution is enough to guarantee the global properties of empirical optimal solutions. Consequently, for this small neighborhood, the size of its covering according to our splitting scheme can be bounded by a constant. The benefit of this local analysis is that we can successfully eliminate the $O(\log \log T)$ factor and achieve a sharper dual convergence bound. The uniform concentration of the second order term in (3.2), together with Lemma 2 enable us to derive the following result.

**Proposition 2.** Under Assumptions 1-3, given any $\varepsilon > 0$, the dual problem satisfies that for $\forall \lambda \in \Omega_\lambda(\varepsilon)$ and $\| \lambda - \lambda^* \|_2 > 2H\varepsilon$, there exists a corresponding $\lambda' \in \Omega_\lambda(\varepsilon)$ such that $\| \lambda' - \lambda^* \|_2 \geq \| \lambda - \lambda^* \|_2$ and

$$
\bar{D}_T(\lambda, d) - \bar{D}_T(\lambda^*, d) \geq \frac{\ell_D}{2} \| \lambda' - \lambda^* \|_2^2 - \ell_D H\varepsilon \| \lambda' - \lambda^* \|_2
$$

with probability at least $1 - 4m \exp\left(-\frac{T\varepsilon^2}{4mc_1}\right) - 2(2\log\left(\frac{1}{2\sqrt{2m}}\right))^2m \exp\left(-\frac{T\varepsilon^2}{2}\right)$, where

$$
H := \left(1 + 2\sqrt{2}(\sqrt{nD} + \sqrt{mG})(1 + \frac{\sqrt{m}(1-q)}{q})\right) / \ell_D, \quad \text{and}
$$

$$
q := \frac{\sqrt{m}}{\sqrt{m} + 1} \wedge \frac{\ell_D}{8\sqrt{L_2 + L_r}} \wedge \frac{1}{2}.
$$

The detailed proof is deferred to Appendix A.4. Proposition 2 delivers a strong message that, under the event where the inequality holds, the dual optimal solution $\lambda_T^*$ must be close to $\lambda^*$ in the sense that $\| \lambda_T^* - \lambda^* \|_2 \leq 2H\varepsilon$. Otherwise:

1. If $\lambda_T^*$ has $2H\varepsilon < \| \lambda_T^* - \lambda^* \|_2 \leq 4H\varepsilon$, then there will be a $\lambda_T^{*'}$ such that $\| \lambda_T^{*'} - \lambda^* \|_2 \geq \| \lambda_T^* - \lambda^* \|_2 > 2H\varepsilon$, and

$$
\bar{D}_T(\lambda_T^{*'}, d) - \bar{D}_T(\lambda^*, d) \geq \frac{\ell_D}{2} \| \lambda_T^{*'} - \lambda^* \|_2^2 - \ell_D H\varepsilon \| \lambda_T^{*'} - \lambda^* \|_2 > 0,
$$

which contradicts the optimality of $\lambda_T^*$.

2. If $\lambda_T^*$ has $\| \lambda_T^* - \lambda^* \|_2 > 4H\varepsilon$, since $\bar{D}_T(\lambda^*, d) - \bar{D}_T(\lambda^*, d) = 0$ and $\bar{D}_T(\lambda_T^*, d) - \bar{D}_T(\lambda^*, d) \leq 0$, by the convexity of $\bar{D}_T$ we have $\bar{D}_T(\lambda, d) - \bar{D}_T(\lambda^*, d) \leq 0$ for any $\lambda = \lambda^* + \alpha(\lambda_T^* - \lambda^*)$ with $0 \leq \alpha \leq 1$. Then we can always find an $\alpha$ such that $2H\varepsilon < \| \lambda - \lambda^* \|_2 \leq 4H\varepsilon$ and $\bar{D}_T(\lambda, d) - \bar{D}_T(\lambda^*, d) \leq 0$. However, according to Proposition 2 we have

$$
\bar{D}_T(\tilde{\lambda}, d) - \bar{D}_T(\lambda^*, d) \geq \frac{\ell_D}{2} \| \tilde{\lambda}' - \lambda^* \|_2^2 - \ell_D H\varepsilon \| \tilde{\lambda}' - \lambda^* \|_2 > 0,
$$

which also ends up with a contradiction.
Thus, under the event in Proposition 2 we get \( \| \lambda_T^* - \lambda^* \|_2 \leq 2H \epsilon \). Since \( 2H \epsilon \) is smaller than the radius of \( \Omega(x) \), we can safely conclude that the optimal solution \( \lambda_T^* \) lies in the small neighborhood \( \Omega(x) \). Consequently, we derive the following \( O(T^{-1}) \) bound for dual convergence.

**Theorem 1** (Dual convergence). Under Assumptions 1-3 the dual optimal solution \( \lambda^*_T \) satisfies

\[
\mathbb{E} \| \lambda_T^* - \lambda^* \|_2^2 \leq C_1 \cdot \frac{1}{T},
\]

where \( C_1 := 4H^2 \left( 16m^2 c_1 + 4(2[\log_q(\frac{1}{2\sqrt{2m}})])^{2m} \right) \).

**Proof.** By the tail expectation formula, for constant \( H > 0 \), we have

\[
\mathbb{E} \| \lambda_T^* - \lambda^* \|_2^2 = 4H^2 \int_0^\infty \mathbb{P}(\| \lambda_T^* - \lambda^* \|_2^2 > 4H^2 z)dz
\]

According to the probabilistic bound in Proposition 2 for any \( z > 0 \),

\[
\mathbb{P}(\| \lambda_T^* - \lambda^* \|_2^2 > 4H^2 z) \leq 4m \exp\left(-\frac{Tz}{4mc_1}\right) + 2(2[\log_q(\frac{1}{2\sqrt{2m}})])^{2m} \exp\left(-\frac{Tz}{2}\right).
\]

Then, calculating the integral, we get

\[
\mathbb{E}(\| \lambda_T^* - \lambda^* \|_2^2) = 4H^2 \int_0^\infty \mathbb{P}(\| \lambda_T^* - \lambda^* \|_2^2 \geq 4H^2z)dz
\]

\[
\leq \int_0^\infty \left[ 4m \exp\left(-\frac{Tz}{4mc_1}\right) + 2(2[\log_q(\frac{1}{2\sqrt{2m}})])^{2m} \exp\left(-\frac{Tz}{2}\right) \right]dz
\]

\[
= \frac{C_1}{T}
\]

\( \square \)

**Remark 1.** Our dual convergence bound is sharper than that in [Li and Ye (2021)]. Under our assumption, the \( O(T^{-1}) \) rate is unimprovable because we can find a distribution \( \mathcal{P} \in \Xi \) that incurs an \( \Omega(T^{-1}) \) dual convergence rate. Let us consider a non-regularized case when \( x \in [0,1] \) and \( f_t(x) := f(x, \xi_t) := -\frac{(x - 2\xi_t)^2}{4} + \xi_t^2 \), with the single constraint \( d = 1/2 \) and cost \( b_t = 1 \). The dual problem is

\[
D_t(\lambda) = \begin{cases} 
\frac{1}{2}\lambda & \text{if } \lambda > \xi_t \\
-\frac{1}{4} + \xi_t - \frac{1}{2}\lambda & \text{if } \lambda \leq \xi_t - \frac{1}{2} \\
\lambda^2 - 2\xi_t - \frac{1}{2}\lambda + \xi_t^2 & \text{if } \xi_t - \frac{1}{2} \leq \lambda \leq \xi_t.
\end{cases}
\]

Let \( \xi_t \) be any distribution varies within \([1/2, 3/4]\) with variance \( \sigma_\xi^2 > 0 \). Then, for any \( t \), we have \( \xi_t - 1/4 \in [1/4, 1/2] \subseteq [\xi_t - 1/2, \xi_t] \). Thus, for the sample average \( \bar{D}_T(\lambda) := T^{-1} \sum_{t=1}^T D_t(\lambda) \), when \( \lambda \in [1/4, 1/2] \), \( \bar{D}_T(\lambda) := \lambda^2 - 2(\xi_t - 1/4)\lambda + \xi_t^2 \) with the optimal solution being \( \lambda_T^* := \xi_t - 1/4 \). We have \( \mathbb{E}(\lambda_T^* - \lambda^*)^2 \geq \text{Var}(\xi_T) = \sigma_\xi^2/T \). This shows that our \( O(T^{-1}) \) dual convergence rate is indeed optimal.

Note that our Proposition 2 holds uniformly for all \( d' \in \Omega_d \). Denote the optimal solutions to problem (2.7) and (2.8), given a certain \( d' \), by \( \lambda_T^*(d') \) and \( \lambda^*(d') \), respectively. Then, we actually have

\[
\mathbb{E} \sup_{d' \in \Omega_d} \| \lambda_T^*(d') - \lambda^*(d') \|_2^2 \leq C_1 \cdot \frac{1}{T},
\]

\( \text{(3.6)} \)
Bound (3.6) plays a critical role in our regret analysis since the re-solving strategy of our adaptive algorithm framework needs to frequently update the resource constraint.

We then discuss $\epsilon$-optimal solutions of dual problem (2.7). Our following finite-sample convergence result of $\epsilon$-optimal solution can be viewed as a non-parametric version of SAA convergence developed by large deviation theory [Ruszczynski and Shapiro, 2003]. Notably, we only make assumptions on the deterministic problem $D(\lambda, d)$, and our result does not rely on restricted tail conditions such as the moment generating function in [Ruszczynski and Shapiro (2003); Shapiro et al. (2009)]. Therefore our result allows more flexible distributions.

**Theorem 2** (Convergence of dual approximate solution). Under Assumptions 1-3, suppose $X_T$ is an $\epsilon$-optimal solution that satisfies $\bar{D}_T(\lambda_T^*, d) - \bar{D}_T(\lambda_T^*, d) \leq \epsilon$. Then we have the following convergence of $\epsilon$-optimal solution:

\[
\mathbb{E} \|X_T - \lambda^*\|^2_2 \leq \frac{C_1}{T} + \frac{4\epsilon}{\sqrt{D}}
\]

Proof. Recall that, by Proposition 2, convex function $\bar{D}_T$ is larger that a quadratic function in a neighborhood of $\lambda^*$ with a high probability claimed there. Then, for any $\epsilon$ satisfying $\epsilon < 4H^2 \epsilon^2 \mathcal{L}_D$, with the same high probability, the $\epsilon$-optimal solution must belong to $\Omega_X(\epsilon)$, because, for all the points in the border $|\lambda - \lambda^*|^2 = 4H \epsilon$, we already have $\bar{D}_T(\lambda, d) - \bar{D}_T(\lambda^*, d) \geq 4H^2 \epsilon^2 \mathcal{L}_D$. Then, with the same high probability, it follows that

\[
\epsilon \geq \bar{D}_T(\lambda_T^*, d) - \bar{D}_T(\lambda_T^*, d) \geq \mathcal{L}_D \|X_T - \lambda^*\|^2_2 - \mathcal{L}_D H \epsilon \|X_T - \lambda^*\|^2_2,
\]

which suggests that $\|X_T - \lambda^*\|^2_2 \leq H \epsilon + (H^2 \epsilon^2 + 2\epsilon/\mathcal{L}_D)^{1/2}$.

Still, applying the tail expectation formula, we get

\[
\mathbb{E}(\|X_T - \lambda^*\|^2_2) = 4H^2 \int_0^{\epsilon \mathcal{L}_D \sqrt{z}} \mathbb{P}(\|X_T - \lambda^*\|^2_2 \geq 2H \sqrt{z}) dz + 4H^2 \int_{\epsilon \mathcal{L}_D \sqrt{z}}^{\infty} \mathbb{P}(\|X_T - \lambda^*\|^2_2 \geq 2H \sqrt{z}) dz \leq \frac{4\epsilon}{\mathcal{L}_D} + 4H^2 \int_{\epsilon \mathcal{L}_D \sqrt{z}}^{\infty} \mathbb{P}(\|X_T - \lambda^*\|^2_2 \geq 2H \sqrt{z}) dz.
\]

Let $2H \sqrt{z} = H \epsilon + \sqrt{H^2 \epsilon^2 + \frac{2\epsilon}{\mathcal{L}_D}}$. When $z > \frac{\epsilon}{H^2 \mathcal{L}_D}$, we have $\epsilon < 4H^2 \epsilon^2 \mathcal{L}_D$, thus $\mathbb{P}(\|X_T - \lambda^*\|^2_2 \geq 2H \sqrt{z})$ can be bounded by $4m \exp(-\frac{T \epsilon^2}{4m^2}) + 2(2\log_2(\frac{1}{2\sqrt{m}}))^{2m} \exp(-\frac{T \epsilon^2}{4})$. Also when $2H \sqrt{z} = H \epsilon + \sqrt{H^2 \epsilon^2 + \frac{2\epsilon}{\mathcal{L}_D}}$, we have $\epsilon^2 \geq z - \frac{\epsilon}{H^2 \mathcal{L}_D}$. By the integral of $z$, we get the second part of the bound. \[Q.E.D.\]

Theorem 2 explains how the approximation of dual solutions affects the dual convergence. The accuracy remains valid as we directly optimize the deterministic dual function $D(\lambda, d)$. Moreover, this theorem reveals that even if the empirical dual function $\bar{D}_T(\lambda, d)$ is not strongly convex or smooth, the dual convergence of approximate solution also holds as long as we choose an appropriate accuracy. We can further show that this property is preserved with a slightly different accuracy if we run stochastic optimization algorithms on $\bar{D}_T$. We describe the convergence of stochastic approximate solution in the following corollary:

**Corollary 1** (Convergence of stochastic dual approximate solution). Under Assumptions 1-3, suppose $X_T^*$ is a stochastic $\epsilon$-optimal solution generated by stochastic optimization algorithm $\mathcal{B}$ that satisfies

\[
\mathbb{E}_\mathcal{B} \left[ \bar{D}_T(X_T^*, d) - \bar{D}_T(X_T^*, d) \mid \bar{D}_T \right] \leq \epsilon
\]
for any given $\bar{D}_T$. Then we have the following convergence of the stochastic $\epsilon$-optimal solution:

$$
\mathbb{E} \|\lambda_T - \lambda^*\|_2^2 \leq \frac{12C_1}{T} + 3 \cdot 2^{\frac{3}{4}} \left( m \left( 2\bar{f} + \bar{r} + G \right) \right)^{\frac{1}{4}} / L_{D,3} \cdot \epsilon^2,
$$

where the expectation is taken with respect to $B$ and $P$.

Corollary 1 points out that the impact of stochastic optimization on the dual convergence is limited, and the order of dual convergence can still be controlled by $\epsilon$. Compared to Theorem 2, the smaller order $\epsilon^{\frac{3}{4}}$ could be viewed as the accuracy loss because of randomness. Even if we do not assume $\bar{D}_T$ to be strongly convex, the difference between stochastic solutions and the deterministic one $\mathbb{E} \|\lambda^*_T - \lambda^*_T\|_2^2$ is still under control just as we optimize a strongly convex function. This inspires us to apply the stochastic approximate solutions to the re-solving heuristic because, in many contexts, the benefits of stochastic algorithms greatly outweigh the lower order of convergence $\epsilon^{\frac{3}{4}}$. With the theory of dual convergence, we are ready to describe our dual-based algorithm framework for online allocation.

4 Algorithm Framework

Our algorithm extends the linear adaptive re-solving strategy in [Li and Ye (2021)] to convex objective functions. The key idea is similar to the frequent re-solving strategy in network revenue management (e.g., [Jasin and Kumar (2012); Bumpensanti and Wang (2020)]) in spirit. We keep re-solving dual problems with updated average remaining capacity inspired by the budget-ratio policy (Arlotto and Gurvich, 2019). Compared to the re-solving strategy in network revenue management, we also need to keep updating the constraints and re-solving the associated optimization programs. But the difference is that our strategy is dual-based, and the size of our optimization problems grows with time. Fortunately, the optimization in our algorithm can be easier as we only need approximate solutions. The resource control in our algorithm is handled more carefully when compared with the simple dual mirror descent. We show that, non-adaptive policies are too greedy and can’t wisely keep the remaining budget balanced in the long run. It is noteworthy that the idea of budget-ratio policy (Arlotto and Gurvich, 2019) featuring average remaining capacity update has actually been implicitly conceived in the frequent re-solving heuristic (Jasin and Kumar, 2012; Wu et al., 2015). If we rescale the variables in the frequent re-solving heuristic in Jasin and Kumar (2012) by the remaining time, we get a very similar constraint update strategy in Bumpensanti and Wang (2020).

Our dual-based online allocation algorithm is in line with other dual-based online algorithms in spirit: we keep maintaining a dual variable $\lambda_t$ and every time when a request comes, we instantly give a response based on the dual variable and the request just received. We choose the primal action $x_t$, given the dual variable $\lambda$, by:

$$
\tilde{x}_t(\lambda) := \arg \max_{x \in X} \{ f_t(x) - (\lambda + \mu)^\top b_t x \} = -\nabla f_t^*(b_t^\top (\lambda + \mu)),
$$

and the primal variable $a$ is set by

$$
\tilde{a}(\mu) := \arg \max_{a \in Z} \{ r(a) + \mu^\top a \} = -\nabla r^*(-\mu).
$$

Note that the primal solution $a$ may not explicitly affect our action $x_t$, but it is helpful for our theoretical analysis of dual-based policies and for algorithm implementation.

We outline our dual-based and history-dependent algorithm framework in Algorithm 1. The algorithm updates dual variables by solving a $t$-sample SAA as shown in equation (4.1). Each $\lambda_t$ is a $\epsilon_t$-optimal solution of the $t$-sample SAA with adaptive resources constraints $d_t$. We emphasize that two ingredients in our algorithm framework are crucial to guarantee an $O(\log T)$ regret: (1) the adaptive
update of resource constraints $d_t$; (2) the careful choice of accuracy $\epsilon_t$ for approximate dual solutions. Without the adaptive update of $d_t$, the worst-case regret will never be optimal for some extreme cases (see Section 5 for more discussion). The dual solution accuracy can be set as either increasing $\epsilon_t = \Theta(t^{-1})$ or decreasing $\epsilon_t = \Theta((T - t)^{-1})$ (or $\epsilon_t = \Theta(t^{-3/2})$, $\epsilon_t = \Theta((T - t)^{-3/2})$ for stochastic optimization algorithms). Approximate solutions help significantly alleviate the total computational cost. Our algorithm is history-dependent, meaning that we exploit all the information we have collected up to time $t$. This is the essence of our adaptive strategy. This history-dependent policy makes our algorithm learn more efficiently compared with other dual-based algorithms that do not learn from history \cite{Devanur2019, Balseiro2022}, at the cost of acceptable extra computation. As is common in the literature on dual-based online algorithms, we assume that both the conjugate $f^*_t$ and corresponding primal variable $\tilde{x}_t$ are easily attainable.

