Submitted to *Operations Research*
manuscript OPRE-2022-09-486

Optimal Regularized Online Allocation by Adaptive Re-Solving

Wanteng Ma
Department of Mathematics, HKUST, wmaah@connect.ust.hk

Ying Cao
Department of Electronic and Computer Engineering, HKUST, ycaoan@connect.ust.hk

Danny H.K. Tsang
Department of Electronic and Computer Engineering, HKUST, eetsang@ust.hk

Dong Xia
Department of Mathematics, HKUST, madxia@ust.hk

This paper introduces a dual-based algorithm framework for solving the regularized online resource allocation problems, which have potentially non-concave cumulative rewards, hard resource constraints, and a non-separable regularizer. Under a strategy of adaptively updating the resource constraints, the proposed framework only requests approximate solutions to the empirical dual problems up to a certain accuracy and yet delivers an optimal logarithmic regret under a locally second-order growth condition. Surprisingly, a delicate analysis of the dual objective function enables us to eliminate the notorious log-log factor in regret bound. The flexible framework renders renowned and computationally fast algorithms immediately applicable, e.g., dual stochastic gradient descent. Additionally, an infrequent re-solving scheme is proposed, which significantly reduces computational demands without compromising the optimal regret performance. 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 demonstrate the merits of the proposed algorithm framework.

Key words: online allocation problems, regret, re-solving strategy

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 attention 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. 2021b, Celli et al. 2022). Compared with non-regularized online resource allocation that maximizes an additively separable objective, non-separable regularization poses new challenges to algorithm design and regret analysis.

In this paper, we study regularized online allocation problem with a, potentially non-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 an irrevocable decision based on the 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. 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 consumption, including the 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 a 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 latter work also proposed adaptive policies that achieve the logarithmic-order regrets up to an additional $\log\log$ 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. 2021b) 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 log-log 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 provide two parallel approaches to derive the convergence rate of the empirical dual solution to its population counterpart in the case of additive non-linear rewards function and non-separable regularizer. By either localized Rademacher complexity or a partition argument, we show that the dual 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 (Li and Ye 2021). The improvement is made possible by a delicate analysis of the local behavior of the empirical dual program near the optimal solution. Different from Balseiro et al. (2021b), the proposed new dual form can depict the impact of constraint update, which is the key to our algorithm design. Our analysis also 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 the approximation accuracy 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. Notably, as a stochastic optimization problem, the derived dual convergence sheds new light on the widely studied Sample Average Approximation (SAA) and Empirical Risk Minimization (ERM) problems and may be of independent interest.

Adaptive algorithm framework with re-solving. 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 featuring re-solving, each iteration mainly consists of two routines: primal decision-making and periodical dual optimization. At a high level, our adaptive algorithm framework generalizes the history-dependent policy in online linear programming (Li and Ye 2021), which evolves from the budget-ratio policy (Arlotto and Gurvich 2019, Balseiro and Gur 2019) and the *re-solving* heuristic in network revenue management (Jasin and Kumar 2012, Wu et al. 2015). There are two key ingredients in the dual optimization of our algorithm framework. First, for each optimization problem, we adaptively update the average remaining resources in the dual problem. Besides fulfilling the resource constraints, this adaptive resource control plays an essential role in achieving a $O(\log T)$ regret rather than the $O(T^{1/2})$ one attained by Balseiro et al. (2021b). Second, our algorithm framework only requires an approximate solution to dual optimization, up to a certain accuracy. This allows a flexible choice of computationally efficient algorithms for dual optimization, be they deterministic or stochastic. Paired with first-order methods, our algorithm enjoys an acceptable polynomial-time cost comparable to prior algorithms. Moreover, we also develop the *infrequent* resolving technique that only requires solving dual optimizations for $O(\log T)$ times while achieving optimal regret and a fast adaptive dual gradient method with *linear* computation costs but a sub-optimal $O(\log^2 T)$ regret. 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 regret is characterized by dual convergence, the aforementioned new result 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 (Hazan et al. 2007) and multi-secretary problem (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 linear programming (Li and Ye 2021) contains an additional $\log\log T$ factor. Distinct from previous work, our non-linear and non-separable primal problem introduces greater challenges when analyzing dual behaviors. By comparing with existing algorithms, we clarify the critical role played by the adaptive resource update 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 on the applications of our method and theory to online linear programming, online max-min fairness allocation and online load balancing, etc. Simulation results will 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.

Apart from AdWords, two major topics related to online linear allocation are online revenue management problems and online multi-secretary problems. In online revenue management, a decision maker aims to find a dynamic pricing policy that maximizes a company's linear total rewards. Among all the strategies for solving online revenue management problems, re-solving stands out for its excellent performance. By combining the re-solving strategy and a trigger-and-threshold mechanism, Reiman and Wang (2008) reduced the regret from previously studied $O(T^{1/2})$ (Cooper 2002) 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 discrete distribution with finite support 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 in 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. Recent advances in contextual linear optimization (Hu

et al. 2022) have demonstrated that this lower bound can be surpassed when instances exhibit more favorable properties, such as finiteness of the hypothesis class and higher-order smoothness, which are not applicable to our problem.

Other independent works of online linear programming also contribute greatly to the understanding of online allocation problems. Agrawal et al. (2014) proposed a dual-based algorithm that dynamically updates dual variables and periodically solves linear programs which 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 frequent resolving 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 exactly solving an offline linear program of growing sizes with high frequency, which may be computationally intractable for large T. An $\Omega(\log T)$ regret lower bound was established, which is consistent with Bray (2019).

Another noteworthy recent development in linear allocation, when the item distribution is known, is the study of greedy policies. Research by Kerimov et al. (2021) has shown that greedy policies can achieve near-optimality in two-way dynamic matching, while Gupta (2022) demonstrated that these policies can produce bounded regret for multi-way matching. However, both of these studies rely on a general position condition for the deterministic approximate linear program, which is not required in our analysis. Furthermore, these greedy policies necessitate knowledge of the optimal solution for the deterministic equivalent, which is infeasible in our scenarios. Importantly, the aforementioned works focus only on limited distributions with discrete support, while our research extends this scope to encompass continuous distributions. In addition, Balseiro et al. (2023) introduced a unified framework known as dynamic resource-constrained reward collection and summarized a group of local notions of smoothness and strong convexity, which can be viewed as general cases for our required assumptions.

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. Recently, partly due to its computational efficiency, dual mirror descent has been 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 proportional to time horizon T. They proposed a

dual-based mirror descent algorithm acting on dual space that achieves $O(T^{1/2})$ regret, which was said to be unimprovable under their assumptions. Dual mirror descent presents a self-correcting mechanism, which naturally prevents resources from depleting too fast. The problem we study in this paper is closer to Balseiro et al. (2021b), 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 they define a new separable dual problem and update dual variables using regularized subgradient descent. 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, it is sub-optimal when the reward functions possess more favorable conditions like strong convexity.

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 has 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 (McMahan 2011, 2017). 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 $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. 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 $\{d_t\}_{t=1}^T$, 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 ℓ_2-norm and ℓ_∞-norm, respectively. We write $\tilde{O}(\cdot)$ as the big-O notation omitting the logarithmic factor.

2. Regularized Online Allocation Problem

We describe the regularized online allocation problem with finite time period T as follows:

$$\begin{aligned} \max_{\{x_t, t\in[T]\}} \quad & \sum_{t=1}^T f_t(x_t) + T \cdot r\left(\frac{\sum_{t=1}^T b_t x_t}{T}\right) \\ \text{s.t.} \quad & \sum_{t=1}^T b_t x_t \preceq dT,\ d \in \mathbb{R}_+^m \\ & x_t \in \mathcal{X}, \forall t \in [T]. \end{aligned} \tag{2.1}$$

where $f_t : \mathbb{R}^n \to \mathbb{R}$ is the reward function which may be potentially non-concave, $r : \mathbb{R}^m \to \mathbb{R}$ 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 potentially non-convex with void action $0 \in \mathcal{X}$.

Following the online sequential learning setting, we assume that at each time $1 \le t \le T$, 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 $\mathcal{H}_{t-1} := \{f_j, b_j, x_j\}_{j=1}^{t-1}$:

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

by taking the total resource constraints $\sum_{j=1}^t b_j x_j \preceq dT$ 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 $\mathcal{P}$ as

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

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

$$R^*(\mathcal{P}) := \mathbb{E}_{\mathcal{P}}\left[\max_{x_t\in\mathcal{X}} \sum_{t=1}^T f_t(x_t) + T \cdot r(\frac{\sum_{t=1}^T b_t x_t}{T}),\ s.t.\ \sum_{t=1}^T b_t x_t \preceq dT\right], \tag{2.3}$$

which serves as the benchmark performance. For a given $\mathcal{P}$, define the *regret* as $\text{Regret}\left(A|\mathcal{P}\right) := R^*(\mathcal{P}) - R\left(A|\mathcal{P}\right)$. We then define the *worst-case regret* of an algorithm A as the worst difference between the expected online reward and offline optimum over all the possible distributions in a certain probability family Ξ:

$$\text{Regret}(A) := \sup_{\mathcal{P}\in\Xi}\left\{R^*(\mathcal{P}) - R\left(A|\mathcal{P}\right)\right\}, \tag{2.4}$$

where the distribution family Ξ will be identified later.

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. 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

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

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 $\mathcal{Z}$ with $b\circ\mathcal{X} := \mathsf{span}\{b\cdot x|$ for all possible b and $x\in\mathcal{X}\}\subseteq\mathcal{Z}$. Define the conjugate function

$$f_t^*(\lambda) := \max_{x\in\mathcal{X}} \left\{ f_t(x) - x^\top\lambda \right\}, \quad r^*(\mu) := \max_{a\in\mathcal{Z}} \left\{ r(a) - a^\top\mu \right\}. \tag{2.6}$$

Then, the dual problem of 2.1 can be written as an additive form:

$$\min_{\mu,\lambda\succeq 0} \qquad \bar{D}_T(\lambda,\mu,d) := \frac{1}{T}\sum_{t=1}^{T} f_t^*(b_t^\top(\mu+\lambda)) + r^*(-\mu) + d^\top\lambda \tag{2.7}$$

Note that, we use two dual variables to pose our problem with μ for the impact of variable separation and λ for the constraint. In contrast to Balseiro et al. (2021b), two dual variables enable us to capture the influence of both variable separation and variation of constraints d, wherein the latter is crucial for our scheme to be optimal.

Under our stochastic input assumption, (2.7) can be viewed as a sample average approximation (SAA) (Shapiro et al. 2009) of the following stochastic program (or *fluid* problem):

$$\min_{\mu,\lambda\succeq 0} \qquad D(\lambda,\mu,d) := \mathbb{E} f_t^*(b_t^\top(\mu+\lambda)) + r^*(-\mu) + d^\top\lambda \tag{2.8}$$

In the following discussion, we will sometimes write $\nu = \lambda + \mu$ and write the dual variable uniformly as $\boldsymbol{\lambda} := [\nu^\top, \mu^\top]^\top$ in shorthand with substitution. If we have known the exact offline solution to (2.7), denoted by $\boldsymbol{\lambda}_T^* := [\nu_T^{*\top}, \mu_T^{*\top}]^\top$, then by choosing the corresponding primal variables we can optimize the primal problem (2.1). However, in the online setting, it is impossible to find such an 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.,

$$\min_{\mu,\lambda\succeq 0} \qquad \bar{D}_t(\lambda,\mu,d) := \frac{1}{t}\sum_{j=1}^{t} f_j^*(b_j^\top(\mu+\lambda)) + r^*(-\mu) + d^\top\lambda \tag{2.9}$$

and then use the dual approximate solution $\boldsymbol{\lambda}_t^*$ to decide the following several primal solutions. Such a re-solving idea has shown its merit in controlling regret both in theory and in practice (Jasin 2015, Ferreira et al. 2018, Li and Ye 2021). Hence we expect that this idea also works in regularized 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 $\boldsymbol{\lambda}_T^*$ for large T ? While $\boldsymbol{\lambda}_T^*$ is random, from the stochastic programming perspective, as T goes large, the optimal solution to the SAA (2.7), $\boldsymbol{\lambda}_T^*$, will converge in probability to the solution to its stochastic program (2.8), denoted by $\boldsymbol{\lambda}^*$. If we want to establish the theory of dual-based algorithms that rely on the approximation of $\boldsymbol{\lambda}_T^*$, we need to first explore the convergence behavior of $\boldsymbol{\lambda}_T^*$ toward $\boldsymbol{\lambda}^*$.
2. How will the dual approximate solutions affect reward and, consequently, regret? This question is the key to the algorithm design. For online allocation problems, a good approximation of $\boldsymbol{\lambda}^*$ or $\boldsymbol{\lambda}_T^*$ 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 explain the influence of dual approximation 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 computational costs while maintaining the optimal order of regret. The influence of our approximation scheme on the regret is, in general, negligible when compared with the exact optimal solutions.