**Algorithm 1** History-based resolving algorithm framework

**Require:** regularizer $r$, iteration number $T$, start point $\lambda_0 := 0$, and initial resource $B_0 := dT$.

for all $t = 1, T$ do

Receive $(f_t, b_t) \sim \mathcal{P}$.

Calculate

$$\tilde{x}_t := \tilde{x}_t(\lambda_{t-1}) := \arg \max_{x \in X} \{f_t(x) - (\lambda_{t-1} + \mu_{t-1})^T b_t x\} = -\nabla f_t^*(b_t^T (\lambda_{t-1} + \mu_{t-1})).$$

Select $x_t := \begin{cases} \tilde{x}_t & \text{if } B_{t-1} \geq b_t x_t \\ 0 & \text{otherwise} \end{cases}$

Update remaining resources: $B_t := B_{t-1} - b_t x_t$

Update average remaining resources: $d_t := \frac{B_t}{t}$

Update dual variable $\lambda_t$ via solving the following dual problem by any approximation algorithm $\tilde{B}_t$ with accuracy $\epsilon_t$:

$$\min_{\lambda \in \Omega_{\lambda} \times \Omega_{\mu}} \left\{ \tilde{D}_t(\lambda, d_t) := \frac{1}{t} \sum_{j=1}^{t} f_j^*(b_j^T (\mu + \lambda)) + r^*(-\mu) + d_t^T \lambda \right\}$$ (4.1)

end for

Our algorithm framework is free of optimizer, that is, we can select any optimizer to get the $\epsilon_t$-optimal solution to dual program (4.1). Since the dual problem $\tilde{D}_t(\lambda, d_t)$ is generally convex with respect to $\lambda$, one favourable choice is stochastic gradient descent that is first order (recall that we assume the gradient of the dual problem, i.e., the primal variable, is easily attainable) and the computational complexity can be free of size $t$. This makes it possible to deal with large scale dual optimization when the total running time $T$ is large.

More specifically, if the dual optimizer is selected as stochastic gradient descent where the accuracy is specified by $\epsilon_t := ct^{-3/2}$, we end up with the following Algorithm \ref{alg:2} by our algorithm framework. Basically, it requires computing $O(t^3)$ stochastic gradients at time $t$. Moreover, if the dual problem $\tilde{D}_t$ is further strongly convex or smooth, we can reduce the computational cost to $O(t)$ for each time $t$. See Section 6.1 for more discussions on the case of strongly convex objectives. In Section 5 we demonstrate that any optimization algorithm $\tilde{B}_t$ that achieves the rate of dual convergence $\mathbb{E}||\lambda_t - \lambda^*(d_t)||^2 = O(t^{-1})$ or $O((T - t)^{-1})$ suffices to guarantee the optimal logarithmic regret in the end.
Algorithm 2 Resolving with Stochastic Gradient Descent

Require: regularizer \( r \), iteration number \( T \), start point \( \lambda_0 := 0 \), and initial resource \( B_0 := dT \).

for all \( t = 1, \ldots, T \) do

Receive \((f_t, b_t) \sim \mathcal{P}\).

Calculate

\[
\tilde{x}_t := \tilde{x}(\lambda_{t-1}) := \arg \max_{x \in X} \{ f_t(x) - (\lambda_{t-1} + \mu_{t-1})^\top b_t x \} = -\nabla f_t^*(b_t^\top (\lambda_{t-1} + \mu_{t-1})).
\]

Select \( x_t \) := \[
\begin{cases} 
\tilde{x}_t & \text{if } B_{t-1} \geq b_t x_t \\
0 & \text{otherwise}
\end{cases}
\]

Update remaining resources: \( B_t := B_{t-1} - b_t x_t \)

Update average remaining resources: \( d_t := \frac{B_t}{T-t} \)

Set \( R := \sqrt{m \left( \frac{2I_f + G}{d^2} \right)} \), \( L := \sqrt{md^2 + 2mb^2D^2 + nG^2} \), \( K := t^3 \), and \( \eta_t := \frac{\sqrt{2}R}{L \sqrt{K}} \). Define \( \lambda_0^t := \lambda_{t-1} \)

for all \( k = 1, \ldots, K \) do

Randomly pick \( \zeta \) from \([t] := \{1, \ldots, t\}\) with uniform distribution.

Calculate the stochastic gradient

\[
\nabla D_{\zeta}(\lambda_t^{k-1}) := \begin{bmatrix} -b_{\zeta} \tilde{x}_{\zeta}(\lambda_t^{k-1}) + d_t \\
-b_{\zeta} \tilde{x}_{\zeta}(\lambda_t^{k-1}) + \tilde{a}(\lambda_t^{k-1}) \end{bmatrix}
\] (4.2)

Update dual variable via stochastic gradient descent:

\[
\lambda_t^k := \arg \min_{\lambda \in \Omega_\lambda \times \Omega_\mu} \left\{ \langle \lambda, \nabla D_{\zeta}(\lambda_t^{k-1}) \rangle + \frac{1}{2\eta_t} \left\| \lambda - \lambda_t^{k-1} \right\|^2 \right\} (4.3)
\]

end for

Update dual variable by averaging: \( \lambda_t := \frac{1}{K} \sum_{k=1}^K \lambda_t^k \)

end for
5 Regret Analysis

5.1 Regret upper bound

In this section, we apply dual convergence established in Section 3 to derive an upper bound of regret. The result is valid for our algorithm framework Algorithm 1 with any dual optimizers. Without loss of generality, we focus on stochastic optimizers $\mathcal{B}_t$, which are independent of future arrivals $\{(f_j, b_j)\}_{j \geq t+1}$. As long as $\mathcal{B}_t$ delivers reasonably accurate dual solutions $\lambda_t$, based on past history $\mathcal{H}_{t-1} = \{f_j, b_j, x_j\}_{j=1}^{t-1}$, new arrival $(f_t, b_t)$ and updated constraint $d_t \in \Omega_d$, our adaptive framework Algorithm 1 achieves a logarithmic-order regret. Precisely, the accuracy of dual solutions shall satisfy the following condition.

**Condition 1. (Accuracy of dual solutions).** Suppose the updated constraints $\{d_j|1 \leq j \leq t\} \subseteq \Omega_d$. We say the algorithm $\{\mathcal{B}_t\}_{t \geq 1}$ satisfies dual convergence condition 1 if

$$
\mathbb{E}_{\mathcal{B},\mathcal{P}} \| \lambda_t - \lambda^\star(d_t) \|^2 \leq C_2 \frac{1}{t+1}, \quad \text{or} \quad \mathbb{E}_{\mathcal{B},\mathcal{P}} \| \lambda_t - \lambda^\star(d_t) \|^2 \leq C_2 \left(\frac{1}{t+1} + \frac{1}{T-t}\right)
$$

for some constant $C_2 > 0$. The expectation is taken with respect to all the $\{\mathcal{B}_t\}_{t \geq 1}$ and $\mathcal{P}$.

Recall that the dual convergence established in Section 3 holds uniformly for any $d \in \Omega_d$. Therefore, any dual optimizers ensuring corresponding dual solution error $\epsilon_t = \Theta(t^{-1})$ or $\epsilon_t = \Theta((T-t)^{-1})$ (or $\epsilon_t = \Theta(t^{-3/2})$ or $\epsilon_t = \Theta((T-t)^{-3/2})$ for stochastic dual optimizers) satisfy Condition 1. If Condition 1 holds, our adaptive framework Algorithm 1 achieves the following optimal regret.

**Theorem 3.** Under Assumptions 1-4, if the algorithm $\{\mathcal{B}_t\}_{t \geq 1}$ we choose satisfies Condition 1, then the regret of Algorithm 1 has the following upper bound:

$$
\text{Regret}(A) \leq \mathcal{C} \cdot \log T
$$

for some constant $\mathcal{C} > 0$ depending on the values in Assumptions 1-4.

Clearly, exact solutions to the SAA program (4.1) is a theoretically valid candidate for $\{\mathcal{B}_t\}_{t \geq 1}$, which is actually the classic idea of re-solving heuristic. However, the computational cost can be high if we want to find an exact solution. Fortunately, by Theorem 3, it suffices to approximately solve SAA program (4.1) as long as the accuracy meets conditions (5.1). We shall show in Section 5.2 that the rate $O(\log T)$ is optimal.

We now briefly sketch the proof of Theorem 3. The proof begins with the decomposition of regret, which shows that regret can be controlled by the cumulative error of dual solutions $\lambda_t - \lambda^\star(d_t)$ and $\mathbb{E}[T - \tau]$ for some stopping time $\tau$. Recall, given a certain distribution $\mathcal{P}$, the definition of regret:

$$
\text{Regret}(A|\mathcal{P}) := R^\star(\mathcal{P}) - R(A|\mathcal{P})
$$

where $R^\star(\mathcal{P})$ and $R(A|\mathcal{P})$ are defined in (2.3) and (2.2), respectively. To upper bound the regret, we need an upper bound of offline maximum reward. To that end, we define

$$
g(\lambda) := \mathbb{E} [f_t(\tilde{x}_t(\lambda)) + r(\tilde{a}(\mu)) + (\tilde{a}(\mu) - b_t\tilde{x}_t(\lambda))\top \mu^\star + (d - b_t\tilde{x}_t(\lambda))\top \lambda^\star].
$$

Here $g(\lambda)$ serves as an upper bound for $R^\star(\mathcal{P})$, characterized by the following lemma.

**Lemma 3.** The offline maximum reward $R^\star(\mathcal{P})$ satisfies $R^\star(\mathcal{P}) \leq T \cdot g(\lambda^\star)$. 

19
Proof. Recall the Lagrangian of program (2.5). By duality, we have

\[
R^*(\mathcal{P}) := \mathbb{E}_p \left[ \max_{x_t \in \mathcal{X}} \sum_{t=1}^{T} f_t(x_t) + T \cdot r \left( \frac{\sum_{t=1}^{T} b_t x_t}{T} \right), \text{ s.t. } \sum_{t=1}^{T} b_t x_t \leq dT \right]
\]

\[
\leq \mathbb{E} \sum_{t=1}^{T} \left[ f_t(x_t^*(\lambda^*)) + r(\tilde{a}(\mu^*)) + (\tilde{a}(\mu^*) - b_t x_t^*(\lambda^*))^\top \mu^* + (d - b_t x_t^*(\lambda^*))^\top \lambda^* \right]
\]

\[
= T \cdot g(\lambda^*)
\]

Since \( f_t \) and \( r(a) \) have trivial upper bounds, we get \( R^*(\mathcal{P}) \leq T(\bar{f} + \bar{r}) \). Thus, for a proper stopping time \( \tau \), we have

\[
R^*(\mathcal{P}) \leq \mathbb{E} \left[ \tau g(\lambda^*) + (T - \tau)(\bar{f} + \bar{r}) \right]. \tag{5.2}
\]

Note that our strategy of dealing the non-separable regularizer is to introduce an additional (i.e., variable split) primal variable \( a \). While its actual value does not directly affect our algorithm framework, it is vital for our theoretical investigation. To this end, denote

\[
a_t := \arg \max_{a \in \mathcal{Z}} \left\{ r(a) + \mu^\top a \right\}
\]

the value of \( a \) at \( t \)-th iteration. By Fenchel conjugate, we actually have \( a_T = -\nabla r^*(-\mu_T) \). The second impact of variable splitting is an equality constraint between \( a_T \) and \( T^{-1} \sum_{t=1}^{T} b_t x_t \). It turns out that their difference can be measured by the difference between \( \mu_T \) and the following quantity

\[
\hat{\mu}_T := \arg \max_{\mu} \left\{ r^*(-\mu) - \mu^\top \frac{\sum_{t=1}^{T} b_t x_t}{T} \right\}. \tag{5.3}
\]

The above maximization is taken without constraints, implying that \( T^{-1} \sum_{t=1}^{T} b_t x_t = -\nabla r^*(-\hat{\mu}_T) \). Due to the property of \( \nabla r^* \) (Assumption 2.3), we have

\[
\|a_T - T^{-1} \sum_{t=1}^{T} b_t x_t\|_2 \leq \bar{L}_r \|\mu_T - \hat{\mu}_T\|_2.
\]

We are in position to describe the following regret decomposition for a general stopping time.

Proposition 3. Under Assumptions 1-3, for a proper stopping time \( \tau \) ensuring that the resource is not depleted before \( t \leq \tau \), the regret of our dual-based adaptive framework Algorithm 1 admits the following upper bound:

\[
\text{Regret} (A|\mathcal{P}) \leq \mathbb{E} \left[ \sum_{t=1}^{\tau} g(\lambda^*) - g(\lambda_t) \right] + \mathbb{E} \left[ 2(\bar{f} + \bar{r} + C_3)(T - \tau) + \left\langle \lambda^*, \sum_{t=1}^{\tau} (d - b_t x_t) \right\rangle \right]
\]

\[
+ \mathbb{E} \left[ \left\langle \mu^* - \hat{\mu}_T, \sum_{t=1}^{\tau} (a_t - b_t x_t) \right\rangle \right], \tag{5.4}
\]

where \( C_3 := \sqrt{mnGD\bar{b}} \).
Lemma 4. Under Assumptions 1-3, there exists a constant $\delta_d > 0$ such that for any $d' \in \Omega_d$, if

$$-\delta_d \leq d'_i - d_i \leq \delta_d \text{ if } i \in I_B, \text{ and } d'_i - d_i \geq -\delta_d \text{ if } i \in I_{NB},$$

the dual problems $D(\lambda, d')$ and $D(\lambda, d)$ share the same binding and non-binding dimensions.

For technical convenience, we assume that, for each non-binding dimensions $i \in I_{NB}$, the updated constraint $d_{it}$ never exceeds the threshold $\bar{d}$ (the uniform bound defined in Assumption 1) at all iterations. This is a mid assumption both for theory and in practice. Indeed, if $d_{it}$ is larger than the $\bar{d}$, this means that the constraint $d_{it}$ is very loose so that its impact to the optimization problem is negligible. In this case, such a constraint can essentially be discarded.

With Lemma 4, we define the required region where binding and non-binding dimensions remain unchanged during iterations by

$$D := \left\{ d' \in \Omega_d \mid -\delta_d \leq d'_i - d_i \leq \delta_d \text{ if } i \in I_B, \text{ and } d'_i - d_i \geq -\delta_d \text{ if } i \in I_{NB} \right\}.$$  

We thereby design the following stopping time.

$$\tau := \min_{t \in [T]} \left\{ T - \left\lfloor \frac{\sqrt{nD\bar{d}}}{\bar{d}} \right\rfloor \right\} \cup \{ t \mid d_t \notin D \}. \tag{5.5}$$

Additionally, this stopping time also guarantees that resource depletion will not happen before $\tau$. We show that $\tau$ rules out early-stopping so that $\mathbb{E}[T - \tau] = O(\log T)$. The following lemmas bound the cumulative dual error and $\mathbb{E}[T - \tau]$.

Lemma 5. Under Assumptions 1-4 Algorithm 1 with selected dual optimizer $\{B_t\}_{t \geq 1}$ satisfying Condition 1 achieves

$$\mathbb{E} \left[ \sum_{t=1}^{\tau} \| \lambda_{t-1} - \lambda^* \|^2 \right] \leq O(\log T) \tag{5.6}$$

Lemma 6. Under Assumptions 1-4, the stopping time (5.5) of Algorithm 1 with selected dual optimizer $\{B_t\}_{t \geq 1}$ satisfying Condition 1 has

$$\mathbb{E} (T - \tau) \leq O(\log T) \tag{5.7}$$
These two lemmas play a key role in our regret analysis. They are proved by investigating the dynamic behavior of constraints $d_t$ for binding and non-binding dimensions, respectively. For binding dimensions, we investigate the recurrence relation of $d_t$ by leveraging the binding relations. For the non-binding dimensions, we exploit the gap between $d_t$ and average resource consumption. Equipped with Lemma 5 and 6, we now continue sketching the proof of Theorem 3. By Proposition 3, it suffices to bound the three terms there.

**Proof of Theorem 3.** The proof continues from Proposition 3.

**Step 1: bounding $R.1$.** By Fenchel conjugate, we re-write the bridging function $g(\lambda)$ by

$$
g(\lambda) = \mathbb{E} \left[ f_t(\tilde{x}_t(\lambda)) + r(\tilde{a}(\mu)) + (\tilde{a}(\mu) - b_t \tilde{x}_t(\lambda))^T \mu^* + (d - b_t \tilde{x}_t(\lambda))^T \lambda^* \right]
$$

$$= \mathbb{E} \left[ f^*_i(b^*_i(\lambda + \mu)) + r^*(-\mu) \right] + \mathbb{E} (\mu^* - \mu)^T (\tilde{a}(\mu) - b_t \tilde{x}_t(\lambda)) + \mathbb{E} (\lambda^* - \lambda)^T (d - b_t \tilde{x}_t(\lambda))
$$

$$= \mathbb{E} \left[ f^*_i(b^*_i(\lambda + \mu)) + r^*(-\mu) \right] - \mathbb{E} \left[ \nabla f^*_i(b^*_i(\lambda + \mu))^T b^*_i(\lambda + \mu) - \lambda^* - \mu^* \right]
$$

$$= \mathbb{E} \left[ D(\lambda, d) - \nabla D(\lambda, d)^T (\lambda - \lambda^*) \right]
$$

By Assumption 2 and 3, we get

$$g(\lambda^*) - g(\lambda) = D(\lambda^*, d) - D(\lambda, d) + \langle \nabla D(\lambda^*, d), \lambda - \lambda^* \rangle + \langle \nabla D(\lambda, d) - \nabla D(\lambda^*, d), \lambda - \lambda^* \rangle \leq (2\mathcal{L}_D - \mathcal{L}_D) \|\lambda - \lambda^*\|_2^2.
$$

Then Lemma 5 gives rise to the following bound.

$$\mathbb{E} \left[ \sum_{t=1}^{\tau} g(\lambda^*) - g(\lambda_{t-1}) \right] \leq O(\log T).
$$

**Step 2: bounding $R.2$.** This term can be controlled by the definition of stopping time and Lemma 4.

$$\mathbb{E} \left[ 2(\bar{f} + \bar{r} + C_3)(T - \tau) + \left\langle \lambda^*, \sum_{t=1}^{\tau} (d - b_t x_t) \right\rangle \right]
$$

$$= \mathbb{E} \left[ 2(\bar{f} + \bar{r} + C_3)(T - \tau) + \left\langle \lambda^*, d_\tau (T - \tau) - d(T - \tau) \right\rangle \right]
$$

$$\leq \mathbb{E} \left[ 2(\bar{f} + \bar{r} + C_3)(T - \tau) + \sum_{i \in l_B} \lambda_i^* (d_i + \delta_d)(T - \tau) \right]
$$

$$\leq (2\bar{f} + 2\bar{r} + 2C_3 + (\|d\| + \sqrt{m \delta_d}) \frac{2(\bar{f} + \bar{r})}{d}) \mathbb{E} (T - \tau) = O(\log T)
$$

**Step 3: bounding $R.3$.** This term requires the most effort. It concerns the combined effects of variable splitting and complementary slackness. The following lemma is important for bounding this term.