We propose the following assumptions that suffice our algorithm to achieve logarithmic regret.

2.2. Assumptions

ASSUMPTION 1 **(Boundedness 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 an unknown distribution* $\mathcal{P}$.

1.2 f_t *is defined in the closed decision region* $\mathcal{X} \subseteq \mathbb{R}_+^n$ *with* $\|x\|_\infty \le D$ *for any* $x \in \mathcal{X}$.

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

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

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

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. Here we assume that the average resource constraints 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 the basis for the subsequent discussion of regret, especially for bounding the stopping time.

Under Assumption 1, one general feasible region of our regularizer $r(a)$ is $\mathcal{Z} := \left\{a \middle| \|a\|_2 \le \sqrt{n}D\bar{b}\right\}$, which satisfies $b \circ \mathcal{X} \subseteq \mathcal{Z}$. We then describe the necessary assumptions on the regularizer r.

ASSUMPTION 2 **(Assumptions on the regularizer)**. *The concave regularizer r satisfies: r is concave, closed, and bounded in $\mathcal{Z}$: $|r| \le \bar{r}$ with bounded (sub)gradient $\|\nabla r(a)\|_\infty \le G$ for any $a \in \mathcal{Z}$.*

The feasible region $\mathcal{Z}$ here can also be chosen in other shape as long as $b \circ \mathcal{X} \subseteq \mathcal{Z}$. Together with Assumption 1, we can show that both the population-version and sample-version optimal solutions, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\lambda}_T^*$, respectively, are uniformly bounded.

LEMMA 1. *Under Assumption 1, 2, the optimal solutions to problem* (2.7) *and* (2.8) *are bounded by:*

$$\begin{gathered} \|\lambda_T^*\|_\infty \le \frac{2(\bar{f}+\bar{r})}{\underline{d}}, \|\lambda^*\|_\infty \le \frac{2(\bar{f}+\bar{r})}{\underline{d}} \\ \|\mu_T^*\|_\infty \le G, \|\mu^*\|_\infty \le G \end{gathered} \tag{2.10}$$

By Lemma 1, we define the regions that contain all the possible optimal dual variables as $\Omega_\lambda := \left\{\lambda \middle| \|\lambda\|_\infty \le \frac{2(\bar{f}+\bar{r})}{\underline{d}}\right\}$, and $\Omega_\mu := \{\mu| \ \|\mu\|_\infty \le G\}$. These regions will be the feasible sets of our dual variables since we do not want them to move far from the optimal solution $\boldsymbol{\lambda}^*$. Assumption 2 can be easily achieved by many popular regularizers enumerated below.

1. ℓ_1**-loss**: $r(a) := -\kappa \|a\|_1$. This regularizer serves as a tool to achieve a sparse resource allocation.
2. **Max-min loss**: $r(a) := \kappa \min_i(a_i/d_i)$. The max-min fairness regularizer allows us to maximize the minimum resource consumption. Resources under 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), Bertsekas and Gallager (2021).
3. **Negative max loss**: $r(a) := -\kappa \max_i(a_i/d_i)$. This regularizer 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). The load-balancing regularizer is widely used in, e.g., network design and cloud computing (Bejerano et al. 2004, Radunovic and Le Boudec 2007).
4. **Entropy loss**: $r(a) := -\kappa \left[\sum_{i=1}^m (a_i+\delta)\log(a_i+\delta) + (1-m\delta-\sum_{i=1}^m a_i)\log(1-m\delta-\sum_{i=1}^m a_i)\right]$ with the corresponding feasible region: $\mathcal{Z} := \left\{a \in \mathbb{R}_+^m \mid \sum_{i=1}^m a_i \le 1-m\delta\right\}$. We use this entropy loss when our problem is related to stochastic strategies and probabilistic assignment, e.g., randomly assigning impressions to advertisers type i with selected probabilities $a_i+\delta$ in online advertising. Here $1-m\delta-\sum_{i=1}^m a_i$ means the probability of no-assigning, and δ is the threshold of minimum assigning. 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. **No regularizer**: $r(a) := 0$. In this case, our problem reduces to the non-regularized online convex allocation problem. Therefore, the theory developed in this paper is immediately applicable to non-regularized cases.

In addition, to achieve optimality, we need the following regularity and non-degeneracy assumption of f_t^*:

ASSUMPTION 3 **(Regularity conditions on the dual problem)**. *We assume that our problem is locally second-order growth and well-conditioned: suppose* (λ^*, μ^*) *is the optimal solution to the problem* (2.8) *when* $d \in \Omega_d$. *Define* ∇f_t^* *as any (sub)gradient of* f_t^*. *Then for any* $d \in \Omega_d$,

3.1 (locally second order growth) Let $\nu := \lambda + \mu$ *and* $\nu^* := \lambda^* + \mu^*$. *The conjugate function* f_t^* *is continuous and satisfies*

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

for any $\lambda \in \Omega_\lambda$, $\mu \in \Omega_\mu$ *and constants* $\underline{\mathcal{L}}_f > 0$, *conditioned on* b_t.

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

Assumption 3.1 requires the expected conjugate of reward function to exhibit a local quadratic growth, conditioning on any given b_t. Assumption 3.1 controls the growth rate of dual function so that it will not degenerate to a line, which is necessary for characterizing dual solutions. Assumption 3.2 is easily satisfied since, oftentimes, the constraints are linearly independent. Note that $-\nabla f_t^*(b_t^\top \nu) = \arg\max_{x \in \mathcal{X}} \{f_t(x) - (\nu)^\top b_t x\}$ represents the primal solution given dual variable ν. Its randomness stems from the stochastic reward f_t. From this perspective, Assumption 3.1 only concerns the effect of dual variables on their corresponding *expected* primal solutions. However, when it comes to the smoothness, 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 moments. The following assumption serves this purpose. Equivalently, it depicts the variation behavior of the random award function f_t. This moment assumption establishes the connection between dual variables and primal performances. Note that Assumption 3 only concerns the deterministic problem (2.8), but the empirical problem does not necessarily share these local properties or is even not differentiable.

ASSUMPTION 4 **(Smoothness of moment)**. *Let* $\nu := \lambda + \mu$ *and* $\nu^* := \lambda^* + \mu^*$ *when we choose* $d \in \Omega_d$ *in* (2.8). *For any random variable* $V \in \mathbb{R}^m$ *that satisfies* $\left\| b_t^\top (V - \nu^*) \right\|_2 \leq \left\| b_t^\top (\nu - \nu^*) \right\|_2$ *a.s., the* 1*-th order moment of the (sub)gradient* ∇f_t^* *satisfies the following smoothness*

$$\mathbb{E}\left[\left\| \nabla f_t^*(b_t^\top V) - \nabla f_t^*(b_t^\top \nu^*) \right\|_2 \middle| b_t\right] \leq L_1 \left\| b_t^\top (\nu - \nu^*) \right\|_2.$$

for any $d \in \Omega_d$, $\lambda \in \Omega_\lambda$, $\mu \in \Omega_\mu$ *and given* b_t, *where* $L_1 > 0$ *is a constant.*

Assumption 4 requires the variation of reward function given b_t: $f_t \sim \mathcal{P}|b_t$ to be mild so that the primal solution changes smoothly. This doesn't mean that f_t^* must be globally smooth. The expectation here is with respect to f_t, $\boldsymbol{\nu}$. A similar description of smoothness can be found in Gorbunov et al. (2020). 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. We note that Assumptions 2-4 assume the corresponding conditions hold for all the $d \in \Omega_d$.

REMARK 1. Here we list several sufficient conditions, any of which can lead to both Assumptions 3 and 4:

1. f_t is linear with reward $v_t \in \mathbb{R}^m$, i.e., $f_t = v_t^\top x_t$. and for each $i \in [m]$, v_{it} is with distribution $\left|\mathbb{P}(v_{it} > b_{it}^\top \nu | b_t) - \mathbb{P}(v_{it} > b_{it}^\top \nu^* | b_t)\right| = \Theta(\left|b_{it}^\top(\nu - \nu^*)\right|)$.
2. f_t is drawn from a finite distribution wherein every possible f_t is locally smooth and strongly concave.
3. every f_t is concave and continuous with first order growth gradient, and $\mathbb{E}f_t(x)$ admits a lower upward quadratic (LUQ) envelope in a local region near the optimal solution (Balseiro et al. 2021a).

We will revisit the linear case in Section 7 for a detailed discussion. For more possible sufficient conditions, we refer the reader to Kakade et al. (2009), Bubeck et al. (2015), Balseiro et al. (2021a), etc.

Define the primal variable given (λ^*, μ^*) as $\tilde{x}_t(\nu^*) := -\nabla f_t^*(b_t^\top(\nu^*))$. In this sequel, all the dimensions that satisfy $d_i - \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)_i = 0$ with respect to the original d in (2.1) are referred to as *binding dimensions*. Denote $I_{\mathsf{B}} = \{i | d_i - \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)_i = 0\}$ the collection of binding dimensions. Similarly, *non-binding dimensions* are written as $I_{\mathsf{NB}} = \{i | d_i - \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)_i > 0\}$. Here for ease of notation, we omit the dependence of I_{B} and I_{NB} on the resource constraint d.

ASSUMPTION 5 **(Non-degeneracy)**. *Let (λ^*, μ^*) be the optimal dual solution given the original d in (2.1). Denote $d^* = \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)$. Then:*

5.1 The optimal solution (λ^*, μ^*) satisfies $\lambda_i^* = 0$ if and only if $i \in I_{NB}$, i.e., $d_i - d_i^* > 0$.

5.2 Further, if $i \in I_{NB}$, then there exists an small constant δ_0 such that the partial gradient $|\nabla_i r(a) - \nabla_i r(d^*)| \le \bar{\mathcal{L}}_r \left\|a - d^*\right\|_2$ for any $\left\|a - d^*\right\|_2 \le \delta_0$.

Assumption 5 states the non-degeneracy condition for dual problems with nonlinear objectives and the regularizer, which is generalized from the non-degeneracy condition of linear programs (Jasin and Kumar 2012, Jasin 2015, Wu et al. 2015, Li and Ye 2021). Assumption 5.1 imposes strong complementary slackness on the resource constraints $d \in \Omega_d$ uniformly. This suggests that when d changes within a certain region of Ω_d, the binding or non-binding dimensions of resource

constraints for the optimal solution will not change. This brings convenience for analyzing adaptive algorithms with frequently updated constraints. Assumption 5.1 ensures that binding and non-binding dimensions can be uniquely determined by the dual solution $\boldsymbol{\lambda}^*$. In our study, Assumption 5.1 is indispensable for the regret analysis because it allows the gap between the fluid benchmark and offline maximum to be well controlled. Assumption 5.2 requires the dual optimal solution $\mu_i^* = -\nabla_i r(d^*)$ to be non-degenerate in the non-binding dimension $i \in I_{\mathsf{NB}}$: it is unique and smooth with the change of resources in a tiny region near d^*. This can be achieved by the aforementioned regularizers as long as d^* is in a good position, e.g., for the max-min loss, we only require d^* to have a unique minimum dimension.