**Lemma 7.** Under Assumptions 4, 5, Algorithm 1 with selected dual optimizer $\{B_t\}_{t \geq 1}$ satisfying Condition 7 and stopping time $\{5.5\}$ ensures

$$\mathbb{E} \|\hat{\mu}_T - \mu^*\|_2^2 \leq O\left(\frac{\log T}{T}\right), \text{ and } \mathbb{E} \left\| \sum_{t=1}^{\tau} (a_t - b_t x_t) \right\|_2^2 \leq O(T \log T).$$
Thus, it holds that \( \sum_{t=1}^{\tau} (a_t - b_t x_t) \). By Cauchy–Schwarz inequality, we get

\[
\mathbb{E} \left[ \left( \mu^* - \tilde{\mu}_T, \sum_{t=1}^{\tau} (a_t - b_t x_t) \right) \right] \leq \left( \mathbb{E} \left\| \tilde{\mu}_T - \mu^* \right\|^2_2 \mathbb{E} \left\| \sum_{t=1}^{\tau} (a_t - b_t x_t) \right\|_2^2 \right)^{1/2}.
\]

(5.8)

Thus R.3 can be controlled by \( \log T \). The proof is concluded. \( \square \)

### 5.2 Lower bound and algorithms without constraint update

Bray (2019) and Li and Ye (2021) have established the logarithmic regret lower bound for online multi-secretary problems and online linear programming, respectively. To show the optimality of Theorem 3 we establish a matching lower bound in this section. We note that there always exists a regularizer function that makes our regularized online allocation problem more challenging than the non-regularized one. For example, consider that \( f_t(x) \) and \( r \) are both monotonic increasing and the hindsight optimal strategy \( \{x_t^*\}_{t=1}^{T} \) that optimizes \( \max_{x_t \in X} \{ \sum_{t=1}^{T} f_t(x_t) \text{ s.t. } \sum_{t=1}^{T} b_t x_t \leq d T \} \), it holds that \( \sum_{t=1}^{T} b_t x_t^* = d T \) and thus \( r(T^{-1} \sum_{t=1}^{T} b_t x_t^*) \geq r(T^{-1} \sum_{t=1}^{T} b_t x_t) \) for any other \( \{x_t\}_{t=1}^{T} \). This renders the regret lower bound of regularized problem larger than that of non-regularized one. Therefore, for the regret lower bound, we only focus on the non-regularized problems.

**Theorem 4 (Regret lower bound).** For any dual-based algorithm \( A \), we have the worst-case regret lower bound:

\[
\text{Regret}(A) \geq \Omega(\log T).
\]

Theorem 4 justifies the optimality of our algorithm in terms of worst-case regret. The logarithmic regret also matches that of classic unrestricted online convex optimization (Hazan et al., 2007). Nevertheless, one may wonder how important the adaptive constraint update is in our adaptive framework Algorithm 3 and whether it is possible to achieve an optimal regret without adaptive constraints update. Here we only present a negative answer partially for two specific but renowned algorithms. For concreteness, we investigate two similar algorithms (Algorithms 3 and 4) without constraints update that have been discussed in the literature for online dual gradient (mirror descent (Balseiro et al., 2021, 2022) and dual SAA (Li and Ye, 2021).

Algorithms 3 and 4 both feature the idea of approximating dual solutions, i.e., iteratively updating \( \lambda_t \) to approach \( \lambda^* \) by Stochastic Approximation (SA) or SAA. But the implementation shows two different approaches. Algorithm 3 is not history-dependent because it updates the dual variable using only the \( t \)-th sample, while Algorithm 4 is history-dependent because it gathers all the information up to time \( t \) to update the dual variable. The following lemma establishes an \( \Omega(T^{1/2}) \) regret lower bound for these two algorithms equipped with a typical stopping time.

**Theorem 5.** Under Assumptions 3 and 4, there exists a constant \( c_2 > 0 \) such that any dual-based algorithm \( A \) attempting to approximate \( \lambda^* \) with \( \mathbb{E} \| \lambda_t - \lambda^* \|_2 \leq c_2 D(t + 1)^{-1/2} \) incurs a worst-case regret lower bound:

\[
\text{Regret}(A) \geq \Omega(T^{1/2}).
\]

We prove this theorem by constructing a one-dimensional strongly convex reward and bound the regret by leveraging the probability estimation of a Binomial distribution. Note that the lower bound can also be controlled by both dual approximate error \( \mathbb{E} \sum_{t=1}^{T} \| \lambda_{t-1} - \lambda^* \|_2 \) and early stopping effect \( \mathbb{E}(T - \tau) \). Here \( \lambda^* \) is the deterministic dual solution when the resource constraint is fixed at \( d \). In sharp contrast, the dual solution \( \lambda_t \) in our adaptive framework Algorithm 1 aims to approximate \( \lambda^*(d_t) \) where \( d_t \) is the updated constraint at time \( t \). Intuitively, the rationale behind constraint update is that,
Algorithm 3 Online dual gradient (mirror) descent without constraint update

Require: regularizer \( r \), iteration number \( T \), step size \( \eta_t := \Theta(\frac{1}{t}) \) for \( t \in [T] \), start point \( \lambda_0 := 0 \), and initial resource \( B_0 := dT \).

for all \( t = 1, T \) do
    Receive \( (f_t, b_t) \sim \mathcal{P} \).
    Calculate
    \[
    \tilde{x}_t := \tilde{x}_t(\lambda_{t-1}) := \operatorname{arg \ max}_{x \in X} \{ f_t(x) - (\lambda_{t-1} + \mu_{t-1})^\top b_t x \} = -\nabla f^*_t(b_t^\top (\lambda_{t-1} + \mu_{t-1}))
    \]
    \[
    \tilde{a}_t := \operatorname{arg \ max}_{a \in Z} \{ r(a) + \mu_{t-1}^\top a \} = -\nabla r^*(-\mu_{t-1}).
    \]
    Select \( x_t := \begin{cases} 
    \tilde{x}_t & \text{if } B_{t-1} \geq b_t x_t \\
    0 & \text{otherwise}
    \end{cases} \)
    Update remaining resources: \( B_t := B_{t-1} - b_t x_t \)
    Calculate the stochastic gradient \( \nabla D_t(\lambda_{t-1}) := \begin{bmatrix} -b_t \tilde{x}_t + d \\ -b_t \tilde{x}_t + \tilde{a}_t \end{bmatrix} \)
    Update dual variable via online gradient descent:
    \[
    \lambda_t := \operatorname{arg \ min}_{\lambda \in \Omega \times \Omega} \left\{ (\lambda, \nabla D_t(\lambda_{t-1})) + \frac{1}{2\eta_t} \| \lambda - \lambda_{t-1} \|_2^2 \right\}
    \]
end for

Algorithm 4 Dual SAA without constraint update

Require: regularizer \( r \), iteration number \( T \), start point \( \lambda_0 := 0 \), and initial resource \( B_0 := dT \).

for all \( t = 1, T \) do
    Receive \( (f_t, b_t) \sim \mathcal{P} \).
    Calculate
    \[
    \tilde{x}_t := \tilde{x}_t(\lambda_{t-1}) := \operatorname{arg \ max}_{x \in X} \{ f_t(x) - (\lambda_{t-1} + \mu_{t-1})^\top b_t x \} = -\nabla f^*_t(b_t^\top (\lambda_{t-1} + \mu_{t-1})).
    \]
    Select \( x_t := \begin{cases} 
    \tilde{x}_t & \text{if } B_{t-1} \geq b_t x_t \\
    0 & \text{otherwise}
    \end{cases} \)
    Update remaining resources: \( B_t := B_{t-1} - b_t x_t \)
    Update dual variable via solving \( t \)-sample SAA:
    \[
    \lambda_t := \operatorname{arg \ min}_{\lambda \in \Omega \times \Omega} \left\{ \frac{1}{t} \sum_{j=1}^{t} f_j^*(b_j^\top (\mu + \lambda)) + r^*(-\mu) + d^\top \lambda \right\}
    \]
end for
at time $t$, the decision should be made in consideration of the remaining resources $d_t$ at hand instead of the initial resource $d$.

**Remark 2.** Theorem 5 suggests that Algorithms 3 and 4 fail to reach the optimal regret under our assumptions because they all seek to approximate a deterministic $\lambda^*$. In fact, even if we know the exact distribution $P$ and its optimal solution $\lambda^*$, we are still unable to make our dual-based algorithm optimal by just choosing $\lambda_t = \lambda^*$. Theorem 5 gives a rigorous evidence that our constraint-update algorithm outperforms other prior ones without constraint update such as the online gradient decent studied by Balseiro et al. (2021, 2022).

Finally, we remark that our theorem pushes forward the understanding of adaptiveness for online algorithms to the dual-based ones. In Arlotto and Gurvich (2019), the authors established an $\Omega(\sqrt{T})$ regret lower bound only for non-adaptive strategies (without adaptively updating the dual solutions). However, our proof demonstrates that, even when the strategy is adaptive, it might still not be sufficient to deliver an optimal regret if the algorithm only focuses on dual updates but neglects the constraint update. Actually, focusing on fixed constraints leads to a sub-optimal early stopping.

6 Applications

6.1 Strongly convex dual problems

We consider a special but practical setting, in which our empirical dual problem $\bar{D}_t(\lambda, d_t)$ in (4.1) is always $L_D$-strongly convex. This assumption can be met if $f^*$ and $r$ are almost-surely strongly convex. In this case, we only need to do stochastic gradient descent for $O(t)$ times at time $t$ to make our algorithm theoretically optimal. The detailed implementations are in Algorithm 5.

Algorithm 5 satisfies Condition 1 but it does not rely on Corollary 1. Notice that $E \| \lambda_t - \lambda^*(d_t) \|^2 \leq 2E \| \lambda_t - \lambda^*_t(d_t) \|^2 + 2E \| \lambda^*_t(d_t) - \lambda^*(d_t) \|^2$ where $\lambda^*_t(d_t)$ is the optimal solution to the empirical dual problem $\bar{D}_t(\lambda, d_t)$. The second term $E \| \lambda^*_t(d_t) - \lambda^*(d_t) \|^2$ represents the dual convergence and can be bounded by $O(t^{-1})$ by Theorem 1 while the first term accounts for the optimization error and can also be bounded by $O(t^{-1})$ (see, Rakhlin et al. (2012)). If $\bar{D}_t(\lambda, d_t)$ is further smooth, we can also ensure Condition 1 by running batch gradient descent for constant steps at each time $t$ to get an $O(t^{-1} + (T - t)^{-1})$ approximate solution, which still requires computing gradients for $O(t)$ times at $t$-th time.

6.2 Online linear programming

Our algorithm framework and theoretical results are also applicable to the classical non-regularized online linear allocation problems, which finds applications in online ad-auction (Buchbinder et al., 2007), network revenue management (Jasin and Kumar, 2012), multi-secretary problem (Kleinberg, 2005), etc. At time $t$, we make a decision $x_t \in X = [0, D]^n$ that returns a linear reward $v_t$ and bears a random cost $b_t \in \mathbb{R}^{m \times n}$ per unit. Online linear programming can be formalized as:

$$\max_{x_t} \sum_{t=1}^{T} v_t^\top x_t$$

$$\text{s.t. } \sum_{t=1}^{T} b_t x_t \leq d T, \ d \in \mathbb{R}^m$$

$$x_t \in [0, D]^n, \forall t \in [T].$$
Algorithm 5 Resolving with SGD for strongly convex dual objective

Require: regularizer \( r \), iteration number \( T \), start point \( \lambda_0 := 0 \), and initial resource \( B_0 := dT \).

for all \( t = 1, \ldots, T \) do
    Receive \((f_t, b_t) \sim P\).
    Calculate
    \[
    \tilde{x}_t := \tilde{x}_t(\lambda_{t-1}) := \arg \max_{x \in X} \{ f_t(x) - (\lambda_{t-1} + \mu_{t-1})^\top b_t x \} = -\nabla f^*_t(b_t^\top (\lambda_{t-1} + \mu_{t-1})).
    \]
    Select \( x_t := \begin{cases} 
    \tilde{x}_t & \text{if } B_{t-1} \geq b_t \tilde{x}_t \\
    0 & \text{otherwise}
    \end{cases} \)
    Update remaining resources: \( B_t := B_{t-1} - b_t x_t \)
    Update average remaining resources: \( d_t := \frac{B_t}{t} \)
    Set \( K := t \) and \( \eta_k := \frac{\ell_k}{k} \).
    Define \( \lambda^0_t := \lambda_{t-1} \)
    for all \( k = 1, \ldots, K \) do
        Randomly pick \( \zeta \) from \([t] := \{1, \ldots, t\}\) with uniform distribution.
        Calculate the stochastic gradient \( \nabla D_{\zeta}(\lambda^{k-1}_t) := \begin{bmatrix}
        -b\zeta \tilde{x}_\zeta(\lambda^{k-1}_t) + d_t \\
        -b\zeta \tilde{x}_\zeta(\lambda^{k-1}_t) + \tilde{a}(\lambda^{k-1}_t)
        \end{bmatrix} \)
        Update dual variable via stochastic gradient descent:
        \[
        \lambda^k_t := \arg \min_{\lambda \in \Omega} \left\{ \langle \lambda, \nabla D_{\zeta}(\lambda^{k-1}_t) \rangle + \frac{1}{2\eta_k} \norm{\lambda - \lambda^{k-1}_t}^2 \right\}
        \]
    end for
    Update the dual variable by \( \lambda_t := \lambda^K_t \)
end for
The empirical dual problem and its population version can be explicitly written as
\[
\tilde{D}_T(\lambda, d) := \sum_{t=1}^{T} \sum_{i=1}^{n} \left( v_{it} - b_{it}^\top \lambda \right) + d^\top \lambda, \quad \text{and} \quad D(\lambda, d) := \mathbb{E} \sum_{t=1}^{n} \left( v_{it} - b_{it}^\top \lambda \right) + d^\top \lambda,
\]
which is in line with [Li and Ye (2021)]. Here the index \( b_{it} \) means the \( i \)-th column of \( b_t \). For a given dual variable \( \lambda \), we make the primal decision by \( x_{it} := D(\lambda) v_{it} > 0 \) if the resource constraints are not violated. Then, under the same locally strongly convex and non-degeneracy assumptions, we can make optimal decisions by choosing \( \lambda_t \) as the \( O(t^{-1}) \)-optimal solution (or \( O(t^{-3/2}) \)-optimal solution for stochastic optimizer) of \( \tilde{D}_t(\lambda_t, d_t) \). Towards that end, an \( O(\log T) \) regret is attainable, which improves prior result [Li and Ye 2021].

### 6.3 Online welfare maximization with costs

Our algorithm framework is also applicable to combinatorial auctions in the existence of production costs and resource constraints. (Blum et al., 2011; Huang and Kim, 2014; Tan et al., 2020). Imagine that we run an online service system where customers arrive with a request of getting a bundle of \( n \) resources. Each customer arriving at time \( t \) has a private valuation function \( v_t = [v_{1t}, v_{2t}, \cdots, v_{nt}]^\top \) on different bundle \( i \), and each bundle \( i \) includes \( m \) types of resources \( b_i \in \mathbb{R}^m \). Denote \( b = [b_1, b_2, \cdots, b_n] \).

At every time \( t \), we make our decision \( x_t \) by choosing which bundle we would like to provide. The decision variable is \( x_t \in \{0, 1\}^n \) and \( \sum_{i=1}^{n} x_{it} \leq 1 \). The cost of consuming resources is given by a convex function \( h_T(x) \). Our goal is to optimize the total social welfare by the following mixed-integer program:

\[
\max_{x_t} \quad \sum_{t=1}^{T} v_{i_t}^\top x_t - h_T(\sum_{t=1}^{T} b x_t)
\]

s.t.
\[
\sum_{t=1}^{T} b x_t \leq d T, \quad d \in \mathbb{R}_+^m
\]

\[
\|x_t\|_1 \leq 1
\]

\[
x_t \in \{0, 1\}^n, \quad \forall t \in [T].
\]

This online program usually formulates practical problems involved in networking and cloud computing, e.g., cloud resource allocation (Dayarathna et al., 2015) and 5G network slicing (Rost et al., 2017). If the convex cost is in the form \( h_T(x) := -T \cdot r(x/T) \) for some strongly convex function \( r \), we can write the corresponding empirical dual problem and its population version as

\[
\tilde{D}_T(\lambda, d) := \frac{\sum_{t=1}^{T} \sup_{i \in [n]} (v_{it} - b_i^\top (\lambda + \mu))}{T} + r^*(\mu) + d^\top \lambda,
\]

\[
D_T(\lambda, d) := \mathbb{E} \sup_{i \in [n]} \left( r_{it} - b_i^\top \lambda \right) + r^*(\mu) + d^\top \lambda.
\]

For a given \( \lambda \), our decision is made by \( x_{it} := 1 \text{ if } i = \arg \max_{i \in [n]} \{ v_{it} - b_i^\top (\lambda + \mu) \} \). Under the similar locally strongly convex and non-degeneracy assumptions, our algorithm framework achieves an \( O(\log T) \) regret. The size of our problem is different from [Tan et al. 2020] because our algorithm focuses more on the regret given linear resources rather than the competitive ratio with highly restricted resources constraints. Here the regularizer can be interpreted as the cost function of resources, which shares an increasing marginal cost.
6.4 Online convex covering and packing problem

We apply our algorithm framework to online covering and packing problems with convex objective functions, which have been discussed in Azar et al. (2013, 2016). Consider an online context that $T$ groups of clients arrive with fixed size and, at each time $t$, we serve the $t$-th group by assigning clients to $n$ different facilities with increasing convex assignment cost $f_{it}$ and a demand $b_{it} > 0$ for each facility $i \in [n]$. Define $x_{it}$ as the number of clients that are assigned to facility $i$ at time $t$, and then $f_{it}(x_{it})$ is the corresponding assignment cost, $b_{it}x_{it}$ is the demand for facility $i$. At each time, the total service must be larger than the group size 1, i.e., $\sum_{i=1}^{n} x_{it} \geq 1$. The average maintenance cost of each facility $h_i$ is an increasing convex function to its congestion, which is the ratio of the total demands of clients assigned to the facility to the total capacity. Our goal is to minimize the sum of assignment costs and maintenance costs:

$$\min_{x_t} \quad \sum_{t=1}^{T} \sum_{i=1}^{n} f_{it}(x_{it}) + T \cdot \sum_{i=1}^{n} h_i(y_i)$$

s.t. $\sum_{t=1}^{T} b_{it}x_{it} \leq Ty_i, \quad \forall i \in [n]$ $\sum_{i=1}^{n} x_{it} \geq 1, \quad \forall t \in [T]$ $x_{it} \in [0, 1]^n, \quad \forall t \in [T], i \in [n]$ $0 \leq y_i \leq 1, \quad \forall i \in [n]$.