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 $\boldsymbol{\lambda}_T^*$ to the sample average approximation (SSA) in eq. (2.7). The Law of Large Numbers dictates that $\boldsymbol{\lambda}_T^*$ converges to $\boldsymbol{\lambda}^*$ in probability as $T \to \infty$. While the asymptotic behaviors of optimal solutions to SAA 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 Assumptions 1-4, for the regularized online problem (2.1). We mainly study the convergence of $\|\nu_T^* - \nu^*\|_2$ since the primal solution is only related to ν, not individually by λ or μ. All the theories can be easily extended to the joint convergence $\|\boldsymbol{\lambda}_T^* - \boldsymbol{\lambda}^*\|_2$ when r^* is also strongly convex. 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 $d' \in \Omega_d$. Here we provide two parallel approaches that can derive the optimal dual convergence rate. One is by localized Rademacher complexity and the other is by partition. Both feature a similar localization idea, which is critical to achieving fast rate $O(T^{-1})$. Define $D_t(\boldsymbol{\lambda}, d) := f_t^*(b_t^\top \nu) + r^*(-\mu) + d^\top(\nu - \mu)$, and the corresponding (sub)gradient

$$\phi_t(\boldsymbol{\lambda}, d) := \nabla D_t(\boldsymbol{\lambda}, d) = \begin{bmatrix} b_t \nabla f_t^*(b_t^\top \nu) + d \\ -\nabla r^*(-\mu) - d \end{bmatrix}.$$

Then we have $\nabla D(\boldsymbol{\lambda}, d) = \mathbb{E}\phi_t(\boldsymbol{\lambda}, d)$. Denote $\bar{\phi}_T(\boldsymbol{\lambda}, d) := T^{-1}\sum_{t=1}^T \phi_t(\boldsymbol{\lambda}, d)$ and the partial derivative w.r.t ν as $\bar{\phi}_{T,\nu}(\nu, d) = T^{-1}\sum_{t=1}^T b_t \nabla f_t^*(b_t^\top \nu) + d$. Both of two approaches focus on the partial second order term of $\bar{D}_T(\boldsymbol{\lambda}, d)$:

$$\bar{s}_T(\nu, d) := \bar{D}_T(\nu, \mu^*, d) - \bar{D}_T(\nu^*, \mu^*, d) - \underbrace{\left\langle \bar{\phi}_{T,\nu}(\nu^*, d), \nu - \nu^* \right\rangle}_{\text{first order term}} \tag{3.1}$$

We may write $\bar{s}_T(\nu)$ for simplicity and $s_t(\nu)$ is defined in a similar spirit.

3.1. Localized Rademacher complexity

We adopt the idea of localization in local Rademacher complexity (Bartlett et al. 2005, Koltchinskii 2006) to derive a tight probability bound of $\|\nu_T^* - \nu^*\|$. Assume ν_T^* is part of the variable that minimizes (2.7), and suppose $\|\nu_T^* - \nu^*\| \geq \varepsilon$. Since the expected second order term $s(\nu) := \mathbb{E}\bar{s}_T(\nu)$ shares a second-order growth property by Assumption 3, then by (3.1) and convexity of $\bar{D}_T(\boldsymbol{\lambda}, d)$, there exists a $\boldsymbol{\lambda} = [\nu^\top, \mu^\top]^\top$ where $\nu \in \mathbb{B}(\nu^*, \varepsilon)$ such that

$$\begin{aligned} \bar{s}_T(\nu) - s(\nu) &\leq \bar{D}_T(\boldsymbol{\lambda}, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) - \left\langle \bar{\phi}_T(\boldsymbol{\lambda}^*, d), \boldsymbol{\lambda} - \boldsymbol{\lambda}^* \right\rangle - s(\nu) + \langle \nabla D(\boldsymbol{\lambda}^*, d), \boldsymbol{\lambda} - \boldsymbol{\lambda}^* \rangle \\ &\leq -\frac{\sigma_{\min}\underline{\mathcal{L}}_f}{2}\varepsilon^2 + \left\| \nabla_\nu D(\boldsymbol{\lambda}^*, d) - \bar{\phi}_{T,\nu}(\nu^*, d) \right\| \varepsilon, \end{aligned} \tag{3.2}$$

where we use convexity and the optimality of $\boldsymbol{\lambda}^*$ for the first inequality and the optimality of $\boldsymbol{\lambda}_T^*$ for the second one. By concentration, it is clear that, for any $\varepsilon > 0$, the gradient $\bar{\phi}_{T,\nu}(\nu^*, d)$ concentrates to $\nabla_\nu D(\boldsymbol{\lambda}^*, d)$ with error upper bounded by $\frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\varepsilon$ with high probability. We can then ensure the empirical process:

$$\sup_{\nu \in \mathbb{B}(\nu^*, \varepsilon)} |\bar{s}_T(\nu) - s(\nu)| \geq \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\varepsilon^2$$

with high probability. Define localized Rademacher complexity of $\bar{s}_T$ within a small neighbourhood of ν^* as $\mathcal{R}_\varepsilon = \mathbb{E}_{\mathcal{P}}\mathbb{E}_\sigma \left[\sup_{\nu \in \mathbb{B}(\nu^*, \varepsilon)} \frac{1}{T} \sum_{t=1}^T \sigma_t s_t(\nu) \right]$, where σ_t are independent Rademacher random variables. By the convergence theory of empirical process (Boucheron et al. 2005, Koltchinskii 2011), we have the follow proposition.

Proposition 1. Under Assumption 1-4, the following inequality holds

$$\mathcal{R}_\varepsilon \leq \sqrt{2m\log(3K)}\frac{2\sqrt{n}\bar{b}D\varepsilon}{\sqrt{T}} + \frac{L_1\varepsilon^2}{K},$$

for any constant $K > 0$. Consequently, if $\varepsilon \geq \frac{64\sqrt{2}\sqrt{n}\bar{b}D}{\sqrt{T}\sigma_{\min}\underline{\mathcal{L}}_f}\sqrt{\log\frac{100M}{\sigma_{\min}\underline{\mathcal{L}}_f}}$, we have the following probabilistic bound:

$$\mathbb{P}\left(\|\nu_T^* - \nu^*\| \geq \varepsilon\right) \leq m\exp\left(-\frac{T\sigma_{\min}^2\underline{\mathcal{L}}_f^2\varepsilon^2}{8mn\bar{b}^2D^2}\right) + \exp\left(-\frac{T\sigma_{\min}^2\underline{\mathcal{L}}_f^2\varepsilon^2}{5000n\bar{b}^2D^2}\right)$$

3.2. Partition

Apart from localized Rademacher complexity, we can also control the dual convergence by providing a uniform lower bound of $\bar{s}_T$ within a small neighborhood of ν^* by partition. This argument will show that, with high probability, the empirical second order term $\bar{s}_T(\nu, d)$ is lower bounded by a quadratic function (Li and Ye 2021). The rationale is obvious. Since $\bar{s}_T(\nu, d)$ is always convex and converges to a deterministic convex function, the shape of $\bar{s}_T(\nu, d)$ in a small neighborhood of ν^* will be very close to $s(\nu, d)$ as long as T is large enough. Moreover, for a convex function, the

local behavior near the deterministic optimal solution is enough to guarantee the global properties of empirical optimal solutions. Consequently, its global optimal solution ν_T^* will lie in a small neighbourhood of ν^*. This delicate analysis also adopts a localization technique, which enables us to reach an optimal result sharper than Li and Ye (2021).

To investigate the second-order term, we focus on a small neighborhood of ν^*. For a constant $H>0$ (to be clarified soon), define $\Omega_\nu(\varepsilon):=\{\nu|\|\nu-\nu^*\|_\infty\le 4H\varepsilon\}$. Actually, it suffices to control the second order term for all dual variables in $\Omega_\nu(\varepsilon)$ since we shall show that ν_T^* belongs to $\Omega_\nu(\varepsilon)$ with a high probability. In order to control the shape of second order term in $\Omega_\nu(\varepsilon)$, we systematically split the region $\Omega_\nu(\varepsilon)$ and derive a uniform concentration bound. This enables us to successfully eliminate the $\mathrm{O}(\log\log T)$ factor and achieve a sharper dual convergence bound.

Proposition 2. Under Assumptions 1-4, we define $H=10\sqrt{nm\log m}\bar{b}D/(\sigma_{\min}\underline{\mathcal{L}}_f)$. Then, given any $\varepsilon>0$, the second order term $\bar{s}_T$ satisfies that for $\forall\nu\in\Omega_\nu(\varepsilon)$ and $\|\nu-\nu^*\|_2>2H\varepsilon$, there exists a corresponding $\nu'\in\Omega_\nu(\varepsilon)$ such that $\|\nu'-\nu^*\|_2\ge\|\nu-\nu^*\|_2$ and

$$\bar{s}_T(\nu,d)\ge\frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\|\nu'-\nu^*\|_2^2-\frac{2}{5}\sigma_{\min}\underline{\mathcal{L}}_f H\varepsilon\|\nu'-\nu^*\|_2$$

with probability at least $1-2\exp(-\frac{m(T\varepsilon^2-1)\log m}{4})$. Consequently, under the event that this inequality holds, if there exists a $\boldsymbol{\lambda}_T^*$ that minimizes $\bar{D}_T$, then it follows that $\|\nu_T^*-\nu^*\|\le 2H\varepsilon$ with probability at least $1-2m\exp(-\frac{T\sqrt{n}\bar{b}D\varepsilon^2\log m}{2})$.

3.3. Convergence result

Equipped with Proposition 1 or 2, we can derive the following $O(T^{-1})$ bound for dual convergence.

Theorem 1 **(Dual convergence)**. *Under Assumptions 1-4, there exists an absolute constant $C_1>0$ such that the empirical dual optimal solution satisfies*

$$\mathbb{E}\|\nu_T^*-\nu^*\|_2^2\le C_1\frac{\bar{b}^2D^2}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2}\frac{nm\log m}{T}. \tag{3.3}$$

Remark 2. Our dual convergence bound is sharper than that in Li and Ye (2021). Under our assumption, the $O(T^{-1})$ rate is unimprovable, as there always exists a distribution $\mathcal{P}\in\Xi$ that incurs an $\Omega(T^{-1})$ dual convergence rate. For further details, please refer to Appendix 10.4. The bound is of order $\widetilde{O}(mn)$ with respect to dimension n and constraint number m. This can be obtained by applying either Proposition 1 or Proposition 2. However, these two approaches offer different advantages for individuals with various purposes. Proposition 1 can be easily extended to other related classical statistical problems, such as ERM or the convergence of M-estimator, which are of concern in general statistics and machine learning communities. On the other hand, the result in Proposition 2 is stronger for optimization and can be adapted to study the behavior of other optimization methods, such as the convergence ϵ-optimal solution, an important result we will discuss later.

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 $\nu_T^*(d')$ and $\nu^*(d')$, respectively. Then, we actually have

$$\mathbb{E} \sup_{d' \in \Omega_d} \|\nu_T^*(d') - \nu^*(d')\|_2^2 \leq C_1 \frac{\bar{b}^2 D^2}{\sigma_{\min}^2 \underline{\mathcal{L}}_f^2} \frac{nm \log m}{T} \tag{3.4}$$

Bound (3.4) plays a critical role in our regret analysis since the re-solving strategy of our adaptive algorithm framework needs to update the resource constraints.