This is a convex and continuous variant of Capacity Constrained Facility Location (CCFL) problem (Azar et al., 2013, 2016) featuring non-negative covering and packing constraints. Here the covering constraint $\sum_{i=1}^{n} x_{it} \geq 1$ represents the minimum service requirement, and thus we can not take void actions; the packing constraints $\sum_{t=1}^{T} b_{it}x_{it} \leq Ty_i$ represents that the congestion of each facility $i$ is bounded by 1. Denote $x_t := [x_{1t}, x_{2t}, \ldots, x_{nt}]^T$ and $\mathcal{X} := \{x \in \mathbb{R}_+^n \mid \|x\| = 1\}$, we can write our convex covering and packing problem as:

$$\max_{x_t \in \mathcal{X}} \quad \sum_{t=1}^{T} f_t(x_t) + T \cdot \frac{\left(\sum_{t=1}^{T} b_t x_t \right)}{T}$$

s.t. $\sum_{t=1}^{T} b_t x_t \leq T \cdot 1$, where $f_t(x_t) := -\sum_{i=1}^{n} f_{it}(x_{it})$, $r(y) := -\sum_{i=1}^{n} h_i(y_i)$, and $b_t := \text{diag}(b_{1t}, b_{2t}, \ldots, b_{nt})$. Thus, we can similarly apply our Algorithm 1 to the CCFL problem and achieve optimal regret control.

7 Numerical Experiments

In this section, we implement Resolving with SGD as a showcase for our proposed algorithmic framework. The performance is assessed under 4 different input models. The implementation details on multiple input models are as follows: the dual update is calculated by closed-form solutions to Equation (4.3) under input I-III and by cvxpy (Diamond and Boyd, 2016) under input IV. See Table 1 for more information. For each $T$, we randomly sample $T$ observations from datasets, implement our algorithm, and calculate the regret. Only the average regret over 10 repetitions is reported. Note that we use the dual objective evaluated at the average gradient $D_T(\frac{1}{T} \sum_{t=1}^{T} \lambda_t)$ as the benchmark to compute the regret.
Table 1: Parameter Settings of Inputs

| Input | \(f_t(x)\) | \(r(x)\) | \(b_{it}\) | \(d_i\) |
|-------|-------------|------------|-------------|--------|
| I     | \(a_t^T x\) | \(-\kappa \|x - d/2\|_2^2\) | \(U(0, 1)\) | 0.1    |
| II    | \(a_t^T x\) | \(-\kappa \|x - d/2\|_2^2\) | Bernoulli(\(p_i\)) | \(U(0.25, 0.75)\) |
| III   | \(-\frac{1}{4}x^2 + \xi_t x\) | 0 | 1 | 0.5 |
| IV    | \(a_t^T x\) | \(-\kappa \sum_{i=1}^{m} \frac{x_i^2}{\sum_{i=1}^{m} d_i} \log \frac{x_i}{\sum_{i=1}^{m} d_i}\) | 1 | From dataset |

Input model I: Online welfare maximization with costs, independent reward and resource consumption. The reward functions are linear as \(f_t(x) = a_t^T x\). The regularization function is the \(\ell_2\) loss \(r(x) = -\kappa \|x - d/2\|_2^2\), which corresponds to the application of online welfare maximization with square costs. The reward coefficients \(a_t\)'s and the constraint coefficients \(b_t\)'s are i.i.d. random variables. More exactly, \(a_{it}\) is generated from the uniform distribution \(U(0, 10)\), and \(b_{it}\) is generated from the uniform distribution \(U(0, 1)\).

To illustrate how the regret scales with the time horizon \(T\), we evaluate the algorithms with different \(T\) chosen from \(\{256, 512, 1024, 1536, 2048, 2560\}\). Here \(m = 6\). We find that Resolving with SGD (Algorithm 2) shows logarithmic regret, while its counterpart without constraint update (\(d_t \equiv d\) in Equation 4.2) shows a much worse regret. We name the latter algorithm as the “Nonadaptive resolving with SGD”. The online gradient descent (OGD) method in Balseiro et al. (2020) exhibits a \(O(\sqrt{T})\) regret as indicated in their theoretical findings. The regret comparison between the algorithms can be found in Figure 1. In Figure 2, we plot the dynamic of resource consumption for one binding dimension of the aforementioned algorithms. Ten curves are displayed, each of which corresponds to one simulation. Being adaptive to the level of remaining resources, Algorithm 2 controls carefully the constraint consumption to ensure that the resources are consumed at a steady rate till they are used up. In comparison, both the OGD and the nonadaptive version of Algorithm 2 stop allocating resources too early, demonstrating the benefits of the constraint updates, which exploit the history of past actions.

Input model II: Online welfare maximization with costs, dependent reward and resource consumption. The parameter setting below is based on Balseiro et al. (2022). The reward functions and...
the regularization function are the same as in input I, whereas input II considers the case when the reward coefficients $a_t$’s are random variables conditional of the constraint coefficients $b_t$’s. We set $a_t = \text{Proj}_{[0,10]}^\top \{\theta_t^\top b_t + \delta_t 1\}$, where $\theta_t$ is generated from a multi-variate Gaussian distribution $N(0, \text{diag}(1))$, and $\delta_t$ is generated from the standard Gaussian distribution $N(0, 1)$. The constraint coefficients $b_t$’s are generated from Bernoulli distribution with probability parameter $p_t$ with $p_t = (1 + \alpha)/2$, and $\alpha$ is generated from the beta distribution $\text{Beta}(1, 3)$. The average resource constraints $d_t$’s are generated from the uniform distribution $U(0.25, 0.75)$.

Similar to the setting of input I, we evaluate the algorithms under input II with different $T$’s and fix $m = 6$. The regret performances and resource consumption are displayed in Figure 3 and Figure 4, respectively. Among the three algorithms (Algorithm 2, the nonadaptive Alg 2 and the OGD method in Balseiro et al. (2020)), Algorithm 2 achieves a logarithmic regret, the nonadaptive Alg 2 suffers from a higher regret while the regret of OGD grows in a much faster speed.

**Input model III: Non-regularized online convex resource allocation with one resource.** In this model, we assess the algorithms’ performance under a non-regularized special case, where there is only one resource, the reward function $f_t(x) = f_t(x, \xi_t) = -\frac{1}{4} x^2 + \xi_t x$, the constraint $d = \frac{1}{2}$ and cost $b_t = 1$. The random variable $\xi_t$ follows a two-point distribution that takes value in $\{\frac{1}{2}, \frac{3}{4}\}$ with equal probability,
Figure 4: Remaining resource of one binding dimension versus time under input model II with $T = 512$. Ten curves are displayed, each of which corresponds to one simulation.

Figure 5: Performance evaluation under input III. No learning and SAA are convex versions of Algorithm 1 and 2 in [Li and Ye (2021)]; OGD is online gradient descent in [Balseiro et al. (2020)]; resolving with SGD is Algorithm 2; nonadaptive resolving with SGD is the nonadaptive version of Algorithm 2.

For input model III, the optimal solution to Problem (2.8) admits a closed-form due to the simple distribution. We compare further with two algorithms: one is “No learning” and the other is “SAA”, which are the convex versions of Algorithm 1 and 2 in [Li and Ye (2021)], respectively. Both of them require the computation of optimal dual solutions, while neither Resolving with SGD (Algorithm 2) nor OGD needs this step. The regret comparison is shown in Figure 5a. All benchmark algorithms show a regret increasing in $T$ while Resolving with SGD exhibits a regret gradually stable with respect to $T$ as $T$ increases. This corroborates the theoretical results that our proposed algorithmic framework can achieve $O(\log T)$ regret and that any algorithm without constraint updates will incur $\Omega(\sqrt{T})$ regret.

We further explain the reason for the performance advantage by plotting the remaining time before stopping in Figure 5b. All benchmark algorithms stop allocating resource $O(\sqrt{T})$ steps earlier than Resolving with SGD (Algorithm 2), which leads to the terrible regret performance.

Input model IV: Display advertisement allocation with entropy regularization. We use the display advertisement dataset in [Balseiro et al. (2021)] as the last input model. Consider $m$ advertisers. In this model, $f_t(x) = q_t^\top x$, where $q_t = \{q_{1t}, ..., q_{mt}\}$ and $q_{it}$ is the expected click through rate from impression $t$ of the $i$th advertiser. The regularization function is $r(x) = -\kappa \sum_{i=1}^m \frac{q_i}{d_i} \log \frac{q_i}{\sum_{i=1}^m d_i}$, i.e., $\Pr[\xi_t = \frac{1}{2}] = \Pr[\xi_t = \frac{3}{4}] = 0.5$. This special case is used in the proof of Theorem 5.
imposing requirements of diversity and fairness on the allocation. The per-time-slot budget of the ith advertiser denoted by \( d_i \) is also given in the dataset. The consumption cost is \( b_t = 1 \). At time \( t \), only one advertiser can be assigned to the impression, i.e., \( x_t \in \{0, 1\}^m \) and \( \sum_{i=1}^{m} x_{it} \leq 1 \).

In Figure 6a, regret curves of Algorithm 2 and OGD algorithm under different \( \kappa \)s are plotted. The regret of Algorithm 2 grows slower than OGD, which shows the advantage of the proposed algorithm under this setting. It is also observed that the regret is very close for different regularization levels \( (\kappa = 0, 0.001, 0.005) \) and \( \kappa = 0.001 \) incurs the lowest regret. Trade-off between the reward (average click through rate) and the regularization term is plotted in Figure 6b.

8 Discussion

In this paper, we investigated regularized online convex allocation problems with a non-separable regularizer. While a polynomial-time adaptive algorithm framework is proved optimal in controlling regret, several interesting yet challenging questions are still open to us. One is the necessity of non-degeneracy assumption. Recently, Bumpensanti and Wang (2020) showed that the non-degeneracy assumption is not necessary for re-solving heuristic to reach a low regret under linear settings. Can a similar optimal result be achieved without the non-degeneracy assumption on constraints in the online convex allocation? Another question is on algorithm implementation. Although our algorithms are of polynomial complexity, we still wonder whether there exists any other adaptive strategy with a linear computational cost that can achieve the (sub)optimal logarithmic regret. We note that in our adaptive strategy, most of our computational complexity comes from the frequent updating of dual solutions. To reduce the computational cost, one possible approach is to reduce the updating frequency. Lastly, throughout this paper, we only discussed online convex allocation problems under the stochastic input model. The behavior of re-solving algorithms for other input models like random permutation inputs or adversarial inputs still remains largely unknown.
References

Agrawal, S. and Devanur, N. R. (2014). Fast algorithms for online stochastic convex programming. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 1405–1424. SIAM.

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

Agrawal, S., Zadimoghaddam, M., and Mirrokni, V. (2018). Proportional allocation: Simple, distributed, and diverse matching with high entropy. In *International Conference on Machine Learning*, pages 99–108. PMLR.

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

Azar, Y., Bhaskar, U., Fleischer, L., and Panigrahi, D. (2013). Online mixed packing and covering. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 85–100. SIAM.

Azar, Y., Buchbinder, N., Chan, T. H., Chen, S., Cohen, I. R., Gupta, A., Huang, Z., Kang, N., Nagarajan, V., Naor, J., et al. (2016). Online algorithms for covering and packing problems with convex objectives. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 148–157. IEEE.

Babaioff, M., Immorlica, N., Kempe, D., and Kleinberg, R. (2007). A knapsack secretary problem with applications. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, pages 16–28. Springer.

Babaioff, M., Immorlica, N., Kempe, D., and Kleinberg, R. (2008). Online auctions and generalized secretary problems. *ACM SIGecom Exchanges*, 7(2):1–11.

Balseiro, S., Lu, H., and Mirrokni, V. (2020). Dual mirror descent for online allocation problems. In *International Conference on Machine Learning*, pages 613–628. PMLR.

Balseiro, S., Lu, H., and Mirrokni, V. (2021). Regularized online allocation problems: Fairness and beyond. In *International Conference on Machine Learning*, pages 630–639. PMLR.

Balseiro, S. R. and Gur, Y. (2019). Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968.

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

Bertsimas, D., Farias, V. F., and Trichakis, N. (2011). The price of fairness. *Operations research*, 59(1):17–31.

Blum, A., Gupta, A., Mansour, Y., and Sharma, A. (2011). Welfare and profit maximization with production costs. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 77–86. IEEE.

Bray, R. (2019). Does the multisecretary problem always have bounded regret? *Available at SSRN 3497056*. 33
Buchbinder, N., Jain, K., and Naor, J. S. (2007). Online primal-dual algorithms for maximizing ad-auctions revenue. In European Symposium on Algorithms, pages 253–264. Springer.

Bumpensanti, P. and Wang, H. (2020). A re-solving heuristic with uniformly bounded loss for network revenue management. Management Science, 66(7):2993–3009.

Cooper, W. L. (2002). Asymptotic behavior of an allocation policy for revenue management. Operations Research, 50(4):720–727.

Dayarathna, M., Wen, Y., and Fan, R. (2015). Data center energy consumption modeling: A survey. IEEE Communications Surveys & Tutorials, 18(1):732–794.

Devanur, N. R. and Hayes, T. P. (2009). The adwords problem: online keyword matching with budgeted bidders under random permutations. In Proceedings of the 10th ACM conference on Electronic commerce, pages 71–78.

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

Diamond, S. and Boyd, S. (2016). Cvxpy: A python-embedded modeling language for convex optimization. The Journal of Machine Learning Research, 17(1):2909–2913.

Ferreira, K. J., Simchi-Levi, D., and Wang, H. (2018). Online network revenue management using thompson sampling. Operations research, 66(6):1586–1602.

Ghosh, A., McAfee, P., Papineni, K., and Vassilvitskii, S. (2009). Bidding for representative allocations for display advertising. In International workshop on internet and network economics, pages 208–219. Springer.

Goel, G. and Mehta, A. (2008). Online budgeted matching in random input models with applications to adwords. In Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms, pages 982–991.

Gorbunov, E., Hanzely, F., and Richtárik, P. (2020). A unified theory of sgd: Variance reduction, sampling, quantization and coordinate descent. In International Conference on Artificial Intelligence and Statistics, pages 680–690. PMLR.

Hazan, E., Agarwal, A., and Kale, S. (2007). Logarithmic regret algorithms for online convex optimization. Machine Learning, 69(2):169–192.

Hazan, E. et al. (2016). Introduction to online convex optimization. Foundations and Trends® in Optimization, 2(3-4):157–325.

Huang, Z. and Kim, A. (2014). Welfare maximization with production costs: A primal dual approach. In Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 59–72. SIAM.

Huber, P. J. (1964). Robust estimation of a location parameter. The Annals of Mathematical Statistics, 35(1):73–101.

Huber, P. J. (1967). Under nonstandard conditions. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather modification, volume 5, page 221. Univ of California Press.
Jasin, S. (2015). Performance of an lp-based control for revenue management with unknown demand parameters. *Operations Research*, 63(4):909–915.

Jasin, S. and Kumar, S. (2012). A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research*, 37(2):313–345.

Jenatton, R., Huang, J., and Archambeau, C. (2016). Adaptive algorithms for online convex optimization with long-term constraints. In *International Conference on Machine Learning*, pages 402–411. PMLR.

Kakade, S., Shalev-Shwartz, S., Tewari, A., et al. (2009). On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization. *Unpublished Manuscript, http://ttic.uchicago.edu/shai/papers/KakadeShalevTewari09.pdf*, 2(1):35.

Kim, S., Pasupathy, R., and Henderson, S. G. (2015). A guide to sample average approximation. *Handbook of simulation optimization*, pages 207–243.

Kleinberg, R. (2005). A multiple-choice secretary algorithm with applications to online auctions. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 630–631. Citeseer.

Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502.

Lee, K.-C., Jalali, A., and Dasdan, A. (2013). Real time bid optimization with smooth budget delivery in online advertising. In *Proceedings of the seventh international workshop on data mining for online advertising*, pages 1–9.

Li, X. and Ye, Y. (2021). Online linear programming: Dual convergence, new algorithms, and regret bounds. *Operations Research*.

Lobos, A., Grigas, P., and Wen, Z. (2021). Joint online learning and decision-making via dual mirror descent. In *International Conference on Machine Learning*, pages 7080–7089. PMLR.

Mahdavi, M., Jin, R., and Yang, T. (2012). Trading regret for efficiency: online convex optimization with long term constraints. *The Journal of Machine Learning Research*, 13(1):2503–2528.

McMahan, B. (2011). Follow-the-regularized-leader and mirror descent: Equivalence theorems and l1 regularization. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 525–533. JMLR Workshop and Conference Proceedings.

McMahan, H. B. (2017). A survey of algorithms and analysis for adaptive online learning. *The Journal of Machine Learning Research*, 18(1):3117–3166.

Mehta, A. (2013). Online matching and ad allocation. *Foundations and Trends® in Theoretical Computer Science*, 8(4):265–368.

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

Molinaro, M. and Ravi, R. (2014). The geometry of online packing linear programs. *Mathematics of Operations Research*, 39(1):46–59.

Nash, J. F. (1950). The bargaining problem. *Econometrica*, 18(2):155–162.
Nesterov, Y. (2009). Primal-dual subgradient methods for convex problems. *Mathematical programming*, 120(1):221–259.

Rakhlin, A., Shamir, O., and Sridharan, K. (2012). Making gradient descent optimal for strongly convex stochastic optimization. In *Proceedings of the 29th International Conference on International Conference on Machine Learning*, pages 1571–1578.

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

Rost, P., Mannweiler, C., Michalopoulos, D. S., Sartori, C., Sciancalepore, V., Sastry, N., Holland, O., Tayade, S., Han, B., Bega, D., et al. (2017). Network slicing to enable scalability and flexibility in 5g mobile networks. *IEEE Communications magazine*, 55(5):72–79.