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

THEOREM 2 **(Convergence of dual approximate solution)**. *Under Assumptions 1-4, suppose $\boldsymbol{\lambda}_T^\epsilon$ is an ϵ-optimal solution that satisfies $\bar{D}_T(\boldsymbol{\lambda}_T^\epsilon, d) - \bar{D}_T(\boldsymbol{\lambda}_T^*, d) \leq \epsilon$. Then we have the following convergence of ϵ-optimal solution:*

$$\mathbb{E} \|\nu_T^\epsilon - \nu^*\|_2^2 \leq C_1 \frac{\bar{b}^2 D^2}{\sigma_{\min}^2 \underline{\mathcal{L}}_f^2} \frac{nm \log m}{T} + \frac{8\epsilon}{\sigma_{\min} \underline{\mathcal{L}}_f}$$

Theorem 2 explains how the approximation of dual solutions affects dual convergence. The accuracy remains valid as we directly optimize the deterministic dual function $D(\boldsymbol{\lambda}, d)$. Moreover, this theorem reveals that even if the empirical dual function $\bar{D}_T(\boldsymbol{\lambda}, d)$ 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 $\boldsymbol{\lambda}_T^\epsilon$ is a stochastic ϵ-optimal solution generated by stochastic optimization algorithm $\mathcal{B}$ that satisfies*

$$\mathbb{E}_{\mathcal{B}} \left[\bar{D}_T(\boldsymbol{\lambda}_T^\epsilon, d) - \bar{D}_T(\boldsymbol{\lambda}_T^*, d) \middle| \bar{D}_T \right] \leq \epsilon$$

for any given $\bar{D}_T$. Then we have the following convergence of the stochastic ϵ-optimal solution:

$$\mathbb{E} \|\nu_T^\epsilon - \nu^*\|_2^2 \leq C_2 \frac{\bar{b}^2 D^2}{\sigma_{\min}^2 \underline{\mathcal{L}}_f^2} \frac{nm \log m}{T} + C_3 \epsilon^{\frac{2}{3}} \left(m \left(2 \frac{\bar{f} + \bar{r}}{\underline{d}} + G \right) \right)^{\frac{1}{3}} / (\sigma_{\min} \underline{\mathcal{L}}_f)^{\frac{2}{3}},$$

where the expectation is taken with respect to $\mathcal{B}$ and $\mathcal{P}$, and C_2, C_3 are absolute constants.

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 ϵ. Compared to Theorem 2, the smaller order $\epsilon^{\frac{2}{3}}$ 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 $\mathbb{E}\left\|\nu_T^{\epsilon} - \nu_T^*\right\|_2^2$ 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 $\epsilon^{\frac{2}{3}}$. Theorem 2 and Corollary 1 are all based on Proposition 2. 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 non-linear and non-separable 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.

Our dual-based online allocation algorithm is in line with other dual-based online algorithms in spirit: we keep maintaining a dual variable $\boldsymbol{\lambda}_t$ by re-solving and whenever 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, and a_t given the dual variables by:

$$\begin{aligned}\tilde{x}_t(\nu) &:= \arg\max_{x\in\mathcal{X}}\{f_t(x) - (\lambda+\mu)^\top b_t x\} = -\nabla f_t^*(b_t^\top(\lambda+\mu)),\\ \tilde{a}(\mu) &:= \arg\max_{a\in\mathcal{Z}}\ \left\{r(a) + \mu^\top a\right\} = -\nabla r^*(-\mu)\end{aligned}$$

Note that ν consists of both λ and μ with different feasible sets, and can not be directly yielded by optimization with respect to a single variable. 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 SAA problems as shown in equation (4.1). Each $\boldsymbol{\lambda}_t$ is

a ϵ_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 ϵ_j 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 than other dual-based algorithms that do not learn from history (Devanur et al. 2019, Balseiro et al. 2022), at the cost of acceptable extra computation. As is common in the literature on dual-based online algorithms, we assume that 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 (λ_0, μ_0), and initial resource $B_0 := dT$.

for all $t = 1, \ldots, T$ **do**

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

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

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: $B_t := B_{t-1} - b_t x_t$ and average remaining resources: $d_t := \frac{B_t}{T-t}$

Update dual variable (λ_t, μ_t) via solving the following dual problem by any approximation algorithm $\mathcal{B}_t$ with accuracy ϵ_t:

$$\min_{(\lambda,\mu) \in \Omega_\lambda \times \Omega_\mu} \left\{ \bar{D}_t(\lambda, \mu, d_t) := \frac{1}{t} \sum_{l=1}^{t} f_l^*(b_l^\top(\mu + \lambda)) + r^*(-\mu) + d_t^\top \lambda \right\} \tag{4.1}$$

end for

Our algorithm framework is free of optimizer, that is, we can select any optimizer to get the ϵ_t-optimal solution to dual program (4.1). Since the dual problem $\bar{D}_t(\boldsymbol{\lambda}, d_t)$ is generally convex with respect to (λ, ν), one favorable 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 2 by our algorithm framework. Basically, it requires computing $O(t^3)$ stochastic gradients at time t. Moreover, if the dual problem $\bar{D}_t$ is further strongly convex or smooth, we can reduce the computational cost to $O(t)$ for each resolving. See Section 7.1 for more discussions on the case of strongly convex objectives. In Section 5, we demonstrate that any optimization algorithm $\mathcal{B}_t$ that achieves the rate of dual convergence $\mathbb{E}\,\|\nu_t - \nu^*(d_t)\|_2^2 = O(t^{-1})$ or $O\big((T-t)^{-1}\big)$ 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 $\boldsymbol{\mu}_0$, where $\boldsymbol{\mu} = [\lambda^\top, \mu^\top]^\top$ and initial resource $B_0 := dT$.

for all $t = 1, \ldots, T$ **do**

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

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

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: $B_t := B_{t-1} - b_t x_t$ and average remaining resources: $d_t := \frac{B_t}{T-t}$

Set $R := \sqrt{m\left(2\frac{\bar{f}+\bar{r}}{\underline{d}} + G\right)}$, $L := \sqrt{m\bar{d}^2 + 2n\bar{b}^2D^2 + nG^2}$, $K := t^3$, and $\eta_t := \frac{\sqrt{2}R}{L\sqrt{K}}$. Define $\boldsymbol{\mu}_t^0 := \boldsymbol{\mu}_{t-1}$

for all $k = 1, \ldots, K$ **do**

Randomly pick ζ from $[t] := \{1, \ldots, t\}$ with uniform distribution and calculate the gradient

$$\nabla D_\zeta(\boldsymbol{\mu}_t^{k-1}) := \begin{bmatrix} -b_\zeta \tilde{x}_\zeta(\nu_t^{k-1}) + d_t \\ -b_\zeta \tilde{x}_\zeta(\nu_t^{k-1}) + \tilde{a}(\mu_t^{k-1}) \end{bmatrix} \tag{4.2}$$

Update dual variable via stochastic gradient descent:

$$\boldsymbol{\mu}_t^k := \underset{\boldsymbol{\mu}\in\Omega_\lambda^+\times\Omega_\mu}{\arg\min}\left\{\langle \boldsymbol{\mu}, \nabla D_\zeta(\boldsymbol{\mu}_t^{k-1})\rangle + \frac{1}{2\eta_t}\left\|\boldsymbol{\mu} - \boldsymbol{\mu}_t^{k-1}\right\|_2^2\right\} \tag{4.3}$$

end for

Update dual variable by averaging: $\boldsymbol{\mu}_t := \frac{\sum_{k=1}^K \boldsymbol{\mu}_t^k}{K}$

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 $\boldsymbol{\lambda}_t$, 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 $\{\mathcal{B}_j\}_{j\geq 1}$ satisfies dual convergence condition 1 if

$$\mathbb{E}_{\mathcal{B},\mathcal{P}} \left\|\nu_t - \nu^*(d_t)\right\|^2 \leq C_4 \frac{1}{t+1}, \text{ or } \quad \mathbb{E}_{\mathcal{B},\mathcal{P}} \left\|\nu_t - \nu^*(d_t)\right\|^2 \leq C_4 (\frac{1}{t+1} + \frac{1}{T-t}) \tag{5.1}$$

for some constant $C_4 = \tilde{O}(nm)$. The expectation is taken with respect to all the $\{\mathcal{B}_j\}_{j\geq 1}$ 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 $\epsilon_t = \Theta\big((T-t)^{-1}\big)$ ($\epsilon_t = \Theta(t^{-3/2})$ 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 **(Regret upper bound)**. *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:*

$$\textit{Regret}(A) \leq \mathring{C} \cdot \log T$$

for some constant $\mathring{C} = O(m^2 n^2 \log m)$ depending on the values in Assumptions 1-5.

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 for the number of periods T. The regret upper bound depends quadratically on the dimension of constraints m and the dimension of action space n.

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 $\nu_t - \nu^*(d_t)$ and by $\mathbb{E}[T - \tau]$ for some stopping time τ. Recall, given a certain distribution $\mathcal{P}$, the definition of regret:

$$\text{Regret}\left(A|\mathcal{P}\right) = R^*(\mathcal{P}) - R\left(A|\mathcal{P}\right),$$

where $R^*(\mathcal{P})$ 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

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

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

LEMMA 2. *The offline maximum reward $R^*(\mathcal{P})$ satisfies $R^*(\mathcal{P}) \le T \cdot g(\nu^*)$.*

Note that, as shown by Bumpensanti and Wang (2020), Vera and Banerjee (2021), as long as our problem is degenerate, there will be a possible $\Omega(\sqrt{T})$ gap between fluid problem $T \cdot g(\nu^*)$ and the optimal $R^*(\mathcal{P})$, which means that this upper bound may not be tight in the degenerate case.

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

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

Note that our strategy of dealing with 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 $a_t = \tilde{a}(\mu_t)$ 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 μ_T and the following quantity:

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

The above minimization is taken without constraints, implying that $T^{-1}\sum_{t=1}^T b_t x_t = -\nabla r^*(-\hat{\mu}_T)$.

We can now describe the following regret decomposition for a general stopping time.

PROPOSITION 3. Under Assumptions 1-5, for a proper stopping time τ ensuring that the resource is not depleted before $t \le \tau$, the regret of our dual-based adaptive framework Algorithm 1 admits the following upper bound:

$$\begin{aligned}\text{Regret}\,(A|\mathcal{P}) \le & \underbrace{\mathbb{E}\left[\sum_{t=1}^{\tau} g(\nu^*) - g(\nu_t)\right]}_{\text{R.1}} + \underbrace{\mathbb{E}\left[2(\bar{f} + \bar{r} + C_3)(T - \tau) + \left\langle \lambda^*, \sum_{t=1}^{\tau}(d - b_t x_t)\right\rangle\right]}_{\text{R.2}}, \\ & + \underbrace{\mathbb{E}\left[\left\langle \mu^* - \hat{\mu}_T, \sum_{t=1}^{\tau}(\tilde{a}(\mu^*) - b_t x_t)\right\rangle\right]}_{\text{R.3}},\end{aligned} \tag{5.4}$$

where $C_3 := \sqrt{mn}GD\bar{b}$.

It remains to bound the two parts in Proposition 3, respectively. The key point is to carefully choose a stopping time that (1) avoids early stopping; (2) enforces the total resource constraints. The first term R.1 is contributed by the algorithm before stopping time, which can be controlled by the cumulative dual error $\mathbb{E}\sum_{t=1}^{\tau}\|\nu_{t-1}-\nu^*\|^2$. The second term R.2 concerns the lost rewards due to resource depletion, which can be controlled by $\mathbb{E}(T-\tau)$. To achieve an $O(\log T)$ regret, the stopping time shall be carefully designed so that $\mathbb{E}(T-\tau)=O(\log T)$. The term R.3 is contributed mainly by the variable splitting, which can be controlled jointly by the cumulative dual error and $\mathbb{E}(T-\tau)$. The three terms capture different sources of regret induced by our adaptive framework Algorithm 1. It turns out that we shall bound $\mathbb{E}\sum_{t=1}^{\tau}\|\nu_{t-1}-\nu^*\|^2$ and $\mathbb{E}(T-\tau)$, for which a smart design of stopping time becomes crucial.

Our design of stopping time is inspired by the budget-ratio stopping time introduced and investigated by Arlotto and Gurvich (2019) and Li and Ye (2021) for online linear allocation problems. At the core of this design is a smart strategy that ensures, as the updated constraint d_t varies within a region $\mathcal{D}\subset\Omega_d$, the binding and non-binding dimensions of the problem $D(\boldsymbol{\lambda},d_t)$ remain unchanged. The region $\mathcal{D}$ is usually a small neighbor of the original budget d. The following lemma dictates that such a region $\mathcal{D}$ exists for our regularized online convex allocation problem. Recall that $\boldsymbol{\lambda}^*(d')$ denotes the optimal dual solution to $D(\boldsymbol{\lambda},d')$, and I_{B} and I_{NB} stand for the binding and non-binding dimension of $D(\boldsymbol{\lambda},d)$.