Ruszczyński, A. and Shapiro, A. (2003). Stochastic programming models. *Handbooks in operations research and management science*, 10:1–64.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). *Lectures on stochastic programming: modeling and theory*. SIAM.

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

Tan, X., Leon-Garcia, A., Wu, Y., and Tsang, D. H. K. (2020). Online combinatorial auctions for resource allocation with supply costs and capacity limits. *IEEE Journal on Selected Areas in Communications*, 38(4):655–668.

Vazirani, U., Vazirani, V., Mehta, A., and Saberi, A. (2005). Adwords and generalized on-line matching. In *Proceedings of FOCS*.

Wu, H., Srikant, R., Liu, X., and Jiang, C. (2015). Algorithms with logarithmic or sublinear regret for constrained contextual bandits. *Advances in Neural Information Processing Systems*, 28.

Xiao, L. (2010). Dual averaging methods for regularized stochastic learning and online optimization. *Journal of Machine Learning Research*, 11:2543–2596.

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

Yuan, J. and Lamperski, A. (2018). Online convex optimization for cumulative constraints. *Advances in Neural Information Processing Systems*, 31.

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 928–936.
A Proofs of Main Results

A.1 Proof of Proposition 1

We prove the bound of the deterministic optimal solution. Consider \( \Omega'_\mu = \{-\nabla r(a) | a \in Z\} \). The bounded subgradient in Assumption 2.1 suggests that the dual variable region \( \Omega'_\mu \) we defined is bounded by \( G \). We explain this definition by the optimal conditions of stochastic programming. Note that for problem (2.8), \( \mu \) is unconstrained. The optimal condition suggests that

\[
\nabla r^*(-\mu^*) = E_b t \nabla f^*_t(b^*_t (\lambda^* + \mu^*))
\]

if we assume fubini theorem holds. Then by the Fenchel conjugate, we have \( \mu^* \in -\nabla r(E_b t x_t) \). This shows that by defining \( \Omega'_\mu \), we indeed define the possible region that contains optimal solution \( \mu^* \), i.e., \( \mu^* \in \Omega'_\mu \). Thus we have \( ||\mu||_\infty \leq G \).

For the second bound of \( ||\lambda^*||_\infty \), we only need to check that \( d^T \lambda^* \leq 2(\bar{f} + \bar{r}) \) always holds. Otherwise if \( d^T \lambda^* > 2(\bar{f} + \bar{r}) \), we have

\[
D(\lambda^*, d) \geq D(0, d),
\]

which suggests that \( \lambda^* \) is not optimal. Thus we have \( d^T \lambda^* \leq 2(\bar{f} + \bar{r}) \), i.e., \( ||\lambda^*||_\infty \leq \frac{2(\bar{f} + \bar{r})}{d} \). The bound of empirical optimal solution \( \lambda_T^* \) follows exactly the same argument.

A.2 Proof of Proposition 2

We consider

\[
D(\lambda, d) - D(\lambda^*, d) - \nabla D(\lambda^*, d)^T (\lambda - \lambda^*) = \int_0^1 [\nabla D(z(\lambda - \lambda^*) + \lambda^*, d) - \nabla D(\lambda^*, d)]^T (\lambda - \lambda^*)dz,
\]

where \( \nabla D(\lambda, d) = \begin{bmatrix} E_b t \nabla f^*_t(b^*_t (\mu + \lambda)) + d E_b t \nabla f^*_t(b^*_t (\mu + \lambda)) \nabla r^*(-\mu) \end{bmatrix} \). Then for any \( z \), we have

\[
[\nabla D(z(\lambda - \lambda^*) + \lambda^*, d) - \nabla D(\lambda^*, d)]^T (\lambda - \lambda^*)
\]

\[
\leq \left[ \left\| E_b t \nabla f^*_t(b^*_t (\mu + \lambda - \mu^* + \lambda^* - \mu^*)) - E_b t \nabla f^*_t(b^*_t (\mu^* + \lambda^*)) \right\|_2 (||\lambda - \lambda^*||_2 + ||\mu - \mu^*||_2) + \left\| \nabla r^*(-\mu) - \nabla r^*(-\mu^*) \right\|_2 ||\mu - \mu^*||_2 \right] \leq z \bar{L}_f \bar{b} E b_t (\mu + \lambda - \mu^* - \lambda^*) \left[ ||\lambda - \lambda^*||_2 + ||\mu - \mu^*||_2 \right] \leq z \bar{L}_f \bar{b}^2 (||\lambda - \lambda^*||_2 + ||\mu - \mu^*||_2^2) + \bar{L}_r \bar{z} ||\mu - \mu^*||_2^2 \leq z (2\bar{b}^2 \bar{L}_f + \bar{L}_r) ||\lambda - \lambda^*||_2^2,
\]

where the second inequality is by Assumption 3.1 when conditioned on \( b_t \) and Assumption 2.3. By the integral of \( z \) we have

\[
D(\lambda, d) - D(\lambda^*, d) - \nabla D(\lambda^*, d)^T (\lambda - \lambda^*) \leq (\bar{b}^2 \bar{L}_f + \bar{L}_r/2) ||\lambda - \lambda^*||_2^2.
\]

For the next direction we have
\[ \nabla D(z(\lambda - \lambda^*) + \lambda^*, d) - \nabla D(\lambda^*, d)]^\top (\lambda - \lambda^*) \\
= \mathbb{E} \left[ \mathbb{E} \left[ (\nabla f_t^*(b_t^\top (z(\mu + \lambda - \mu^* - \lambda^*) + \mu + \lambda^*)) - \nabla f_t^*(b_t^\top (\mu^* + \lambda^*)) , b_t\top (\mu + \lambda - \mu^* - \lambda^*)) \right| b_t \right] + (\nabla r^*(-(z(\mu - \mu^*) + \mu^*)) - \nabla r^*(-(\mu^*))) - \mu^* - \mu) \right] \\
\geq z L_f^2 \mathbb{E} \left[ \|b_t\| (\mu + \lambda - \mu^* - \lambda^*)\|^2_2 + z L_f \|\mu - \mu^*\|^2_2 \right] \geq z L_f \|\mu - \mu^*\|^2_2 + z L_f. \]

Here the first inequality is by Assumption 3.1 when conditioned on \( b_t \) and Assumption 2.2. With the inequality \( \|a + b\|^2 = \|a\|^2 + \|b\|^2 + 2(a, b) \geq \|a\|^2 + \|b\|^2 - (1 + \delta) \|a\|^2 - \frac{1}{1+\delta} \|b\|^2 = -\delta \|a\|^2 + \frac{\delta}{1+\delta} \|b\|^2 \) for any positive \( \delta \), we choose \( \delta = \frac{L_f}{L_f^2 \sigma_{\min}} \) and \( a = \mu - \mu^* \), \( b = \lambda - \lambda^* \). Then we have

\[ \|\lambda - \lambda^*\|^2 \leq D(\lambda, d) - D(\lambda^*, d) \leq L_D \|\lambda - \lambda^*\|^2, \]

where the constant \( L_D = \frac{L_f}{L_f^2 \sigma_{\min}} \), \( \bar{L} = \bar{b} \bar{L}_f + L_f/2 \).

**A.3 Proof of Lemma 2**

Since \( \phi_t(\lambda^*, d) = \begin{bmatrix} b_t \nabla f_t^*(b_t^\top (\mu^* + \lambda^*)) + d \\ b_t \nabla f_t^*(b_t^\top (\mu^* + \lambda^*)) - \nabla r^*(-(\mu^*)) \end{bmatrix} = \begin{bmatrix} \partial_\lambda D_t(\lambda^*, d) \\ \partial_\mu D_t(\lambda^*, d) \end{bmatrix} \), we consider the partial gradient of \( \lambda \), \( \mu \) separately.

For any dimension \( i \in [m] \), \( |(\partial_\lambda D_t(\lambda^*, d))_i - d_i| = |(b_t \nabla f_t^*(b_t^\top (\mu^* + \lambda^*)))_i| \leq \sqrt{n} b D \), then

\[ |(\partial_\lambda D_t(\lambda^*, d))_i| \leq \sqrt{n} b D + \bar{d}; \]

we also have

\[ |(\partial_\mu D_t(\lambda^*, d))_i| \leq \sqrt{n} b D + G, \]

According to Hoeffding’s inequality, we have

\[ \mathbb{P}(\left| (\partial (\phi_T(\lambda^*, d))_i - (\nabla D(\lambda^*, d))_i > \varepsilon / \sqrt{2m} \right| \leq 2 \exp(-\frac{\varepsilon^2}{4mc^2}) \]

for all \( i \in [2m] \).

Combining all \( 2m \) dimensions together we conclude that

\[ \mathbb{P}(\left| \frac{1}{T} \sum_{t=1}^T \phi_t(\lambda^*, d) - \nabla D(\lambda^*, d) \right| > \varepsilon \leq 2m \cdot 2 \exp(-\frac{\varepsilon^2}{4mc^2}) = 4m \exp(-\frac{\varepsilon^2}{4mc^2}). \]

**A.4 Proof of Proposition 2**

For any given \( \varepsilon > 0 \), we define the neighbourhood of \( \lambda^* \) for given \( \varepsilon \) as 3.4. We then construct a good event \( \mathcal{E}(\varepsilon) \) that only depends on \( \varepsilon \) to guarantee that under this good event, the convex function \( \hat{D}_T(\lambda, d) - \hat{D}_T(\lambda^*, d) \) is larger than a quadratic function in \( \Omega_\lambda(\varepsilon) \), which serves as a lower bound of dual function. The construction of this good event \( \mathcal{E}(\varepsilon) \) is based on the following splitting scheme and concentration of objective function:
1. Firstly we confine the first order term so that \( \tilde{\phi}_T(\lambda, d) = \frac{1}{T} \sum_{t=1}^{T} \phi_t(\lambda, d) \) will not vary too much. This can be guaranteed by the concentration inequality Lemma \[2\]

2. Then we split \( \Omega_\lambda(\varepsilon) \) into multiple cubes layer by layer and in each single cube, we control the difference of second order terms between all the \( \lambda \) in the cube and the central point of the cube.

3. Finally we uniformly control the deviation of second order terms for all central points.

For the first order term, denote event \( \mathcal{E}_0(\varepsilon) = \{ \| \tilde{\phi}_T(\lambda, d) - \nabla D(\lambda^*, d) \|_2 > \varepsilon \} \). Then by Lemma \[2\], we have \( \mathbb{P}(\mathcal{E}_0(\varepsilon)) \leq 4m \exp\left(-\frac{T \varepsilon^2}{4m^2} \right) \). Under event \( \mathcal{E}_0(\varepsilon) \), we have

\[
\langle \tilde{\phi}_T(\lambda^*, d), \lambda - \lambda^* \rangle \geq \langle \nabla D(\lambda^*, d), \lambda - \lambda^* \rangle - \varepsilon \| \lambda - \lambda^* \|_2 \geq -\varepsilon \| \lambda - \lambda^* \|_2. \tag{A.1}
\]

We now discuss the second order term of \( \tilde{D}_T(\lambda, d) \). Define the second order term

\[
s_t(\lambda, d) = D_t(\lambda, d) - D_t(\lambda^*, d) - \langle \phi_t(\lambda^* d), \lambda - \lambda^* \rangle
\]

\[
\bar{s}_T(\lambda, d) = \frac{1}{T} \sum_{t=1}^{T} s_t(\lambda, d).
\]

To derive an uniform lower bound of \( \bar{s}_T(\lambda, d) \), we do the following split on \( \Omega_\lambda(\varepsilon) \) according to \[Huber\;1967\].

Define set \( \Omega^k_\lambda(\varepsilon) = \{ (\lambda, \mu) \mid \| \lambda - \lambda^* \|_\infty \leq q^k 4H \varepsilon, \| \mu - \mu^* \|_\infty \leq q^k 4H \varepsilon \} \), \( 0 \leq k \leq N \), where \( q \in (0, 1) \) and \( N \in \mathbb{N}_+ \) will be identified later. This split divides \( \Omega_\lambda(\varepsilon) \) into \( N \) layers \( \{ \Omega^k_\lambda(\varepsilon) \}_{k=1}^{N} \) and a center cube \( \Omega^N_\lambda(\varepsilon) \). We then split each layer into disjoint cubes \( \{ \Omega^{kl}_\lambda(\varepsilon) \}_{k=1}^{N} \) with edges of length \( (1 - q)q^{k-1}4H \varepsilon \), and denote the center cube by \( \Omega^{N1}_\lambda(\varepsilon) \). \[Huber\;1967\] shows that there are at most \((2N)^{2m}\) cubes. This split is not unique to get the desired convergence order but it makes our result tighter. The center of each cube \( \Omega^{kl}_\lambda(\varepsilon) \) is \( \lambda_{kl} = (\lambda_{kl}, \mu_{kl}) \). Define \( \lambda_{kl} = \arg\max_{\lambda \in \Omega^{kl}_\lambda(\varepsilon)} \| \lambda - \lambda^* \|_2 \), and

\[
\Gamma_{kl} = \max_{\lambda \in \Omega^{kl}_\lambda(\varepsilon)} [s_t(\lambda_{kl}, d) - s_t(\lambda, d)]
\]

\[
= \max_{\lambda \in \Omega^{kl}_\lambda(\varepsilon)} [D_t(\lambda_{kl}, d) - D_t(\lambda, d) - \langle \phi_t(\lambda^* d), \lambda_{kl} - \lambda \rangle]. \tag{A.2}
\]

Then for \( k \in \{0, \ldots, N-1\} \), and \( \forall \lambda \in \Omega^{kl}_\lambda(\varepsilon) \), \( \bar{s}_T \) can be decomposed as

\[
\bar{s}_T(\lambda, d) = \frac{1}{T} \sum_{t=1}^{T} s_t(\lambda, d) - \frac{1}{T} \sum_{t=1}^{T} s_t(\lambda_{kl}, d) + \frac{1}{T} \sum_{t=1}^{T} s_t(\lambda_{kl}, d)
\]

\[
\geq \underbrace{\mathbb{E}s_t(\lambda_{kl}, d) - \mathbb{E}G_{kl}^t}_{A.3.1} + \underbrace{-\frac{1}{T} \sum_{t=1}^{T} \Gamma_{kl}^t + \mathbb{E}G_{kl}^t}_{A.3.2} + \underbrace{\frac{1}{T} \sum_{t=1}^{T} s_t(\lambda_{kl}, d) - \mathbb{E}s_t(\lambda_{kl}, d)}_{A.3.3}. \tag{A.3}
\]

We study lower bounds of these 3 terms in [A.3] respectively.
Lower bound of [A.3.1]:

\[ \mathbb{E} \Gamma_t^{kl} = \mathbb{E} \max_{\lambda \in \Omega^{kl}(\varepsilon)} \left[ f_t^*(b_t^T (\lambda_{kl} + \mu_{kl})) + r^*(-\mu_{kl}) - f_t^*(b_t^T (\lambda + \mu)) - r^*(-\mu) \right] \]

\[ - \nabla f_t^*(\lambda^* + \mu^*)^T b_t^T (\lambda_{kl} + \mu_{kl} - \lambda - \mu) + \nabla r^*(-\mu^*)^T (\mu_{kl} - \mu) \]

\[ = \mathbb{E} \max_{\lambda \in \Omega^{kl}(\varepsilon)} \left[ \int_0^1 v_1^T(\lambda) \left[ \nabla f_t^*(b_t^T (\lambda + \mu) + v_1 \cdot z) - \nabla f_t^*(b_t^T (\lambda^* + \mu^*)) \right] dz \right. \]

\[ + \left. \int_0^1 (\mu_{kl} - \mu)^T \left[ -\nabla r^*(-\mu - (\mu_{kl} - \mu) \cdot z) + \nabla r^*(-\mu^*) \right] dz \right] \]

\[ = \mathbb{E} \left[ \int_0^1 v_1^T(\tilde{\lambda}, \tilde{\mu}) \left[ \nabla f_t^*(b_t^T (\tilde{\lambda} + \tilde{\mu} + v_1 \cdot z) - \nabla f_t^*(b_t^T (\lambda^* + \mu^*)) \right] dz \right. \]

\[ + \left. \int_0^1 (\mu_{kl} - \tilde{\mu})^T \left[ -\nabla r^*(-\tilde{\mu} - (\mu_{kl} - \tilde{\mu}) \cdot z) + \nabla r^*(-\mu^*) \right] dz \right] \]

\[ \leq \max_{\lambda \in \Omega^{kl}(\varepsilon)} \|v_1\|_2 \cdot \sqrt{L_2 b} \max_{\lambda \in \Omega^{kl}(\varepsilon)} \|\lambda + \mu - \lambda^* - \mu^*\|_2 + \max_{\lambda \in \Omega^{kl}(\varepsilon)} \|\mu_{kl} - \mu\|_2 \cdot \tilde{L}_r \max_{\lambda \in \Omega^{kl}(\varepsilon)} \|\mu - \mu^*\|_2 \]

\[ \leq \sqrt{L_2 b^2} \sqrt{\tilde{L}_r} \cdot \left( \max_{\lambda \in \Omega^{kl}(\varepsilon)} \|\lambda_{kl} - \lambda\|_2 + \|\mu_{kl} - \mu\|_2 \right) \cdot \left( \|\lambda_{kl} - \lambda^*\|_2 + \|\mu_{kl} - \mu^*\|_2 \right) \]

\[ \text{(A.4)} \]

where \( v_1(\lambda) = b_t^T (\lambda_{kl} + \mu_{kl} - \lambda - \mu) \) is the direction vector, and the first inequality is obtained by Assumption 2 and Assumption 4.

According to Proposition 1, we have

\[ \mathbb{E} s_t(\lambda_{kl}, d) = D(\lambda_{kl}, d) - D(\lambda^*, d) - \nabla D(\lambda^*, d)^T \left[ \begin{array}{c} \lambda_{kl} - \lambda^* \\ \mu_{kl} - \mu^* \end{array} \right] \]

\[ \geq L_D (\|\mu_{kl} - \mu^*\|_2^2 + \|\lambda_{kl} - \lambda^*\|_2^2) \]

So for the first term we have

\[ - \mathbb{E} \Gamma_t^{kl} + \mathbb{E} s_t(\lambda_{kl}, d) \]

\[ \geq L_D (\|\mu_{kl} - \mu^*\|_2^2 + \|\lambda_{kl} - \lambda^*\|_2^2) \]

\[ - \sqrt{L_2 b^2} \sqrt{\tilde{L}_r} \cdot \left( \max_{\lambda \in \Omega^{kl}(\varepsilon)} \|\lambda_{kl} - \lambda\|_2 + \|\mu_{kl} - \mu\|_2 \right) \cdot \left( \|\lambda_{kl} - \lambda^*\|_2 + \|\mu_{kl} - \mu^*\|_2 \right) \]

\[ \text{(A.5)} \]

**Lower bound of [A.3.2]:** Since the gradients \( \|\nabla f_t^*\|_\infty, \|\nabla r^*\|_\infty \) are bounded by \( D \) and \( G \), by the integral form of \( \Gamma_t^{kl} \) in the second equality of [A.4] we also have:

\[ \left\| \Gamma_t^{kl} \right\|_2 \leq 2\sqrt{n} b D \max_{\lambda \in \Omega^{kl}(\varepsilon)} \|\lambda + \mu - \lambda_{kl} - \lambda_{kl}\| + 2\sqrt{n} G \max_{\lambda \in \Omega^{kl}(\varepsilon)} \|\mu - \mu_{kl}\|_2 \]

\[ \leq 2(\sqrt{n} b D + \sqrt{n} G) (\max_{\lambda \in \Omega^{kl}(\varepsilon)} \|\lambda - \lambda_{kl}\|_2 + \|\mu - \mu_{kl}\|_2), \]

for any \( t \in [T] \).

Define event

\[ \mathcal{E}_{kl,1}(\varepsilon) = \left\{ -\frac{1}{T} \sum_{t=1}^T \Gamma_t^{kl} + \mathbb{E} \Gamma_t^{kl} < -\varepsilon (\sqrt{n} b D + \sqrt{n} G) (\max_{\lambda \in \Omega^{kl}(\varepsilon)} \|\lambda - \lambda_{kl}\|_2 + \|\mu - \mu_{kl}\|_2) \right\}. \]

\[ \text{(A.6)} \]

Then according to Hoeffding’s inequality, \( \mathbb{P}(\mathcal{E}_{kl,1}(\varepsilon)) \leq \exp\left(-\frac{T \varepsilon^2}{2}\right) \)

**Lower bound of [A.3.3]:** We calculate the norm of each \( s_t(\lambda_{kl}, d) \):
\[ \|s_t(\lambda_{kl}, d)\|_2 = \left\| \int_0^1 v_2^T [\nabla f_t^*(b_t(\lambda^* + \mu^*) + v_2 \cdot z) - \nabla f_t^*(\lambda^* + \mu^*)] \, dz \right\|_2 \]

\[ + \int_0^1 (\mu_{kl} - \mu^*)^T [-\nabla r^*(-\mu^* - (\mu_{kl} - \mu^*) \cdot z) + \nabla r^*(-\mu^*)] \, dz \right\|_2 \]

\[ \leq 2\sqrt{n}bD(\|\lambda_{kl} - \lambda^*\|_2 + \|\mu_{kl} - \mu^*\|_2) + 2\sqrt{mg} \|\mu_{kl} - \mu^*\|_2 \]

\[ \leq 2(\sqrt{n}bD + \sqrt{mg})(\|\lambda_{kl} - \lambda^*\|_2 + \|\mu_{kl} - \mu^*\|_2) \]

for any \( t \in [T] \), where \( v_2 = b_t^T (\lambda_{kl} + \mu_{kl} - \lambda^* - \mu^*) \) is the direction vector.

Define event

\[ \mathcal{E}_{kl,2} = \left\{ \frac{1}{T} \sum_{t=1}^T s_t(\lambda_{kl}, d) - \mathbb{E}s_t(\lambda_{kl}, d) < -2\varepsilon(\sqrt{n}bD + \sqrt{mg})(\|\lambda_{kl} - \lambda^*\|_2 + \|\mu_{kl} - \mu^*\|_2) \right\}. \] (A.8)

Then we have \( \mathbb{P}(\mathcal{E}_{kl,2}) \leq \exp(-\frac{T\varepsilon^2}{2}) \) by Hoeffding’s inequality.

Now we would like to make all the quantities in the lower bound uniform by leveraging the splitting scheme. From the split, we have

\[ \max_{\lambda \in \Omega^k(\varepsilon)} \|\lambda - \lambda_{kl}\|_2 = \sqrt{m}(1 - q)q^{k-1}4H\varepsilon, \]

\[ \max_{\mu \in \Omega^k(\varepsilon)} \|\mu - \mu_{kl}\|_2 = \sqrt{m}(1 - q)q^{k-1}4H\varepsilon, \]

\[ \|\lambda^* - \lambda_{kl}\|_2 \geq q^k4H\varepsilon, \]

\[ \|\mu^* - \mu_{kl}\|_2 \geq q^k4H\varepsilon. \]

And also

\[ \|\lambda^* - \bar{\lambda}_{kl}\|_2 \leq \|\lambda^* - \lambda_{kl}\|_2 + \max_{\lambda \in \Omega^k(\varepsilon)} \|\lambda - \lambda_{kl}\|_2 \]

\[ \leq \left( 1 + \frac{\sqrt{m}(1 - q)}{q} \right) \|\lambda - \lambda_{kl}\|_2, \]

\[ \|\mu^* - \bar{\mu}_{kl}\|_2 \leq \left( 1 + \frac{\sqrt{m}(1 - q)}{q} \right) \|\mu - \mu_{kl}\|_2, \]

\[ \max_{\lambda \in \Omega^k(\varepsilon)} \|\lambda - \lambda_{kl}\|_2 \leq \frac{\sqrt{m}(1 - q)}{q} \|\lambda^* - \lambda_{kl}\|_2 \leq \frac{\sqrt{m}(1 - q)}{q} \|\lambda^* - \bar{\lambda}_{kl}\|_2 \] (so as \( \mu \)).

Thus we have the following result for the A.3.1 term in (A.5).

\[ -\mathbb{E}\Gamma_t^{kl} + \mathbb{E}s_t(\lambda_{kl}, d) \geq L_D(\|\mu_{kl} - \mu^*\|_2^2 + \|\lambda_{kl} - \lambda^*\|_2^2) \]

\[ - \sqrt{L_2b^2 \vee L_r} \cdot (\max_{\lambda \in \Omega^k(\varepsilon)} \|\lambda_{kl} - \lambda\|_2 + \|\mu_{kl} - \mu\|_2) \cdot (\|\lambda_{kl} - \lambda^*\|_2 + \|\mu_{kl} - \mu^*\|_2) \]

\[ \geq \frac{L_D}{(1 + \frac{\sqrt{m}(1 - q)}{q})^2} \|\lambda_{kl} - \lambda^*\|_2^2 - 2\sqrt{m}(1 - q) \cdot \sqrt{L_2b^2 \vee L_r} \|\lambda_{kl} - \lambda^*\|_2^2 \]

So there exists \( q = \frac{\sqrt{m}E_D}{\sqrt{L_2b^2 \vee L_r}} \) such that when \( q \geq q^*, \frac{\sqrt{m}(1 - q)}{q} \leq 1 \wedge \frac{L_D}{8\sqrt{L_2b^2 \vee L_r}} \), and

\[ \frac{L_D}{(1 + \frac{\sqrt{m}(1 - q)}{q})^2} - 2\sqrt{m}(1 - q) \cdot \sqrt{L_2b^2 \vee L_r} \geq \frac{L_D}{2}. \]
Choose \( q = q' \cup \frac{1}{2} \). Then for the \( A.3 \) term in \( A.5 \) we have

\[
- \varepsilon \Gamma_{kl}^t + \mathbb{E} \mathbb{E}_t(\lambda_{kl}, d) \geq \frac{L_D}{2} (\|\mu_{kl} - \mu^*\|^2_2 + \|\lambda_{kl} - \lambda^*\|^2_2) = \frac{L_D}{2} \|\lambda_{kl} - \lambda^*\|^2_2. \tag{A.9}
\]

For \( A.3 \), under event \( \mathcal{E}_{k,l,1}^c(\varepsilon) \) in \( A.6 \) we have

\[
- \frac{1}{T} \sum_{t=1}^T \Gamma_{kl}^t + \mathbb{E} \mathbb{E}_t(\lambda_{kl}, d) \geq - 2 \varepsilon (\sqrt{n} \bar{b} D + \sqrt{m} G) (\max_{\lambda \in \Omega_{kl}(\varepsilon)} \|\lambda - \lambda_{kl}\|_2 + \|\mu - \mu_{kl}\|_2)
\geq - 2 \sqrt{2} \varepsilon (\sqrt{n} \bar{b} D + \sqrt{m} G) \frac{\sqrt{m}(1 - q)}{q} \|\lambda_{kl} - \lambda^*\|_2.
\tag{A.10}
\]

For \( A.3 \), under event \( \mathcal{E}_{k,l,2}^c(\varepsilon) \) in \( A.8 \) we have

\[
\frac{1}{T} \sum_{t=1}^T s_t(\lambda_{kl}, d) - \mathbb{E} s_t(\lambda_{kl}, d) \geq - 2 \sqrt{2} \varepsilon (\sqrt{n} \bar{b} D + \sqrt{m} G) \|\lambda_{kl} - \lambda^*\|_2.
\tag{A.11}
\]

Now we combine first order lower bound in \( A.1 \) and second order lower bound in \( A.9 \), \( A.10 \), \( A.11 \) together under the desired good event

\[
\mathcal{E}(\varepsilon) = \mathcal{E}_0^c(\varepsilon) \cap \bigcap_{k=1}^N (\mathcal{E}_{k,1}^c(\varepsilon) \cap \mathcal{E}_{k,2}^c(\varepsilon)),
\]

where we choose \( N \) by setting the radius of \( \hat{\Omega}^{N-1}(\varepsilon) : \sqrt{2} m q^N 4 H \varepsilon \leq 2 H \varepsilon \), i.e., \( N = \lceil \log_q(\frac{1}{2 \sqrt{2m}}) \rceil \).

Under \( \mathcal{E}(\varepsilon) \), for any \( \lambda \in \Omega_{\lambda}(\varepsilon) \) satisfying \( \|\lambda - \lambda^*\|_2 > 2 H \varepsilon \), there exists \( k \in \{0, \ldots, N - 1\} \) and \( l \) such that \( \lambda \in \hat{\Omega}_{kl}(\varepsilon) \), and

\[
D_T(\lambda, d) - \tilde{D}_T(\lambda^*, d) = \langle \bar{\phi}_T(\lambda^*, d), \lambda - \lambda^* \rangle + \bar{s}_T(\lambda, d)
\geq \frac{L_D}{2} \|\lambda_{kl} - \lambda^*\|^2_2 - 2 \sqrt{2} \varepsilon (\sqrt{n} \bar{b} D + \sqrt{m} G) \frac{\sqrt{m}(1 - q)}{q} \|\lambda_{kl} - \lambda^*\|_2
- 2 \sqrt{2} \varepsilon (\sqrt{n} \bar{b} D + \sqrt{m} G) \|\lambda_{kl} - \lambda^*\|_2 - \varepsilon \|\lambda_{kl} - \lambda^*\|_2
= \frac{L_D}{2} \|\lambda_{kl} - \lambda^*\|^2_2 - H \varepsilon : L_D \|\lambda_{kl} - \lambda^*\|_2
\]

where \( H = \left( 1 + 2 \sqrt{2} (\sqrt{n} \bar{b} D + \sqrt{m} G)(1 + \frac{\sqrt{m}(1 - q)}{q}) \right) / L_D \).

Compute the probability of \( \mathcal{E}(\varepsilon) \) we can show that

\[
\mathbb{P}(\mathcal{E}(\varepsilon)) \geq 1 - \mathbb{P}(\mathcal{E}_0(\varepsilon)) - \sum_{0 \leq k \leq N - 1, \ l} (\mathbb{P}(\mathcal{E}_{k,1}(\varepsilon)) + \mathbb{P}(\mathcal{E}_{k,2}(\varepsilon)))
\geq 1 - 4 m \exp\left(\frac{T^2 \varepsilon^2}{4 mc_1} - 2 (2 \lceil \log_q(\frac{1}{2 \sqrt{2m}}) \rceil)^2 m \exp\left(\frac{T^2 \varepsilon^2}{2}\right)\right)
\]

**A.5 Proof of Corollary 1**

Recall the proof of Theorem 2 that when \( \varepsilon \) satisfying \( \varepsilon < 4 H^2 \varepsilon^2 L_D \), with high probability the deterministic \( \varepsilon \)-optimal solution must be in \( \Omega_{\lambda}(\varepsilon) \). Similarly, for the stochastic \( \varepsilon \)-optimal solution, we try to confine it in a larger region so that with high probability \( \mathbb{E} [\|\lambda_{T} - \lambda_{T}^*\|^2_2 \tilde{D}_T] \) can still be bounded by \( \varepsilon \). Notice that, although our Proposition 2 only focus on \( \Omega_{\lambda}(\varepsilon) \), it also bring us information outside \( \Omega_{\lambda}(\varepsilon) \). For any \( \varepsilon \) and \( \varepsilon' \), under the event when Proposition 2 holds, for any \( \tilde{D}_T \) we have:

1. If \( \tilde{D}_T(\lambda_{T}, d) - \tilde{D}_T(\lambda^*, d) \leq 4 H^2 \varepsilon^2 L_D \), then \( \|\lambda_{T} - \lambda_{T}^*\|_2 \leq 4 H \varepsilon \).
2. If $\bar{D}_T(\lambda_T, d) - \bar{D}_T(\lambda^*, d) > 4H^2\varepsilon^2 L_D$, then we have $\|\lambda_T - \lambda^*_T\|_2 \leq \frac{1}{H\varepsilon L_D} (\bar{D}_T(\lambda_T, d) - \bar{D}_T(\lambda^*, d))$. Because the convex function $\bar{D}_T(\lambda, d) - \bar{D}_T(\lambda^*, d) = 0$ when $\lambda = \lambda^*$, and $\bar{D}_T(\lambda, d) - \bar{D}_T(\lambda^*, d) \geq 4H^2\varepsilon^2 L_D$ when $\|\lambda - \lambda^*\|_2 = 4H\varepsilon$.

We conclude that under the event when Proposition 2 holds, for any $\varepsilon < 4H^2\varepsilon^2 L_D$,

$$
\mathbb{E}_B \left[ \|\lambda_T^e - \lambda^*\|_2^2 | \bar{D}_T \right] \leq 16H^2\varepsilon^2 + 2\sqrt{m \left( \frac{2\bar{f} + \bar{r}}{d} + G \right)} \cdot \varepsilon
$$

because $\|\lambda_T^e - \lambda^*\|_2 \leq 2\sqrt{m \left( \frac{2\bar{f} + \bar{r}}{d} + G \right)}$. The RHS term has a minimum value

$$
z_0 = 3 \cdot 2^4\varepsilon^2 \left( m \left( \frac{2\bar{f} + \bar{r}}{d} + G \right) \right) / L_D^2
$$

when $\varepsilon_0 = \varepsilon^2 \left( m \left( \frac{2\bar{f} + \bar{r}}{d} + G \right) \right) / (2^4H\varepsilon^2 L_D^2)$. When the RHS term is larger than is minimum value, we can always take the corresponding $\varepsilon$ at the right side where $\varepsilon > \varepsilon_0$ and it follows that

$$
z = 16H^2\varepsilon^2 + 2\sqrt{m \left( \frac{2\bar{f} + \bar{r}}{d} + G \right)} \cdot \varepsilon \leq 48H^2\varepsilon^2.
$$

Then by the tail expectation formula we have

$$
\mathbb{E}_{B,P} \|\lambda_T^e - \lambda^*\|_2^2 = \int_0^{z_0} \mathbb{P}(\mathbb{E}_B \left[ \|\lambda_T - \lambda^*\|_2^2 | \bar{D}_T \right] \geq z) dz + \int_{z_0}^{\infty} \mathbb{P}(\mathbb{E}_B \left[ \|\lambda_T^e - \lambda^*\|_2^2 | \bar{D}_T \right] \geq z) dz
$$

$$
\leq z_0 + \int_{z_0}^{\infty} \left[ 4m \exp(- \frac{T^2}{48H^2 m_c^2}) + 2(2|\log_2(\frac{1}{2\sqrt{2m}})|)^2 m \exp(- \frac{T^2}{2m_c^2}) \right] dz.
$$

$$
\leq z_0 + 12C_1 T.
$$

#### A.6 Proof of Proposition 3

By Fenchel conjugate, the definition of $\mu_T$ implies

$$
r(\frac{\sum_{t=1}^T b_t x_t}{T}) + \mu_T^T \frac{\sum_{t=1}^T b_t x_t}{T} = \max_a r(a) + \mu_T^T a = r^*(-\mu_T) \geq r\left( \frac{\sum_{t=1}^T a_t}{T} \right) + \mu_T^T \frac{\sum_{t=1}^T a_t}{T}
$$

$$
\geq \frac{\sum_{t=1}^T r(a_t)}{T} + \mu_T^T \frac{\sum_{t=1}^T a_t}{T}.
$$

Combined with $R(A|P) = \mathbb{E}_{A,P} \left[ \sum_{t=1}^T f_t(x_t) + T \cdot r\left( \frac{\sum_{t=1}^T b_t x_t}{T} \right) \right]$, we have

$$
R(A|P) \geq \mathbb{E} \left[ \sum_{t=1}^T f_t(x_t) + \sum_{t=1}^T r(a_t) + \mu_T^T \left( \sum_{t=1}^T (a_t - b_t x_t) \right) \right].
$$

The Assumption 2 suggests that

$$
\|\mu_T\|_2 \leq \sqrt{m} G, \quad \|a_t\|_2 = \|\nabla r^*(-\mu_t)\|_2 \leq \sqrt{n} D\bar{b}.
$$

43
Thus
\[
R(A|\mathcal{P}) \geq \mathbb{E} \left[ \sum_{t=1}^{\tau} \left[ f_t(\hat{x}_t(\lambda_{t-1})) + r(a_t) \right] - (\hat{f} + \tilde{r} + 2\sqrt{\text{mn}GD\tilde{b}})(T - \tau) + \left\langle \mu_T, \sum_{t=1}^{\tau} (a_t - b_tx_t) \right\rangle \right]
\]
\[
= \mathbb{E} \left[ \sum_{t=1}^{\tau} g(\lambda_{t-1}) + \left\langle \mu_T - \mu^*, \sum_{t=1}^{\tau} (a_t - b_tx_t) \right\rangle \right]
\]
\[-\mathbb{E} \left[ (\hat{f} + \tilde{r} + 2C_3)(T - \tau) - \left\langle \lambda^*, \sum_{t=1}^{\tau} (d - b_tx_t) \right\rangle \right]
\]
Combined with (5.2), we conclude the proof.

A.7 Proof of Lemma 4

We start with a lemma on the continuity of dual optimal solution to prove Lemma 4.