LEMMA 3. *Under Assumptions 1-5, there exists a constant $\delta_d>0$ such that for any $d'\in\Omega_d$, if*

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

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

With Lemma 3, we define the required region where binding and non-binding dimensions remain unchanged during iterations by

$$\mathcal{D}:=\left\{d'\in\Omega_d\middle|-\delta_d\le d'_i-d_i\le\delta_d \text{ if } i\in I_{\mathsf{B}}, \text{ and } d'_i-d_i\ge-\delta_d \text{ if } i\in I_{\mathsf{NB}}\right\}.$$

We thereby design the following stopping time.

$$\tau:=\min_{t\in[T]}\left\{T-\left\lceil\frac{\sqrt{n}D\bar{b}}{\underline{d}}\right\rceil\right\}\bigcup\{t|d_t\notin\mathcal{D}\}. \tag{5.5}$$

Additionally, this stopping time also guarantees that resource depletion will not happen before τ. We show that τ 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 4. *Under Assumptions 1-5, Algorithm 1 with selected dual optimizer* $\{\mathcal{B}_t\}_{t\geq 1}$ *satisfying Condition 1 achieves*

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

LEMMA 5. *Under Assumptions 1-5, the stopping time (5.5) of Algorithm 1 with selected dual optimizer* $\{\mathcal{B}_t\}_{t\geq 1}$ *satisfying Condition 1 has*

$$\mathbb{E}\left(T-\tau\right)\leq O(\log T) \tag{5.7}$$

These two lemmas play a key role in our regret analysis. They are proved by investigating the dynamic behavior of constraints d_{it} for binding and non-binding dimensions, respectively. For binding dimensions, we investigate the recurrence relation of d_{it} by leveraging the binding relations. For the non-binding dimensions, we exploit the δ_d gap between d_{it} and average resource consumption. Since the regret is jointly controlled by $\mathbb{E}\sum_{t=1}^{\tau}\left\|\nu_{t-1}-\nu^*\right\|^2$ and $\mathbb{E}(T-\tau)$, we conclude the $O(\log T)$ regret order. We defer the detailed proof to Appendix 10.12.

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. We establish a matching lower bound in this section to show the optimality of Theorem 3. 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 $\max_{x_t\in\mathcal{X}}\left\{\sum_{t=1}^T f_t(x_t)\text{ s.t. }\sum_{t=1}^T b_t x_t\leq dT\right\}$, it holds that $\sum_{t=1}^T b_t x'_t = dT$ and thus $r(T^{-1}\sum_{t=1}^T b_t x'_t)\geq r(T^{-1}\cdot\sum_{t=1}^T b_t x_t)$ for any other $\{x_t\}_{t=1}^T$. This renders the regret lower bound of regularized problem larger than that of a non-regularized one. Therefore, we only focus on the non-regularized problems for the regret lower bound.

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 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 1 and whether achieving an optimal regret without the adaptive constraints update is possible. Here we only present a negative answer for algorithms that do not update constraints, partially for two specific but renowned algorithms. In this discussion, we call "algorithms without

constraint update" as dual algorithms seeking to approximately optimize the fluid problem $D(\boldsymbol{\lambda}, d)$ in the whole process without changing d and the objective. For concreteness, we investigate two similar algorithms (described in Algorithms 3) without constraints update that have been discussed in the literature for online dual gradient (mirror descent (Balseiro et al. 2021b, 2022) and dual SAA (Li and Ye 2021). Algorithm 3 try to minimize $D(\boldsymbol{\lambda}, d)$ by either Stochastic Approximation (SA) or by solving SAA. In the strongly convex case, it is clear that they all converge with $\mathbb{E}\left\|\boldsymbol{\lambda}_t - \boldsymbol{\lambda}^*\right\| = O(1/\sqrt{t})$ by Rakhlin et al. (2012) and Section 3.

Algorithm 3 Online dual gradient (mirror) descent or SAA without constraint update

Require: regularizer r, iteration number T, step size $\eta_t := \Theta(\frac{1}{t})$ for $t \in [T]$, start point $\boldsymbol{\mu}_0 = [\lambda_0^\top, \mu_0^\top]^\top$, and initial resource $B_0 := dT$.

for all $t = 1, \ldots, T$ **do**

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

Calculate $\tilde{x}_t := -\nabla f_t^*(b_t^\top(\lambda_{t-1} + \mu_{t-1}))$, $\tilde{a}_t := -\nabla r^*(-\mu_{t-1})$.

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 $\nabla D_t(\boldsymbol{\mu}_{t-1}, d) := \begin{bmatrix} -b_t \tilde{x}_t + d \\ -b_t \tilde{x}_t + a_t \end{bmatrix}$

Update dual variable via online gradient descent:

$$\boldsymbol{\mu}_t := \underset{\boldsymbol{\mu} \in \Omega_\lambda^+ \times \Omega_\mu}{\arg\min} \left\{ \langle \boldsymbol{\mu}, \nabla D_t(\boldsymbol{\mu}_{t-1}, d) \rangle + \frac{1}{2\eta_t} \left\| \boldsymbol{\mu} - \boldsymbol{\mu}_{t-1} \right\|_2^2 \right\}$$

Or via solving t-sample SAA:

$$\boldsymbol{\mu}_t := \underset{\boldsymbol{\mu} \in \Omega_\lambda^+ \times \Omega_\mu}{\arg\min} \left\{ \frac{1}{t} \sum_{j=1}^{t} f_j^*(b_j^\top(\mu + \lambda)) + r^*(-\mu) + d^\top \lambda \right\}$$

end for

The following lemma establishes an $\Omega(T^{1/2})$ regret lower bound for these two algorithms.

THEOREM 5. *Under Assumptions 1-4, there exists a constant $c_2 > 0$ such that any dual-based algorithm A attempting to approximate ν^* with $\mathbb{E}\left\|\nu_t - \nu^*\right\|_2 \leq c_2 D(t+1)^{-1/2}$ 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 Binomial distribution. Note that the lower bound can also be controlled by both dual approximate error $\mathbb{E}\sum_{t=1}^{\tau} \left\|\nu_{t-1} - \nu^*\right\|_2^2$ and early stopping effect

$\mathbb{E}(T-\tau)$. Here ν^* is the deterministic dual solution when the resource constraint is fixed at d. In sharp contrast, the dual solution ν_t in our adaptive framework Algorithm 1 aims to approximate $\nu^*(d_t)$ where d_t is the updated constraint at time t. Intuitively, the rationale behind constraint update is that at time t, and the decision should be made considering the remaining resources d_t at hand instead of the initial resource d. Algorithm 3, however, without constraint update, suffers from early stopping $\mathbb{E}(T-\tau) \geq \Omega(\sqrt{T})$ and is thus not optimal.

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

Finally, we remark that our theorem pushes forward the understanding of adaptiveness for online algorithms to 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, as in Algorithm 3, focusing on fixed constraints leads to a sub-optimal early stopping.

6. Infrequent Resolving and Fast algorithms

Although Algorithm 1 only requires inexact resolving, its frequency of solving convex programming is of order $O(T)$ for T periods. This raises the question of whether it is possible to further reduce the computational burden through infrequent resolution or other faster algorithms. Here we extend our previous result to both infrequent resolving and fast algorithm design, showing that (i) we can still achieve optimal regret by infrequent resolving given a good initialization; (ii) we can reach sub-optimal $O(\log^2 T)$ regret under linear computational cost. These two algorithms significantly alleviate the computational cost while ensuring a good regret performance.

Define rate $\rho \in (0,1)$ for resolving that satisfies: $\rho \geq \left\lceil \frac{\sqrt{n}D\bar{b}}{\underline{d}} \right\rceil / T \vee (1 - \delta_d/(\bar{d} + \sqrt{n}\bar{b}D + \delta_d))$, i.e., $t \leq T - \left\lceil \frac{\sqrt{n}D\bar{b}}{\underline{d}} \right\rceil$ and $d_t \in \mathcal{D}$ for any $t \leq (1-\rho)T$. We describe the infrequent resolving algorithm in Algorithm 4 and the fast algorithm in Algorithm 5. The ideas behind Algorithm 4 and 5 are similar: we split the total period T into $O(\log T)$ decreasing epochs, and only update the remaining constraints d_t for resolving at the beginning of each epoch $t = T_j$. The distinction is that, Algorithm 4 inexactly solve the convex programming at time T_j, but Algorithm 5 exploits the following epoch from T_j to T_{j+1} to perform stochastic approximation as to minimize $D(\boldsymbol{\lambda}, d_{T_j})$.

Algorithm 4 Infrequent resolving

Require: regularizer r, iteration number T, start point (λ_0, μ_0), and initial resource $B_0 := dT$.

Set $J = \lceil \log_{\frac{1}{\rho}} T \rceil$, and $T_j = T - \lceil \rho^j T \rceil$ for $j \in [J]$

for all $t = 1, \ldots, T$ **do**

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

Calculate $\tilde{x}_t := -\nabla f_t^*(b_t^\top(\lambda_{t-1} + \mu_{t-1}))$.

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

if $t = T_j$ for some j **then**

Compute average remaining resources: $d_t := \frac{B_t}{T-t}$

Update dual variable (λ_t, μ_t) via solving the following dual problem by any approximation algorithm $\mathcal{B}_t$ with accuracy ϵ_t:

$$\min_{(\lambda,\mu)\in\Omega_\lambda\times\Omega_\mu} \left\{ \bar{D}_t(\lambda, \mu, d_t) := \frac{1}{t}\sum_{l=1}^{t} f_l^*(b_l^\top(\mu + \lambda)) + r^*(-\mu) + d_t^\top \lambda \right\}$$

else

Let $(\lambda_t, \mu_t) = (\lambda_{t-1}, \mu_{t-1})$, $d_t = d_{t-1}$ with no update

end if

end for

THEOREM 6 **(Infrequent resolving)**. *Suppose Assumption 1-5 hold. Given the initial point* $\boldsymbol{\lambda}_0 = [\nu_0^\top, \mu_0^\top]^\top$ *satisfying* $\mathbb{E}\,\|\nu_0 - \nu^*\|_2^2 = O(1/T)$*, and under Condition 1, Algorithm 4 enjoys an optimal regret upper bound*

$$\textit{Regret}(A) \leq \mathring{C} \cdot \log T,$$

for some constant $\mathring{C} = O(m^2 n^2 \log m)$ *depending on the values in Assumptions 1-5. Here in the regret, the expectation is taken also with respect to* ν_0.

THEOREM 7 **(Sub-optimal fast algorithm)**. *Suppose Assumption 1-5 hold. Algorithm 5 achieves sub-optimal regret bound*

$$\textit{Regret}(A) \leq \widetilde{C} \log^2 T,$$

for some constant $\widetilde{C} = O(mn^2)$ *depending on the values in Assumptions 1-5.*

REMARK 4. The initialization in Theorem 6 serves to ensure the performance in the first epoch. This initialization requirement can be met, for instance, by employing dual optimization based on previously collected data as side information. The fast algorithm presented in Algorithm 5 updates the dual variable using one-step online gradient descent, resulting in a linear computational

Algorithm 5 Fast algorithm

Require: regularizer r, iteration number T, start point $\boldsymbol{\mu}_0 = [\lambda_0^\top, \mu_0^\top]^\top$, and initial resource $B_0 := dT$. Set $l = 0$

Set $J = \lceil \log_{\frac{1}{\rho}} T \rceil$, and $T_j = T - \lceil \rho^j T \rceil$ for $j \in [J]$

for all $t = 1, \ldots, T$ **do**

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

Calculate $\tilde{x}_t := -\nabla f_t^*(b_t^\top(\lambda_{t-1} + \mu_{t-1}))$, $\tilde{a}_t := -\nabla r^*(-\mu_{t-1})$.