Lemma 8 (Continuity of dual optimal solution (Li and Ye 2021)). Under Assumption 1 2 3 for the stochastic program \(\min_{\mu, \lambda \geq 0} D(\lambda, d') = \mathbb{E} f^*_t(b^T_\tau (\mu + \lambda) + r^*(-\mu) + d^T \lambda)\), let \(d'\) be \(d'_1, d'_2 \in \Omega_d\) separately, then the corresponding optimal solution \(\lambda^*(d'_1), \lambda^*(d'_2)\) satisfies
\[
\|\lambda^*(d'_1) - \lambda^*(d'_2)\|_2^2 \leq \frac{1}{4L_D^2} \sum_{i \in I_B} (d'_{1i} - d'_{2i})^2,
\]

If further \(d'_1, d'_2\) identify the same binding/non-binding dimensions, then
\[
\|\lambda^*(d'_1) - \lambda^*(d'_2)\|_2^2 \leq \frac{1}{4L_D^2} \sum_{i \in I_B} (d'_{1i} - d'_{2i})^2,
\]

where the binding dimension \(I_B\) is with respect to \(d'_1\) and \(d'_2\).

Proof (Lemma 8). By Proposition 1 and the uniform assumption on \(d\), we have
\[
D(\lambda^*(d'_2), d'_1) - D(\lambda^*(d'_1), d'_1) \geq L_D \|\lambda^*(d'_2) - \lambda^*(d'_1)\|_2^2
\]
\[
D(\lambda^*(d'_1), d'_2) - D(\lambda^*(d'_2), d'_2) \geq L_D \|\lambda^*(d'_1) - \lambda^*(d'_2)\|_2^2.
\]

Summing up two inequality we have
\[
(d'_1 - d'_2)^T(\lambda^*(d'_2) - \lambda^*(d'_1)) \geq 2L_D \|\lambda^*(d'_2) - \lambda^*(d'_1)\|_2^2,
\]
or equivalently, \(\sum_{i \in I_B} (d'_{1i} - d'_{2i})(\lambda^*_i(d'_2) - \lambda^*_i(d'_1)) \geq 2L_D \|\lambda^*(d'_2) - \lambda^*(d'_1)\|_2^2\) if further \(d'_1, d'_2\) share the same binding/non-binding dimensions. From (A.12), we can show that
\[
2L_D \|\lambda^*(d'_2) - \lambda^*(d'_1)\|_2^2 \leq (d'_1 - d'_2)^T(\lambda^*(d'_2) - \lambda^*(d'_1)) \leq \|d'_1 - d'_2\|_2 \|\lambda^*(d'_2) - \lambda^*(d'_1)\|_2
\]
\[
\|\lambda^*(d'_2) - \lambda^*(d'_1)\|_2 \leq \frac{1}{2L_D} \|d'_1 - d'_2\|_2.
\]

Thus we get the first statement. For the second statement we focus on the binding dimensions
\[
2L_D \|\lambda^*(d'_2) - \lambda^*(d'_1)\|_2^2 \leq \sum_{i \in I_B} (d'_{1i} - d'_{2i})(\lambda^*_i(d'_2) - \lambda^*_i(d'_1))
\]
\[
\leq \sqrt{\sum_{i \in I_B} (d'_{1i} - d'_{2i})^2} \sqrt{\sum_{i \in I_B} (\lambda^*_i(d'_2) - \lambda^*_i(d'_1))^2}
\]
\[
\leq \sqrt{\sum_{i \in I_B} (d'_{1i} - d'_{2i})^2} \|\lambda^*(d'_2) - \lambda^*(d'_1)\|_2,
\]
which completes the proof of Lemma 8.
Then, we return to Lemma 4 and consider the original constraints \(d\) and the its binding/non-binding dimensions: \(I_B = \{i | d_i - \mathbb{E}(b_t x_t(\lambda^*))_i = 0\}\), and \(I_{NB} = \{i | d_i - \mathbb{E}(b_t x_t(\lambda^*))_i > 0\}\). Here we write the corresponding optimal solution to \(\min_{\mu, \lambda \geq 0} D(\lambda, d)\) as \(\lambda^*\), and write \(\lambda^*(d')\) if we change \(d\) to \(d'\). Then if \(i \in I_B\) and \(i\) changes to non-binding dimensions for \(d'\), by Lemma 8 we have

\[
\|d - d'\|_2 \geq 2L_D \|\lambda^*(d') - \lambda^*\|_2 \geq 2L_D \lambda^*,
\]

where \(\lambda = \min \{\lambda^*_i | i \in I_B\}\). If on the other hand, \(i \in I_{NB}\) and \(i\) changes to binding dimensions for \(d'\), by Assumption 3.1 we have

\[
\lambda_i^* = \min d_i - \mathbb{E}(b_t x_t(\lambda^*))_i = \frac{1}{2b^2 L_f} \lambda_i^* (|d_i - d_i' - |d_i' - d_i|).
\]

Denote the minimum of remaining resources in non-binding dimensions by

\[
\gamma = \min_{i \in I_{NB}} \{d_i - \mathbb{E}(b_t x_t(\lambda^*))_i\}.
\]

By Lemma 8 we have

\[
\|d - d'\|_2 \geq 2L_D \mathbb{E}[\|\lambda^*(d') - \lambda^*\|_2] \geq \frac{L_D}{b^2 L_f} (\gamma - |d_i' - d_i|) \geq \frac{L_D}{b^2 L_f} (\gamma - \|d - d'\|_2),
\]

i.e., \(\|d - d'\|_2 \geq \frac{\gamma L_D}{L_D + b^2 L_f}\). Combined with (A.13), taking \(\delta_d = \frac{1}{\sqrt{m}} \left(\frac{\gamma L_D}{L_D + b^2 L_f}\right) \wedge (2L_D \lambda)\) we can conclude that when \(|d_i - d_i'| \leq \delta_d\), the binding/non-binding dimensions will never change. Moreover, enlarging the constraint in a non-binding dimension will never change this constraint to binding dimension. So, for the non-binding dimensions, \(d_i' - d_i\) can be any large. This finishes the proof.

### A.8 Proof of lemma 5

\[
\mathbb{E} \left[ \sum_{t=1}^\tau \|\lambda_{t-1} - \lambda^*\|_2^2 \right] \leq 2 \mathbb{E} \left[ \sum_{t=1}^\tau \|\lambda_{t-1} - \lambda^*(d_{t-1})\|_2^2 + \|\lambda^*(d_{t-1}) - \lambda^*\|_2^2 \right]
\]

By the definition of stopping time \(\tau\) and Condition 1, the first term in the RHS has

\[
\mathbb{E} \left[ \sum_{t=1}^\tau \|\lambda_{t-1} - \lambda^*(d_{t-1})\|_2^2 \right] \leq \sum_{t=1}^\tau 2C_2 \frac{1}{T} \text{ or } C_2 \frac{1}{T - t + 1} \leq 2C_2(\log T + 1)
\]

For the second term, we apply lemma 8 to it.

\[
2 \mathbb{E} \left[ \sum_{t=1}^\tau \|\lambda^*(d_{t-1}) - \lambda^*\|_2^2 \right] \leq \frac{1}{2L_D^2} \mathbb{E} \left[ \sum_{t=1}^\tau \sum_{i \in I_0} (d_{it} - d_i)^2 \right].
\]

Thus we transform the perturbation of \(\lambda^*(d_{t})\) into the derivation of \(d_t\) in the binding dimensions.

To ease our analysis, we define a new sequence \(d'_t\)

\[
d'_t = \begin{cases} 
    d_t, & \text{if } t \leq \tau \\
    d_{t-1}, & \text{if } t > \tau
\end{cases}
\]
which shares the same stopping time with \( d_t \) and define \( \tau_i = \min\{ T - \left\lceil \frac{mb}{d} \right\rceil \} \cup \{ t|d'_t \notin D_i \} \) for \( i \in [m] \) as the stopping time on each dimension with \( \tau = \min\{ \tau_1, ..., \tau_m \} \).

We first consider the binding dimensions. For any \( i \in I_B \), we follow a similar procedure in \cite{Li and Ye 2021} to derive:

\[
d'_{i,t+1} = d'_{it} + \frac{d'_{it} - (b_{t+1}\tilde{x}_{t+1}(\lambda_{t+1}))}{T - t - 1} \mathbb{I}(\tau > t)
\]

\[
\mathbb{E}\left( d'_{t,t+1} - d_i \right)^2 = \mathbb{E}\left( d'_{it} - d_i \right)^2 + \mathbb{E}\left( \frac{d'_{it} - (b_{t+1}\tilde{x}_{t+1}(\lambda_{t+1}))}{T - t - 1} \mathbb{I}(\tau > t) \right)^2
\]

\[
+ 2\mathbb{E}\left( \frac{d'_{it} - d_i (b_{t+1}\tilde{x}_{t+1}(\lambda^*(d_i)) - b_{t+1}\tilde{x}_{t+1}(\lambda_{t+1}))}{T - t - 1} \right) \mathbb{I}(\tau > t)
\]

For the term \( A' \) we have \( A' \leq \frac{(\tilde{d} + \sqrt{n}Db)^2}{(T - t - 1)^2} \). For the term \( B' \), since \( i \in I_B \) and \( d_i \in \sigma(H_t) \), conditioned on past history \( H_t \), we always have \( \mathbb{E}[d'_{it} - d_i (b_{t+1}\tilde{x}_{t+1}(\lambda^*(d_i)))] \mathbb{I}(\tau > t) | H_t = 0 \), thus \( B' = 0 \). For the term \( C' \), we apply Assumption [3] and Condition [1]

\[
C' \leq 2\mathbb{E}\left( d'_{it} - d_i \right)^2 \sqrt{\mathbb{E}\|b_{t+1}\tilde{x}_{t+1}(\lambda^*(d_i)) - b_{t+1}\tilde{x}_{t+1}(\lambda_{t+1})\|_2^2}
\]

\[
\leq 2\mathbb{E}\left( d'_{it} - d_i \right)^2 \sqrt{2b^4L_2 \mathbb{E}\|\lambda_{t} - \lambda^*(d_i)\|_2^2} \leq \frac{2\sqrt{2C_2L_2b^2}}{T - t - 1}. 
\]

Here the second inequality is because of Assumption [3] and the third inequality is from Condition [1].

Here in the derivation, we can treat \( \{\lambda_t\} \) as a new sequence generated by \( \{d'_t\} \), which has the same value with the original one when \( t \leq \tau \), and takes \( \lambda_t = B_t(\mathcal{H}_t, d'_t) \) when \( t > \tau \). Then we get the recurrence relation of \( d'_t - d_i \):

\[
\mathbb{E}\left( d'_{t,t+1} - d_i \right)^2 \leq \mathbb{E}\left( d'_{it} - d_i \right)^2 + \frac{\tilde{d} + \sqrt{n}Db}{(T - t - 1)^2} + \frac{2\sqrt{2C_2L_2b^2}}{T - t - 1} \mathbb{E}\left( d'_{it} - d_i \right)^2.
\]

Since \( d_0 = d \), by induction we have \( \mathbb{E}\left( d'_{it} - d_i \right)^2 \leq C_3 \frac{T + 1}{(T + 1)(T - t)} \), where

\[
C_3 = \left( 2 \cdot \left( \tilde{d} + \sqrt{n}Db \right)^2 \vee \left( 2\sqrt{2C_2L_2b^2} \right) + 1 \right)^2.
\]

So, we have

\[
2\mathbb{E}\left[ \sum_{t=1}^{\tau} \|\lambda^*(d_{t-1}) - \lambda^*\|^2 \right] \leq 2\mathbb{E}\left[ \sum_{t=1}^{\tau} \sum_{i \in I_B} (d_{i,t-1} - d_i)^2 \right]
\]

\[
\leq 2m\mathbb{E}\sum_{t=1}^{T} \left[ (d'_{i,t+1} - d_i)^2 \right] \leq 2mC_3 \log T, \text{ and}
\]

\[
\mathbb{E}\left[ \sum_{t=1}^{\tau} \|\lambda_{t-1} - \lambda^*\|^2 \right] \leq \left( 2C_2 + \frac{mC_3}{L^2b^2} \right) \log T + 2C_2,
\]

which completes the proof.
A.9 Proof of lemma\textsuperscript{6}

Since \( \tau = \min\{\tau_1, \ldots, \tau_m\} \), we only need to show \( \mathbb{E}(T - \tau_i) \leq C \log T \) for any \( i \) in binding dimensions and non-binding dimensions.

For the binding dimensions, applying Chebyshev’s inequality, we have

\[
\mathbb{E}(T - \tau_i) \leq \sum_{i=1}^{T} P(\tau_i \leq t) \leq 1 + \frac{\sqrt{nDB}}{d} + \sum_{i=1}^{T} P\left(|d'_it - d_i| \leq \delta_d\right)
\]

\[
\leq 1 + \frac{\sqrt{nDB}}{d} + \frac{\sum_{t=1}^{T} \mathbb{E}(d'_it - d_i)^2}{\delta_d^2}
\]

\[
\leq 1 + \frac{\sqrt{nDB}}{d} + \frac{C_3}{\delta_d^2} \log T \tag{A.14}
\]

For the non-binding dimensions, \( D \) ensures that binding/non-binding dimensions remain unchanged when \( d' \in D \). Then for \( d' \in D \), we define

\[
d'_i = \begin{cases} d'_i, & \text{if } i \in I_B \\ d_i - \delta_d, & \text{if } i \in I_{NB} \end{cases}
\]

We know that \( \lambda^*(d') = \lambda^*(\tilde{d}') \) because the non-binding constraints are loose, then

\[
\mathbb{E}\left(b_t \tilde{x}_t(\lambda^*(d'))\right)_i = \mathbb{E}\left(b_t \tilde{x}_t(\lambda^*(\tilde{d}))\right)_i < d_i - \delta_d
\]

Recall that \( \tilde{x}_t(\cdot) \perp \mathcal{H}_{t-1} \), thus \( \mathbb{E}\left[(b_t \tilde{x}_t(\lambda^*(d_{t-1})))_i | \mathcal{H}_{t-1}\right] < d_i - \delta_d \) for \( i \in I_{NB} \) and \( d_{t-1} \in D \). This implies that

\[
\mathbb{P}(\tau_i \leq t) = \mathbb{P}\left(\sum_{j=1}^{t'} (b_j \tilde{x}_j(\lambda_{j-1}))_i \geq t'(d_i - \delta_d) + T\delta_d \text{ for some } 1 \leq t' \leq t\right)
\]

\[
\leq \mathbb{P}\left(\sum_{j=1}^{t'} [(b_j \tilde{x}_j(\lambda_{j-1}))_i - \mathbb{E}\left[(b_j \tilde{x}_j(\lambda^*(d_{j-1})))_i | \mathcal{H}_{j-1}\right]] \geq T\delta_d \text{ for some } 1 \leq t' \leq t\right)
\]

\[
\leq \mathbb{P}\left(\sum_{j=1}^{t'} [(b_j \tilde{x}_j(\lambda_{j-1}))_i - \mathbb{E}\left[(b_j \tilde{x}_j(\lambda_{j-1}))_i | \mathcal{H}_{j-1}\right]] + \sum_{j=1}^{t'} \mathbb{E}\left[(b_j \tilde{x}_t(\lambda^*(d_{j-1})))_i | \mathcal{H}_{t-1}\right] - \mathbb{E}\left[(b_j \tilde{x}_j(\lambda_{j-1}))_i | \mathcal{H}_{j-1}\right] \geq T\delta_d \text{ for some } 1 \leq t' \leq t\right)
\]

\[
\leq \mathbb{P}\left(\sum_{j=1}^{t'} [(b_j \tilde{x}_j(\lambda_{j-1}))_i - \mathbb{E}\left[(b_j \tilde{x}_j(\lambda_{j-1}))_i | \mathcal{H}_{j-1}\right]] \geq T\delta_d \frac{2}{\delta_d} \text{ for some } 1 \leq t' \leq t\right)
\]

\[
+ \mathbb{P}\left(\sum_{j=1}^{t'} \mathbb{E}\left[(b_j \tilde{x}_t(\lambda^*(d_{j-1})))_i | \mathcal{H}_{j-1}\right] - \mathbb{E}\left[(b_j \tilde{x}_j(\lambda_{j-1}))_i | \mathcal{H}_{j-1}\right] \geq T\delta_d \frac{2}{\delta_d} \text{ for some } 1 \leq t' \leq t\right)
\]

Since sequences in the last two lines are martingales/sub-martingales, we use Doob’s martingale inequality and get the following derivation:
\[ P(\tau_i \leq t) \leq \frac{4}{T^2 \delta_d^2} \sum_{j=1}^{t} \mathbb{E} \left[ (b_j \tilde{x}_j(\lambda_{j-1}))_i - \mathbb{E} \left[ (b_j \tilde{x}_j(\lambda_{j-1}))_i | \mathcal{H}_{t-1} \right] \right]^2 + \frac{4}{T^2 \delta_d^2} \mathbb{E} \left[ \sum_{j=1}^{t} \mathbb{E} \left[ (b_j \tilde{x}_t(\lambda^*(d_{j-1}))_i | \mathcal{H}_{j-1} \right] - \mathbb{E} \left[ (b_j \tilde{x}_j(\lambda_{j-1}))_i | \mathcal{H}_{j-1} \right] \right]^2 \]
\[ \leq \frac{16nb^2D^2t}{T^2 \delta_d^2} + \frac{8 \tilde{C}^2 \tilde{b}^4 t}{T^2 \delta_d^2} \sum_{j=1}^{t} \mathbb{E} \left\| \lambda_{j-1} - \lambda^*(d_{j-1}) \right\|_2^2 \quad \text{(by Assumption 3)} \]
\[ \leq \frac{16nb^2D^2t}{T^2 \delta_d^2} + \frac{16nb^2D^2C_2^t}{T^2 \delta_d^2} (\log t + \log T - \log(T - t + 1) + 2) \]

We now go back to calculate the \( \mathbb{E}(T - \tau_i) \):

\[
\mathbb{E}(T - \tau_i) \leq 1 + \frac{\sqrt{nD\tilde{b}}}{d} + \sum_{i=1}^{T} P(\tau_i \leq t) \leq \sum_{t=1}^{T} \frac{16nb^2D^2t}{T^2 \delta_d^2} + \frac{16nb^2D^2C_2t}{T^2 \delta_d^2} (\log t + \log T + 2) \quad (A.15)
\]

Putting together (A.14) and (A.15) we conclude the proof of lemma 6.

**A.10 Proof of Lemma 7**

For the \( \mathbb{E} \|\hat{\mu}_T - \mu^*\|_2^2 \), the optimality of \( \mu^* \) implies \( \hat{a}(\mu^*) = \mathbb{E}b_t \tilde{x}_t(\lambda^*) \), thus by conjugate we have

\[
\mathbb{E} \|\hat{\mu}_T - \mu^*\|_2^2 = \mathbb{E} \left\| \nabla r(\hat{a}(\mu^*)) - \nabla r\left( \sum_{t=1}^{T} \frac{b_t x_t}{T} \right) \right\|_2^2 \leq \mathbb{E} \left( \frac{1}{\varepsilon_r} \right)^2 \left\| \hat{a}(\mu^*) - \sum_{t=1}^{T} \frac{b_t x_t}{T} \right\|_2^2
\]

\[
= \left( \frac{1}{\varepsilon_r} \right)^2 \mathbb{E} \left\| \sum_{t=1}^{T} \frac{b_t x_t}{T} - \sum_{t=1}^{T} \frac{b_t \tilde{x}_t(\lambda^*)}{T} + \sum_{t=1}^{T} \frac{b_t \tilde{x}_t(\lambda^*)}{T} - \mathbb{E}b_t \tilde{x}_t(\lambda^*) \right\|_2^2 \leq \frac{3}{\varepsilon_r^2} \mathbb{E} \left\| \sum_{t=1}^{T} \frac{b_t \tilde{x}_t(\lambda_{t-1}) - b_t \tilde{x}_t(\lambda^*)}{T} \right\|_2^2 \quad \text{(part A.10 1)}
\]

\[
+ \frac{3}{\varepsilon_r^2} \mathbb{E} \left\| \sum_{t=1}^{T} \frac{b_t \tilde{x}_t(\lambda^*) - T\mathbb{E}b_t \tilde{x}_t(\lambda^*)}{T} \right\|_2^2 \quad \text{(part A.10 2)}
\]

\[
+ \frac{3}{\varepsilon_r^2} \mathbb{E} \left\| \sum_{t=T+1}^{T} \frac{b_t x_t - b_t \tilde{x}_t(\lambda^*)}{T} \right\|_2^2 \quad \text{(part A.10 3)}.\]
For the part [A.10] 1, applying Assumption [4] we can yield

\[
\mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T} b_t \tilde{x}_t(\lambda_{t-1}) - b_t \tilde{x}_t(\lambda^*) \right\|^2 \leq \mathbb{E} \left[ \frac{1}{T^2} \sum_{t=1}^{T} \| b_t \tilde{x}_t(\lambda_{t-1}) - b_t \tilde{x}_t(\lambda^*) \|^2 \right]
\]

\[
\leq \frac{\mathbb{E} \sum_{t=1}^{T} \| b_t \tilde{x}_t(\lambda_{t-1}) - b_t \tilde{x}_t(\lambda^*) \|^2}{T} = \frac{\mathbb{E} \sum_{t=1}^{T} \| b_t \tilde{x}_t(\lambda_{t-1}) - b_t \tilde{x}_t(\lambda^*) \|^2}{T} I(t \leq \tau)
\]

\[
= \frac{\sum_{t=1}^{T} \mathbb{E} \left[ \| b_t \tilde{x}_t(\lambda_{t-1}) - b_t \tilde{x}_t(\lambda^*) \|^2 I(t \leq \tau) \right]}{T}
\]

\[
\leq L_2 b^2 \frac{\sum_{t=1}^{T} \mathbb{E} \left[ \| \lambda_{t-1} - \lambda^* \|^2 I(t \leq \tau) \right]}{T}
\]

\[
\leq O \left( \frac{\log T}{T} \right).
\]

(a) is by Assumption [4] and the fact \( \{ t \leq \tau \} \subseteq \sigma(\mathcal{H}_{t-1}) \), and \( \tilde{x}_t(\cdot) \perp \lambda_{t-1} \). (b) is by Lemma 5.

For the part [A.10] 2, since \( b_t \tilde{x}_t(\lambda^*) \) is bounded by \( |(b_t \tilde{x}_t(\lambda^*))_i| \leq \sqrt{n}Db \), using Hoeffding’s inequality on each dimension we have

\[
P \left( \left\| \frac{1}{T} \sum_{t=1}^{T} b_t \tilde{x}_t(\lambda^*) - \mathbb{E} b_t \tilde{x}_t(\lambda^*) \right\| > \varepsilon \right) \leq m \exp \left( -\frac{2\varepsilon^2 T}{mnD^2b^2} \right).
\]

Thus

\[
\mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T} b_t \tilde{x}_t(\lambda^*) - \mathbb{E} b_t \tilde{x}_t(\lambda^*) \right\|^2 = \int_0^\infty \mathbb{P} \left( \left\| \frac{1}{T} \sum_{t=1}^{T} b_t \tilde{x}_t(\lambda^*) - \mathbb{E} b_t \tilde{x}_t(\lambda^*) \right\| > \varepsilon \right) d\varepsilon
\]

\[
\leq \int_0^\infty m \exp \left( -\frac{2\varepsilon T}{mnD^2b^2} \right) d\varepsilon \leq \frac{m^2nD^2b^2}{2T}.
\]

For the part [A.10] 3, since \( \| b_t x_t - \mathbb{E} b_t \tilde{x}_t(\lambda^*) \|^2 \leq nD^2b^2 \), by Lemma 6 we have

\[
\mathbb{E} \left\| \frac{1}{T} \sum_{t=\tau+1}^{T} b_t x_t - b_t \tilde{x}_t(\lambda^*) \right\|^2 \leq \mathbb{E} \left\| (T-\tau) \sum_{t=\tau+1}^{T} (b_t x_t - b_t \tilde{x}_t(\lambda^*)) \right\|^2 \leq nD^2b^2 \mathbb{E} \frac{(T-\tau)}{T} \leq O \left( \frac{\log T}{T} \right)
\]

We then go back to control the next term \( \mathbb{E} \left\| \sum_{t=1}^{\tau} (a_t - b_t x_t) \right\|^2 \).

\[
\mathbb{E} \left\| \sum_{t=1}^{\tau} (a_t - b_t x_t) \right\|^2 \leq 2\mathbb{E} \left\| \sum_{t=1}^{\tau} (a_t - \tilde{a}(\mu^*) + \mathbb{E} b_t \tilde{x}_t(\lambda^*) - b_t x_t) \right\|^2
\]

\[
\leq 2\mathbb{E} \left\| \sum_{t=1}^{\tau} (a_t - \tilde{a}(\mu^*)) \right\|^2 + 2\mathbb{E} \left\| \sum_{t=1}^{\tau} (\mathbb{E} b_t \tilde{x}_t(\lambda^*) - b_t x_t) \right\|^2
\]

\[
\leq 2\mathbb{E} \tau \sum_{t=1}^{\tau} \| \mu_t - \mu^* \|^2 + 2T^2 \mathbb{E} \left\| \frac{1}{T^2} \sum_{t=1}^{T} (\mathbb{E} b_t \tilde{x}_t(\lambda^*) - b_t x_t) \right\|^2
\]

From the argument above, we show that the first term is controlled by \( O(T \log T) \), and the second term can also be controlled by \( O(T \log T) \) (this proof follows previous derivation of part [A.10] 1-3). Thus we finish the proof.
A.11 Proof of Theorem 4 and 5

We specify a non-regularized case where \( f_t(x) = -\frac{1}{2}(x - 2\xi_t)^2 + \xi_t^2 \), with fixed cost \( b_t = 1 \), average resource capacity \( d = \frac{1}{2}D \), and \( \xi_t \) following two-point distribution \( \mathbb{P} (\xi_t = \frac{1}{2}D) = \mathbb{P} (\xi_t = \frac{3}{4}D) = \frac{1}{2} \). Then the dual problem is

\[
D_t(\lambda) = \begin{cases} 
\frac{1}{2}D\lambda & \text{if } \lambda > \xi_t \\
-\frac{1}{4}D + \xi_t - \frac{1}{2}D\lambda & \text{if } \lambda < \xi_t - \frac{1}{2}D \\
\lambda^2 - 2(\xi_t - \frac{1}{4}D)\lambda + \xi_t^2 & \text{if } \xi_t - \frac{1}{2}D \leq \lambda \leq \xi_t.
\end{cases}
\]

Suppose \( \lambda^* \) is the optimal solution to the deterministic problem \( \min_{\lambda \geq 0} D(\lambda) = \mathbb{E}D_t(\lambda) \). Without loss of generality, we assume that our dual variable \( \lambda \) is taken within \( \left[ \frac{1}{4}D, \frac{1}{2}D \right] \) since we know that \( \lambda^* = \mathbb{E}\xi_t - \frac{1}{4}D = \frac{3}{8}D \in \left[ \frac{1}{4}D, \frac{1}{2}D \right] \).

\[
D_t(\lambda) = f_t^*(\lambda) + d^T \lambda = \lambda^2 - 2(\xi_t - \frac{1}{4}D)\lambda + \xi_t^2.
\]

For the dual-based police \( \{\lambda_t\}_{t=0}^{T-1} \), the corresponding primal variable is \( x_t = \tilde{x}_t(\lambda_{t-1}) = 2\xi_t - 2\lambda_{t-1} \) or void if the resource is depleted. We have the following regret:

\[
\text{Regret}(A) = R^*(\mathcal{P}) - R(A|\mathcal{P})
= \mathbb{E} \left[ \max_{x_t \in [0,D]} \left\{ \sum_{t=1}^{T} f_t(x_t) \text{ s.t. } \sum_{t=1}^{T} x_t \leq \frac{1}{2}DT \right\} \right] - \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right]
= \mathbb{E} \left[ \min_{\lambda \geq 0} \left\{ \sum_{t=1}^{T} D_t(\lambda) \right\} \right] - \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right]
= \mathbb{E} \left[ \sum_{t=1}^{T} D_t(\lambda_t^*) \right] - \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right]
\]

Define the corresponding

\[
g(\lambda) = \mathbb{E} [f_t(\tilde{x}_t(\lambda)) + \langle d - b_t\tilde{x}_t(\lambda), \lambda^* \rangle] = D(\lambda) - \langle \nabla D(\lambda), \lambda - \lambda^* \rangle
\]

We have \( g(\lambda^*) = D(\lambda^*) \) and \( g(\lambda^*) - g(\lambda) = (\lambda^* - \lambda)^2 \). For the quadratic function \( D_t \), we always have \( D_t(\lambda_1) - D_t(\lambda_2) = \nabla D_t(\lambda_2)(\lambda_1 - \lambda_2) + (\lambda_1 - \lambda_2)^2 \). Thus it follows that

\[
\text{Regret}(A) = \mathbb{E} \left[ \sum_{t=1}^{T} D_t(\lambda_t^*) \right] - TD(\lambda^*) + TD(\lambda^*) - \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right]
= \mathbb{E} \left[ \sum_{t=1}^{T} D_t(\lambda_t^*) - D_t(\lambda^*) \right] + Tg(\lambda^*) - \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right]
= -\mathbb{E} \left[ \sum_{t=1}^{T} [\nabla D_t(\lambda_t^*)(\lambda^* - \lambda_t^*)] + T(\lambda^* - \lambda_t^*)^2 \right] + Tg(\lambda^*) - \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right]
= -T\mathbb{E}(\lambda^* - \lambda_t^*)^2 + Tg(\lambda^*) - \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right].
\]

By the dual convergence in Theorem 1, we know that the first term \( T\mathbb{E}(\lambda^* - \lambda_t^*)^2 \) can be bounded by a constant. Now we handle the second term by controlling the stopping time. Define the stopping time
\( \tau_0 = \min \{ t \in [T] | \sum_{t=1}^{t} x_t \geq \frac{1}{2} DT - D \} \cup \{ T \} \). Then when \( t \leq \tau_0 \), we always have \( x_t = \tilde{x}_t(\lambda_{t-1}) = 2\xi_t - 2\lambda_{t-1} \), and \( 0 \leq \sum_{t=\tau_0+1}^{T} x_t \leq D \) for \( t > \tau \). Then we have

\[
\mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right] \leq \mathbb{E} \left[ \sum_{t=1}^{\tau_0} f_t(\tilde{x}_t(\lambda_{t-1})) + \left( \frac{1}{2} D - \tilde{x}_t(\lambda_{t-1}), \lambda^* \right) \right] + \mathbb{E} \left[ \sum_{t=\tau_0+1}^{T} f_t(x_t) + \left( \frac{1}{2} D - x_t(\lambda), \lambda^* \right) \right]
\]

\[
\leq \mathbb{E} \sum_{t=1}^{\tau} g(\lambda_{t-1}) + \mathbb{E} \left[ \sum_{t=\tau_0+1}^{T} \frac{3}{4} D x_t + \frac{1}{2} D \lambda^* \right]
\]

\[
\leq \mathbb{E} \sum_{t=1}^{\tau} g(\lambda_{t-1}) + \frac{3}{16} D^2 \mathbb{E}[T - \tau_0] + \frac{3}{4} D^2.
\]

The first inequality is because of the resource constraint, and the second one is because \( f_t(x) \leq f_t'(0)(x - 0) \leq \frac{3}{4} D x \). If we specify \( \lambda_{t-1} = \xi_t \) when the resource constraints are violated, we also have \( \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right] \leq \mathbb{E} \sum_{t=1}^{T} g(\lambda_{t-1}) \). Then

\[
Tg(\lambda^*) - \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right] \geq \mathbb{E} \left[ \sum_{t=1}^{\tau_0} g(\lambda^*) - g(\lambda_{t-1}) \right] + \mathbb{E}(g(\lambda^*) - \frac{3}{16} D^2) \mathbb{E}[T - \tau_0] - \frac{3}{4} D^2
\]

\[
= \mathbb{E} \left[ \sum_{t=1}^{\tau_0} (\lambda^* - \lambda_{t-1})^2 \right] + \frac{5}{64} D^2 \mathbb{E}[T - \tau_0] - \frac{3}{4} D^2,
\]

or \( Tg(\lambda^*) - \mathbb{E} \left[ \sum_{t=1}^{T} f_t(x_t) \right] \geq \mathbb{E} \left[ \sum_{t=1}^{T} (\lambda^* - \lambda_{t-1})^2 \right] \). Applying van Trees inequality to the estimation of \( \lambda^* \), we can prove the Theorem 4. To prove the Theorem 5, we only need to show the stopping time \( \mathbb{E}[T - \tau_0] \geq \Omega(\sqrt{T}) \) given the convergence condition. This proof is inspired by Arlotto and Gurvich [2019]. Denote \( t' = \lfloor T - \sqrt{T} \rfloor \). We show that \( \mathbb{P}(\tau_0 \leq t') \) is larger that a constant \( c \) so that \( \mathbb{E}\tau_0 \leq (1 - c)T + c(T - \sqrt{T}) \leq T - c\sqrt{T} \).

\[
\mathbb{P}(\tau_0 \leq t') = \mathbb{P} \left( \sum_{t=1}^{t'} 2(\xi_t - \lambda_{t-1}) \geq \frac{DT}{2} - D \right)
\]

\[
\geq \mathbb{P} \left( \left\{ \sum_{t=1}^{t'} 2(\xi_t - \lambda^*) \geq \frac{DT}{2} - D + \varepsilon D \sqrt{t'} \right\} \cap \left\{ \sum_{t=1}^{t'} |\lambda_{t-1} - \lambda^*| < \varepsilon D \sqrt{t'} \right\} \right)
\]

\[
\geq \mathbb{P} \left( \left\{ \sum_{t=1}^{t'} 2(\xi_t - \lambda^*) \geq \frac{DT}{2} - D + \varepsilon D \sqrt{t'} \right\} \right) - \mathbb{P} \left( \left\{ \sum_{t=1}^{t'} |\lambda_{t-1} - \lambda^*| \geq \varepsilon D \sqrt{t'} \right\} \right)
\]

With the condition \( \mathbb{E}|\lambda_t - \lambda^*| \leq c_2 D / \sqrt{T+1} \), we have \( \mathbb{P} \left( \sum_{t=1}^{t'} |\lambda_{t-1} - \lambda^*| \geq \varepsilon D \sqrt{t'} \right) \leq \frac{2c_2}{\varepsilon} \) by Chebyshev’s inequality. Then it holds that

\[
\mathbb{P}(\tau_0 \leq t') \geq \mathbb{P} \left( \left\{ \sum_{t=1}^{t'} 2(\xi_t - \lambda^*) \geq \frac{DT}{2} - D + \varepsilon D \sqrt{t'} \right\} \right) - \frac{2c_2}{\varepsilon}
\]

\[
= \mathbb{P} \left( \left\{ \sum_{t=1}^{t'} \frac{4}{D} (\xi_t - D) \geq \frac{t'}{2} + (T - t') - 2 + 2\varepsilon \sqrt{t'} \right\} \right) - \frac{2c_2}{\varepsilon}
\]

\[
\geq \mathbb{P} \left( \left\{ \sum_{t=1}^{t'} \frac{4}{D} (\xi_t - D) \geq \frac{t'}{2} + (1 + 2\varepsilon) \sqrt{t'} \right\} \right) - \frac{2c_2}{\varepsilon},
\]
where \( \sum_{t=1}^{t'} \frac{4}{D} (\xi_t - \frac{D}{2}) \) follows the binomial distribution \( B(t', \frac{1}{2}) \), with mean \( \mu = \frac{t'}{2} \) and standard deviation \( \sigma = \frac{\sqrt{t'}}{2} \). The second inequality is because \( T - t' \leq \sqrt{T} + 1 \) and \( \sqrt{T} - \sqrt{t'} \leq \sqrt{T} - \sqrt{T - \sqrt{T}} = \frac{\sqrt{T}}{\sqrt{T} - \sqrt{T} + \sqrt{T}} \leq 1 \). For the binomial distribution, \( \mathbb{P}(X \geq \mu + x\sigma) \) converge to \( \Phi(-x) \) for any \( x \) with known \( O(\frac{1}{\sqrt{n}}) \) speed by Berry-Esseen CLT where \( \Phi(x) \) is the distribution function of standard normal distribution. We let \( c_2 = \sup_{\varepsilon > 0} \varepsilon \Phi(-2 - 4\varepsilon)/4 \). Then there exists \( \varepsilon_0 > 0 \) such that when \( T \) is large enough, \( \mathbb{P} \left( \left\{ \sum_{t=1}^{t'} \frac{4}{D} (\xi_t - \frac{D}{2}) \geq \frac{t'}{2} + (1 + 2\varepsilon_0)\sqrt{t'} \right\} \right) \geq \frac{3c_2}{\varepsilon_0} \), which indicates that \( \mathbb{P}(\tau_0 \leq t') \geq \frac{c_2}{\varepsilon_0} \). This makes our proof complete.