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

if $t = T_j$ for some j **then**

Let $l = T_j$. Compute average remaining resources: $d_t := \frac{B_t}{T-t}$

else

Let $d_t = d_{t-1}$ with no constraint update

end if

Calculate the stochastic gradient with updated constraint $\nabla D_t(\boldsymbol{\mu}_{t-1}, d_t) := \begin{bmatrix} -b_t \tilde{x}_t + d_t \\ -b_t \tilde{x}_t + a_t \end{bmatrix}$

Set step size $\eta_t := \Theta(\frac{1}{t-l+1})$ and update dual variable via online gradient descent:

$$\boldsymbol{\mu}_t := \operatorname*{arg\,min}_{\boldsymbol{\mu} \in \Omega_\lambda^+ \times \Omega_\mu} \left\{ \langle \boldsymbol{\mu}, \nabla D_t(\boldsymbol{\mu}_{t-1}, d_t) \rangle + \frac{1}{2\eta_t} \left\| \boldsymbol{\mu} - \boldsymbol{\mu}_{t-1} \right\|_2^2 \right\}$$

end for

cost. In contrast to Balseiro et al. (2022), our Algorithm 5 incorporates constraint updates, which effectively helps us avoid the $\Omega(\sqrt{T})$ lower bound stated in Theorem 4 and achieve logarithmic regret. Compared with the previous results, our Algorithm 5 only depends linearly on the number of constraints. Due to the reliance of dual convergence, the polynomial dependence of dimensions ($O(mn)$) is unavoidable in regret analysis of all the methods in our framework without further assumptions on the problem.

7. Applications

7.1. Strongly convex dual problems

We consider a special but practical setting, in which our empirical dual problem $\bar{D}_t(\boldsymbol{\lambda}, d_t)$ in (4.1) is always $\underline{\mathcal{L}}_D$-strongly convex. This assumption can be met if f_t^* and r are almost surely strongly convex. In this case, we only need to do SGD for $O(t)$ times at time t to make our algorithm theoretically optimal. Simply modify algorithm by setting $K := t$, and $\eta_k := \frac{\underline{\mathcal{L}}_D}{k}$, and take $\boldsymbol{\mu}_t := \boldsymbol{\mu}_t^K$

Notice that $\mathbb{E}\left\|\nu_t-\nu^*(d_t)\right\|_2^2 \le 2\mathbb{E}\left\|\nu_t-\nu_t^*(d_t)\right\|_2^2 + 2\mathbb{E}\left\|\nu_t^*(d_t)-\nu^*(d_t)\right\|_2^2$ where $\nu_t^*(d_t)$ is the optimal solution to the empirical dual problem $\bar{D}_t(\boldsymbol{\lambda}, d_t)$. The second term $\mathbb{E}\left\|\nu_t^*(d_t)-\nu^*(d_t)\right\|_2^2$ 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)).

7.2. Online linear programming

An instant application of our algorithm is 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 $x_t \in \mathcal{X} = [0, D]^n$ 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:

$$\begin{aligned} \max_{x_t} \quad & \sum_{t=1}^{T} v_t^\top x_t \\ \text{s.t.} \quad & \sum_{t=1}^{T} b_t x_t \preceq dT,\ d \in \mathbb{R}_+^m \\ & x_t \in [0, D]^n, \forall t \in [T]. \end{aligned}$$

The empirical dual problem and its population version can be explicitly written as

$$\bar{D}_T(\lambda, d) := \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\left(v_{it} - b_{it}^\top\lambda\right)^+}{T} + d^\top\lambda, \text{ and } \ D(\lambda, d) := \mathbb{E}\sum_{i=1}^{n}\left(v_{it} - b_{it}^\top\lambda\right)^+ + d^\top\lambda,$$

which is in line with Li and Ye (2021). Here the index b_{it} means the i-column of b_t. For a given dual variable λ, we make the primal decision by $x_{it} := D\mathbb{I}(v_{it} - b_{it}^\top\lambda > 0)$ 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 approximate solution λ_t. Towards that end, an $O(\log T)$ regret is attainable, which improves prior result (Li and Ye 2021). Assumption 4 seems stronger than the smoothness of expected primal solutions because the former implies the latter. However, in the case of linear programming they are actually equivalent because, for any $\delta > 0$, we have

$$\mathbb{E}\left[\sup_{\lambda:\left\|b_t^\top(\lambda-\lambda^*)\right\|\le\delta} D\mathbb{I}(b_{it}^\top\lambda^* \le v_{it} \le b_{it}^\top\lambda)\middle| b_t\right] \le D\,\mathbb{P}(b_{it}^\top\lambda^* \le v_{it} \le b_{it}^\top\lambda + \delta) \le O(\delta),$$

i.e., the smooth of the expected primal decision also implies Assumption 4.

7.3. Online max-min fairness and load-balancing allocation

Our algorithm framework applies to online max-min fairness allocation, which is a well-accepted fairness criterion used in various real-world problems, including bandwidth allocation (Salles and Barria 2008), routing and load-balancing (Nace et al. 2006), and classroom allocations (Kurokawa et al. 2015). To satisfy the condition on regularizer in Assumption 5, we set $r(a) = \kappa\min_i(a_i/d_i)$

and require the (scaled) expected optimal constraint $\{d_i^*/d_i := \mathbb{E}\left(b_t\tilde{x}_t(\nu^*)\right)_i/d_i,\ i\in[m]\}$ to have one unique minimum element.

We can also apply our algorithm framework to online load-balancing problems, which have been studied in various fields such as bandwidth allocation (Bejerano et al. 2004), network design (Radunovic and Le Boudec 2007), distributed system design (Xu et al. 2011), and other various scenarios. To avoid over-exploitation, we select negative max loss: $r(a) := -\kappa\max_i(a_i/d_i)$ as our regularizer. To satisfy the regularizer condition, we only need the (scaled) expected optimal constraint $\{d_i^*/d_i := \mathbb{E}\left(b_t\tilde{x}_t(\nu^*)\right)_i/d_i,\ i\in[m]\}$ to have one unique maximum element.

Other regularizers like ℓ_1-loss, hinge loss (Balseiro et al. 2021b), and entropy loss are also applicable as long as the optimal resource consumption d^* is located in a small smooth region of regularizer r only for non-binding dimensions.

8. Numerical Experiments

We implement Resolving with SGD as a showcase for our proposed algorithmic framework. The performance is assessed under 4 different stochastic input models. Due to limited space, the numerical results are relegated to Appendix 11. The results show the superiority of our algorithms in regret control compared with other aforementioned algorithms. Our algorithm exhibits $O(\log T)$ regret within all different input models, which corroborates our Theorem 3 and 4. We also display the resource consumption dynamic to visualize our algorithmic framework's optimal resource control. To compare the impact of different regularizers, we plot the regrets on different regularization levels and show the trade-off between maximizing the reward and penalizing the regularization. The regret performance is very robust to different regularization levels.

9. Discussion

This paper investigated regularized online convex allocation problems with a non-separable regularizer. While a polynomial-time adaptive algorithm framework is proven optimal in controlling regret, several interesting yet challenging questions remain open. One is the necessity of the non-degeneracy assumption. Recently, Bumpensanti and Wang (2020) showed that the non-degeneracy assumption is unnecessary for re-solving heuristics 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 challenge is the input model. Throughout this paper, we only discussed online allocation problems under the stochastic input model. The behavior of re-solving algorithms for other input models like random permutation or adversarial inputs 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.

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.

Balseiro, S., Besbes, O., and Pizarro, D. (2021a). Survey of dynamic resource constrained reward collection problems: Unified model and analysis. *Available at SSRN 3963265*.

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. (2021b). Regularized online allocation problems: Fairness and beyond. In *International Conference on Machine Learning*, pages 630–639. PMLR.

Balseiro, S. R., Besbes, O., and Pizarro, D. (2023). Survey of dynamic resource-constrained reward collection problems: Unified model and analysis. *Operations Research*.

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*.

Bartlett, P., Bousquet, O., and Mendelson, S. (2005). Local rademacher complexities. *Annals of Statistics*, 33(4):1497–1537.

Bejerano, Y., Han, S.-J., and Li, L. (2004). Fairness and load balancing in wireless lans using association control. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 315–329.

Bertsekas, D. and Gallager, R. (2021). *Data networks*. Athena Scientific.

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

Boucheron, S., Bousquet, O., and Lugosi, G. (2005). Theory of classification: A survey of some recent advances. *ESAIM: probability and statistics*, 9:323–375.

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

Bubeck, S. et al. (2015). Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357.

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.

Celli, A., Colini-Baldeschi, R., Kroer, C., and Sodomka, E. (2022). The parity ray regularizer for pacing in auction markets. In *Proceedings of the ACM Web Conference 2022*, pages 162–172.

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

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.

Gupta, V. (2022). Greedy algorithm for multiway matching with bounded regret. *Operations Research*.

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.

Hu, Y., Kallus, N., and Mao, X. (2022). Fast rates for contextual linear optimization. *Management Science*, 68(6):4236–4245.

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.

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.

Kerimov, S., Ashlagi, I., and Gurvich, I. (2021). On the optimality of greedy policies in dynamic matching. *Available at SSRN*.

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.

Koltchinskii, V. (2006). Local rademacher complexities and oracle inequalities in risk minimization. *The Annals of Statistics*, pages 2593–2656.

Koltchinskii, V. (2011). *Oracle inequalities in empirical risk minimization and sparse recovery problems: École D'Été de Probabilités de Saint-Flour XXXVIII-2008*, volume 2033. Springer Science & Business Media.

Kurokawa, D., Procaccia, A. D., and Shah, N. (2015). Leximin allocations in the real world. In *Proceedings of the sixteenth ACM conference on economics and computation*, pages 345–362.

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*.

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

Nace, D., Doan, N.-L., Gourdin, E., and Liau, B. (2006). Computing optimal max-min fair resource allocation for elastic flows. *IEEE/ACM Transactions on Networking*, 14(6):1272–1281.

Nash, J. F. (1950). The bargaining problem. *Econometrica*, 18(2):155–162.

Radunovic, B. and Le Boudec, J.-Y. (2007). A unified framework for max-min and min-max fairness with applications. *IEEE/ACM Transactions on networking*, 15(5):1073–1083.

Rakhlin, A., Shamir, O., and Sridharan, K. (2012). Making gradient descent optimal for strongly convex stochastic optimization. In *Proceedings of the 29th International Coference 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.

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

Salles, R. M. and Barria, J. A. (2008). Lexicographic maximin optimisation for fair bandwidth allocation in computer networks. *European Journal of Operational Research*, 185(2):778–794.

Shapiro, A., Dentcheva, D., and Ruszczynski, 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.

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

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

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.

Xu, J., Lam, A. Y., and Li, V. O. (2011). Chemical reaction optimization for task scheduling in grid computing. *IEEE Transactions on Parallel and Distributed systems*, 22(10):1624–1631.

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.

Supplement to " Optimal Regularized Online Allocation by Adaptive Re-Solving "

10. Proofs of Main Results

10.1. Proof of Lemma 1

We prove the bound of the deterministic optimal solution. Consider $\Omega'_\mu = \{-\nabla r(a) | a \in \mathcal{Z}\}$. The bounded subgradient in Assumption 2 suggests that the dual variable region Ω'_μ we defined is bounded by G. We explain this definition by the optimal conditions of stochasic programming. Note that for problem (2.8), μ is unconstrained. The optimal condition suggests that

$$\nabla r^*(-\mu^*) = \mathbb{E} b_t \nabla f_t^*(b_t^\top(\lambda^* + \mu^*))$$

if we assume fubini theorem holds. Then by the Fenchel conjugate, we have $\mu^* \in -\nabla r(\mathbb{E} b_t \tilde{x}_t)$. This shows that by defining Ω'_μ we indeed define the possible region that contains optimal solution μ^*, i.e., $\mu^* \in \Omega_\mu$. Thus we have $\|\mu\|_\infty \le G$.

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

$$\begin{aligned} D(\boldsymbol{\lambda}^*, d) &= \mathbb{E} \sup_x \{f_t(x) - (\lambda^* + \mu^*)^\top b_t x_t\} + \sup_a \{r(a) + a^\top \mu^*\} + d^\top \lambda^* \ge \mathbb{E} f_t(0) + r(0) + d^\top \lambda^* \\ &> (\bar{f} + \bar{r}) \ge D(\mathbf{0}, d), \end{aligned}$$

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

The following proposition on the growth of second order term $s(\nu, d)$ will be useful in the development of our theory.

Proposition 4. Under Assumptions 1-3, the second order objective function $s(\nu, d)$ in stochastic program satisfies the following growth condition:

$$\underline{\mathcal{L}}_s \|\nu - \nu^*\|_2^2 \le s(\nu) \le \bar{\mathcal{L}}_s \|\nu - \nu^*\|_2^2, \tag{10.1}$$

where the constant $\underline{\mathcal{L}}_s := \sigma_{\min} \underline{\mathcal{L}}_f / 2$, $\bar{\mathcal{L}}_s := \bar{b}^2 \bar{\mathcal{L}}_f / 2$.

Proof: By definition, we have

$$\begin{aligned} s(\nu) &= D(\nu, \mu^*, d) - D(\nu^*, \mu^*, d) - \nabla_\nu D(\nu^*, \mu^*, d)^\top (\nu - \nu^*) \\ &= \int_0^1 \left[\nabla D_\nu(z(\nu - \nu^*) + \nu^*, \mu^*, d) - \nabla D_\nu(\nu^*, \mu^*, d)\right]^\top (\nu - \nu^*) dz, \end{aligned}$$

where $\nabla_\nu D(\nu, d) = \mathbb{E} b_t f_t^*(b_t^\top v)$. Then for any z, we have

$$
\begin{aligned}
&\left[\nabla D_\nu(z(\nu-\nu^*)+\nu^*,\mu^*,d)-\nabla D_\nu(\nu^*,\mu^*,d)\right]^\top(\nu-\nu^*)\\
\le&\left\|\mathbb{E}b_t\nabla f_t^*(b_t^\top(z(\mu+\lambda-\mu^*-\lambda^*)+\mu^*+\lambda^*)-\mathbb{E}b_t\nabla f_t^*(b_t^\top(\mu^*+\lambda^*))\right\|_2\|\nu-\nu^*\|_2\\
\le&\left\|z\bar{\mathcal{L}}_f\bar{b}\mathbb{E}\left[b_t^\top(\mu+\lambda-\mu^*-\lambda^*)\right]\right\|_2\|\nu-\nu^*\|\\
\le&z\bar{\mathcal{L}}_f\bar{b}^2\|\nu-\nu^*\|^2,
\end{aligned}
$$

where the second inequality is by Assumption 3.1 when conditioned on b_t. By the integral of z we have $s(\nu)\le\bar{\mathcal{L}}_s\|\nu-\nu^*\|_2^2$. For the next direction, it is also clear that

$$
\begin{aligned}
&\left[\nabla D_\nu(z(\nu-\nu^*)+\nu^*,\mu^*,d)-\nabla D_\nu(\nu^*,\mu^*,d)\right]^\top(\nu-\nu^*)\\
=&\left(\mathbb{E}b_t\nabla f_t^*(b_t^\top(z(\mu+\lambda-\mu^*-\lambda^*)+\mu^*+\lambda^*)-\mathbb{E}b_t\nabla f_t^*(b_t^\top(\mu^*+\lambda^*))\right)^\top(\nu-\nu^*)\\
=&\mathbb{E}\left[\mathbb{E}\left[\langle\nabla f_t^*(b_t^\top(z(\mu+\lambda-\mu^*-\lambda^*)+\mu^*+\lambda^*))-\nabla f_t^*(b_t^\top(\mu^*+\lambda^*)),b_t^\top(\mu+\lambda-\mu^*-\lambda^*)\rangle\right]\middle|b_t\right]\\
\ge&z\underline{\mathcal{L}}_f\mathbb{E}\left\|b_t^\top(\mu+\lambda-\mu^*-\lambda^*)\right\|_2^2\ge z\underline{\mathcal{L}}_f\sigma_{\min}\|\mu+\lambda-\mu^*-\lambda^*\|_2^2.
\end{aligned}
$$

10.2. Proof of Proposition 1: Empirical Risk Minimization

In this proof, we generalize our discussion to a broader setting: we consider the convergence of classical ERM method. In many areas of statistics and machine learning research, we aims to solve the following problem, named Empirical Risk Minimization (ERM): given T empirical convex risk functions $\ell(\boldsymbol{\lambda},\xi_t):\mathbb{R}^m\to\mathbb{R},t\in[T]$ where ξ_t are i.i.d realizations from an unknown distribution, with its population version $\mathfrak{L}(\boldsymbol{\lambda})=\mathbb{E}_\xi\ell(\boldsymbol{\lambda},\xi_t)$, we seek to find a good parameter $\widehat{\boldsymbol{\lambda}}$ by minimizing the empirical risk

$$
\widehat{\boldsymbol{\lambda}}=\arg\min_{\lambda\in\mathbb{R}^m}\bar{\ell}(\boldsymbol{\lambda},\{\xi_t\}_{t=1}^T)=\arg\min_{\lambda\in\mathbb{R}^m}\frac{1}{T}\sum_{t=1}^T\ell(\boldsymbol{\lambda},\xi_t)
$$

as a proxy of the parameter that minimize the population risk $\boldsymbol{\lambda}^*=\arg\min_{\boldsymbol{\lambda}}\mathfrak{L}(\boldsymbol{\lambda})$. In statistic, such $\widehat{\boldsymbol{\lambda}}$ is also called M-estimator.

The following approach will show that, under second order growth condition of $\mathfrak{L}(\boldsymbol{\lambda})$, the ERM method provides a estimate $\widehat{\boldsymbol{\lambda}}$ with optimal convergence rate $\mathbb{E}\left\|\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^*\right\|^2=O(\frac{1}{T})$

Assumption 6 **(Lipschitz continuity)**. *Suppose the subgradient of each $\ell(\boldsymbol{\lambda},\xi_t)$ satisfies*

$$
\|\nabla\ell(\boldsymbol{\lambda},\xi_t)\|\le L
$$

Assumption 7 **(Second order growth)**. *The population risk satisfies the following second order growth*

$$
\langle\nabla\mathfrak{L}(\boldsymbol{\lambda})-\nabla\mathfrak{L}(\boldsymbol{\lambda}^*),\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\rangle\ge\underline{\mathcal{L}}_\ell\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\|_2^2
$$

ASSUMPTION 8 **(Smoothness of the first moment)**. *For any* $\boldsymbol{\lambda} \in \mathbb{R}^m$, *with* $\delta = \|\boldsymbol{\lambda} - \boldsymbol{\lambda}^*\|$, *we have*

$$\mathbb{E} \sup_{\boldsymbol{\lambda}' \in \mathbb{B}(\boldsymbol{\lambda}^*, \delta)} \left\| \nabla \ell(\boldsymbol{\lambda}', \xi_t) - \nabla \ell(\boldsymbol{\lambda}^*, \xi_t) \right\| \leq M\delta,$$

where $\mathbb{B}(\boldsymbol{\lambda}^*, r) = \left\{ \boldsymbol{\lambda}' : \left\| \boldsymbol{\lambda}' - \boldsymbol{\lambda}^* \right\| \leq r \right\}$.

To investigate the convergence of $\widehat{\boldsymbol{\lambda}}$, one classical approach is to compute the statistical complexity of function group $\{\ell(\boldsymbol{\lambda}, \xi_t), \boldsymbol{\lambda} \in \Theta\}$ to get an uniform bound of $\sup_{\boldsymbol{\lambda} \in \Theta} \left[\bar{\ell}(\boldsymbol{\lambda}, \{\xi_t\}_{t=1}^T) - \mathfrak{L}(\boldsymbol{\lambda}) \right]$. However, this approach may fail to reach the optimal convergence rate because it neglects the second order information. Instead, we consider the second order part of losses and improve our analyses by a localized argument near the optimal solution $\boldsymbol{\lambda}^*$. Equipped with localized Rademacher complexity which shares a similar idea as Bartlett et al. (2005), we are able to derive a sharp local probabilistic bound of $\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^*\|$. To fix this idea, we define the second order part of our loss function:

$$s(\boldsymbol{\lambda}, \xi_t) = \ell(\boldsymbol{\lambda}, \xi_t) - \langle \nabla \ell(\boldsymbol{\lambda}^*, \xi_t), \boldsymbol{\lambda} - \boldsymbol{\lambda}^* \rangle - \ell(\boldsymbol{\lambda}^*, \xi_t), \ \bar{s}(\boldsymbol{\lambda}) = \frac{1}{T} \sum_{t=1}^T s(\boldsymbol{\lambda}, \xi_t)$$

with its population version $S(\boldsymbol{\lambda}) = \mathbb{E} s(\boldsymbol{\lambda}, \xi_t) = \mathfrak{L}(\boldsymbol{\lambda}) - \langle \nabla \mathfrak{L}(\boldsymbol{\lambda}), \boldsymbol{\lambda} - \boldsymbol{\lambda}^* \rangle - \mathfrak{L}(\boldsymbol{\lambda}^*)$. Define localized Rademacher complexity of s within a small neighbourhood of $\boldsymbol{\lambda}^*$ as

$$\mathcal{R}_\varepsilon = \mathbb{E}_\xi \mathbb{E}_\sigma \left[\sup_{\boldsymbol{\lambda} \in \mathbb{B}(\boldsymbol{\lambda}^*, \varepsilon)} \frac{1}{T} \sum_{t=1}^T \sigma_t s(\boldsymbol{\lambda}, \xi_t) \right],$$

where σ_t are independent Rademacher random variables. We have the following result:

PROPOSITION 5. Under Assumption 6-8, the following inequality holds

$$\mathcal{R}_\varepsilon \leq \sqrt{2m \log(3K)} \frac{2L\varepsilon}{\sqrt{T}} + \frac{M\varepsilon^2}{K},$$

for any constant $K > 0$. Consequently, if $\varepsilon \geq \frac{64\sqrt{2}L}{\sqrt{T}\underline{\mathcal{L}}_\ell} \sqrt{\log \frac{100M}{\underline{\mathcal{L}}_\ell}}$, we have the following probabilistic bound:

$$\mathbb{P}\left(\left\| \widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^* \right\| \geq \varepsilon \right) \leq m \exp\left(-\frac{T\underline{\mathcal{L}}_\ell^2 \varepsilon^2}{8L^2 m} \right) + \exp\left(-\frac{T\underline{\mathcal{L}}_\ell^2 \varepsilon^2}{5000L^2} \right)$$

D efine $\mathcal{K}$ as a $\frac{\varepsilon}{K}$-cover of the set $\boldsymbol{\lambda} \in \mathbb{B}(\boldsymbol{\lambda}^*, \varepsilon)$, then by the covering number of a ball, it is valid that $\log |\mathcal{K}| \leq m \log(3K)$. Define a projection $\mathcal{K}(\boldsymbol{\lambda})$ that project each $\boldsymbol{\lambda} \in \mathbb{B}(\boldsymbol{\lambda}^*, \varepsilon)$ onto the closest element in the cover $\mathcal{K}$. By Assumption 6, we have a uniform bound $|s(\boldsymbol{\lambda}, \xi_t)| \leq 2L\varepsilon$. Then it follows that:

$$\begin{aligned}\mathcal{R}_\varepsilon =&\mathbb{E}_\xi\mathbb{E}_\sigma\left[\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon)}\frac{1}{T}\sum_{t=1}^T\sigma_t s(\boldsymbol{\lambda},\xi_t)\right]\\ \le&\mathbb{E}_\xi\left[\mathbb{E}_\sigma\sup_{\boldsymbol{\lambda}\in\mathcal{K}}\frac{1}{T}\sum_{t=1}^T\sigma_t s(\boldsymbol{\lambda},\xi_t)+\mathbb{E}_\sigma\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon),\boldsymbol{\lambda}'=\mathcal{K}(\boldsymbol{\lambda})}\frac{1}{T}\sum_{t=1}^T\sigma_t(s(\boldsymbol{\lambda},\xi_t)-s(\boldsymbol{\lambda}',\xi_t))\right]\\ \le&\sqrt{2m\log(3K)}\frac{2L\varepsilon}{\sqrt{T}}+\mathbb{E}_\xi\left[E_\sigma\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon),\boldsymbol{\lambda}'=\mathcal{K}(\boldsymbol{\lambda})}\frac{1}{T}\sum_{t=1}^T\sigma_t(s(\boldsymbol{\lambda},\xi_t)-s(\boldsymbol{\lambda}',\xi_t))\right],\end{aligned}$$

where the second inequality is by Massart's finite class lemma. We focus on controlling the second term by computing the first order moment of $\left|s(\boldsymbol{\lambda},\xi_t)-s(\boldsymbol{\lambda}',\xi_t)\right|$:

$$\begin{aligned}&\mathbb{E}\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon),\boldsymbol{\lambda}'=\mathcal{K}(\boldsymbol{\lambda})}\left|s(\boldsymbol{\lambda},\xi_t)-s(\boldsymbol{\lambda}',\xi_t)\right|=\mathbb{E}\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon),\boldsymbol{\lambda}'=\mathcal{K}(\boldsymbol{\lambda})}\left|\ell(\boldsymbol{\lambda},\xi_t)-\ell(\boldsymbol{\lambda}',\xi_t)-\langle\nabla\ell(\boldsymbol{\lambda}^*,\xi_t),\boldsymbol{\lambda}-\boldsymbol{\lambda}'\rangle\right|\\ &=\mathbb{E}\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon),\boldsymbol{\lambda}'=\mathcal{K}(\boldsymbol{\lambda})}\left(\int_0^1 v^\top(\nabla\ell(\boldsymbol{\lambda}'+vz,\xi_t)-\nabla\ell(\boldsymbol{\lambda}^*,\xi_t))dz\right),\ \text{where } v=\boldsymbol{\lambda}-\boldsymbol{\lambda}'\\ &\le\sup\|v\|\cdot\mathbb{E}\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon),\boldsymbol{\lambda}'=\mathcal{K}(\boldsymbol{\lambda})}\int_0^1\left\|\nabla\ell(\boldsymbol{\lambda}'+vz,\xi_t)-\nabla\ell(\boldsymbol{\lambda}^*,\xi_t)\right\|dz\\ &\le\frac{\varepsilon}{K}\int_0^1\mathbb{E}\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon),\boldsymbol{\lambda}'=\mathcal{K}(\boldsymbol{\lambda})}\left\|\nabla\ell(\boldsymbol{\lambda}'+vz,\xi_t)-\nabla\ell(\boldsymbol{\lambda}^*,\xi_t)\right\|dz\le\frac{M\varepsilon^2}{K},\end{aligned}$$

where the last inequality we use the Assumption 8. Then it follows that

$$\begin{aligned}&\mathbb{E}_\xi\left[E_\sigma\sup_{\boldsymbol{\lambda}\in\mathbb{B},\boldsymbol{\lambda}'=\mathcal{K}(\boldsymbol{\lambda})}\frac{1}{T}\sum_{t=1}^T\sigma_t(s(\boldsymbol{\lambda},\xi_t)-s(\boldsymbol{\lambda}',\xi_t))\right]\le\frac{\sum_{t=1}^T\mathbb{E}_\xi\sup_{\boldsymbol{\lambda}\in\mathbb{B},\boldsymbol{\lambda}'=\mathcal{K}(\boldsymbol{\lambda})}\left|s(\boldsymbol{\lambda},\xi_t)-s(\boldsymbol{\lambda}',\xi_t)\right|}{T}\\ &\le\frac{M\varepsilon^2}{K},\end{aligned}$$

which proves the first statement. We then prove the second statement by choosing a suitable K.

If $\widehat{\boldsymbol{\lambda}}$ which minimizes the empirical risk satisfies $\left\|\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^*\right\|\ge\varepsilon$, then by the convexity of $\bar{\ell}$, there exists a $\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon)$ such that $\bar{\ell}(\boldsymbol{\lambda})-\bar{\ell}(\boldsymbol{\lambda}^*)\le 0$. Together with the second order growth Assumption 7, we have

$$\begin{aligned}\bar{s}(\boldsymbol{\lambda})-S(\boldsymbol{\lambda})&=\bar{\ell}(\boldsymbol{\lambda})-\bar{\ell}(\boldsymbol{\lambda}^*)-(\mathfrak{L}(\boldsymbol{\lambda})-\mathfrak{L}(\boldsymbol{\lambda}^*))-\langle\nabla\bar{\ell}(\boldsymbol{\lambda}^*)-\nabla\mathfrak{L}(\boldsymbol{\lambda}^*),\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\rangle\\ &\le-\frac{\underline{\mathcal{L}}_\ell}{2}\varepsilon^2+\left\|\nabla\bar{\ell}(\boldsymbol{\lambda}^*)-\nabla\mathfrak{L}(\boldsymbol{\lambda}^*)\right\|\varepsilon.\end{aligned}$$

By Hoeffding's concentration inequality, we have

$$\mathbb{P}\left(\left\|\nabla\bar{\ell}(\boldsymbol{\lambda}^*)-\nabla\mathfrak{L}(\boldsymbol{\lambda}^*)\right\|\ge\frac{\underline{\mathcal{L}}_\ell}{4}\varepsilon\right)\le m\exp\left(-\frac{T\underline{\mathcal{L}}_\ell^2\varepsilon^2}{8L^2m}\right)$$

Define the event that inequality $\left\|\nabla\bar{\ell}(\boldsymbol{\lambda}^*)-\nabla\mathfrak{L}(\boldsymbol{\lambda}^*)\right\|\ge\frac{\underline{\mathcal{L}}_\ell}{4}\varepsilon$ holds as $\mathcal{E}_1$. Then under $\{\left\|\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^*\right\|\ge\varepsilon\}\cap\mathcal{E}_1^c$, we have

$$\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon)} |\bar{s}(\boldsymbol{\lambda}) - S(\boldsymbol{\lambda})| \geq \frac{\underline{\mathcal{L}}_\ell}{4}\varepsilon^2.$$

Choosing $K = \frac{32M}{\underline{\mathcal{L}}_\ell}$. When $\varepsilon \geq \frac{64\sqrt{2}L}{\sqrt{T}\underline{\mathcal{L}}_\ell}\sqrt{\log\frac{100M}{\underline{\mathcal{L}}_\ell}}$, we have

$$2\mathcal{R}_\varepsilon \leq \frac{\underline{\mathcal{L}}_\ell}{8}\varepsilon^2,$$

thus we have the following inequality:

$$\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon)} |\bar{s}(\boldsymbol{\lambda}) - S(\boldsymbol{\lambda})| \geq 2\mathcal{R}_\varepsilon + \frac{\underline{\mathcal{L}}_\ell}{8}\varepsilon^2.$$

By the convergence theory of empirical process (Koltchinskii 2011, Boucheron et al. 2005),

$$\mathbb{P}\left(\sup_{\boldsymbol{\lambda}\in\mathbb{B}(\boldsymbol{\lambda}^*,\varepsilon)} |\bar{s}(\boldsymbol{\lambda}) - S(\boldsymbol{\lambda})| \geq 2\mathcal{R}_\varepsilon + \frac{6L\varepsilon z}{\sqrt{T}}\right) \leq \exp\left(-\frac{z^2}{2}\right),$$

thus we conclude that $\mathbb{P}(\{\left\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*\right\| \geq \varepsilon\} \cap \mathcal{E}_1^c) \leq \exp\left(-\frac{T\underline{\mathcal{L}}_\ell^2\varepsilon^2}{5000L^2}\right)$, i.e.,

$$\mathbb{P}\left(\left\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*\right\| \geq \varepsilon\right) \leq m\exp\left(-\frac{T\underline{\mathcal{L}}_\ell^2\varepsilon^2}{8L^2m}\right) + \exp\left(-\frac{T\underline{\mathcal{L}}_\ell^2\varepsilon^2}{5000L^2}\right).$$

THEOREM 8. *The following bound holds for the convergence rate of* $\widehat{\boldsymbol{\lambda}}$*:*

$$\mathbb{E}\left\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*\right\|^2 \leq \left(\frac{512L^2}{\underline{\mathcal{L}}_\ell^2}\log\frac{100M}{\underline{\mathcal{L}}_\ell} + \frac{8m^2L^2 + 5000L^2}{\underline{\mathcal{L}}_\ell^2}\right)\frac{1}{T}$$

T his is a direct consequence of Proposition 5. By the integral formula of expectation, we have

$$\begin{aligned}
&\mathbb{E}\left\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*\right\|^2 = \int_0^\infty \mathbb{P}(\left\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*\right\| \geq \sqrt{z})dz \\
&\leq \left(\frac{512L^2}{\underline{\mathcal{L}}_\ell^2}\log\frac{100M}{\underline{\mathcal{L}}_\ell}\right)\frac{1}{T} + \int_c^\infty \mathbb{P}(\left\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*\right\|^2 \geq z)dz, \text{ where } c = \left(\frac{512L^2}{\underline{\mathcal{L}}_\ell^2}\log\frac{100M}{\underline{\mathcal{L}}_\ell}\right)\frac{1}{T} \\
&\leq \left(\frac{512L^2}{\underline{\mathcal{L}}_\ell^2}\log\frac{100M}{\underline{\mathcal{L}}_\ell}\right)\frac{1}{T} + \int_0^\infty m\exp\left(-\frac{T\underline{\mathcal{L}}_\ell^2 z}{8L^2m}\right) + \exp\left(-\frac{T\underline{\mathcal{L}}_\ell^2 z}{5000L^2}\right)dz \\
&\leq \left(\frac{512L^2}{\underline{\mathcal{L}}_\ell^2}\log\frac{100M}{\underline{\mathcal{L}}_\ell} + \frac{8m^2L^2 + 5000L^2}{\underline{\mathcal{L}}_\ell^2}\right)\frac{1}{T},
\end{aligned}$$

which finishes the proof.

By simply equating $\ell(\boldsymbol{\lambda}, \xi_t)$ with Fenchel conjugate $f_t^*(\boldsymbol{\lambda})$ in online convex allocation, we are able to prove the optimal dual convergence rate $O(\frac{1}{T})$. Notice that, here our Assumption 4 is equivalent to the Assumption 8 we used in the proof.

10.3. Proof of Proposition 2

For any given $\varepsilon > 0$, we define the neighbourhood of ν^* for given ε as

$$\Omega_\nu(\varepsilon) := \{\nu : \|\nu - \nu^*\|_\infty \le 4H\varepsilon\}.$$

We then construct a good event $\mathcal{E}(\varepsilon)$ with pro only depends on ε that under this good event, the convex function $\bar{s}_T(\nu, d)$ is larger than a quadratic function in $\Omega_\nu(\varepsilon)$, 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. We first split $\Omega_\nu(\varepsilon)$ into multiple cubes layer by layer and in each single cube, we control the difference of second order terms between all the ν in the cube and the central point of the cube.
2. Then we uniformly control the deviation of second order terms for all central points.

We now discuss the second order term $\bar{s}_T(\nu, d)$ defined in (3.1). To derive an uniform lower bound of $\bar{s}_T(\nu, d)$, we do the following split on $\Omega_{\boldsymbol{\nu}}(\varepsilon)$ according to Huber (1967).

Define set $\Omega_\nu^k(\varepsilon) = \left\{\nu \middle| \|\nu - \nu^*\|_\infty \le q^k 4H\varepsilon, \|\mu - \mu^*\|_\infty \le q^k 4H\varepsilon\right\}, 0 \le k \le N$, where $q \in (0,1)$ and $N \in \mathbb{N}_+$ will be identified later. This split divides $\Omega_\nu(\varepsilon)$ into N layers $\{\Omega_\nu^{k-1}(\varepsilon) \backslash \Omega_\nu^k(\varepsilon)\}_{k=1}^N$ and a center cube $\Omega_\nu^N(\varepsilon)$. We then split each layer into disjoint cubes $\{\bar{\Omega}^{kl}(\varepsilon)\}_{l=1}^{l_k}$ with edges of length $(1-q)q^{k-1}4H\varepsilon$, and denote the center cube by $\bar{\Omega}^{N1}(\varepsilon)$. Huber (1967) shows that there are at most $(2N)^m$ cubes. This split is not unique to get the desired convergence order but it makes our result tighter. The center of each cube $\bar{\Omega}^{kl}(\varepsilon)$ is defined as ν_{kl}. Define $\bar{\nu}_{kl} = \arg\max_{\nu \in \bar{\Omega}^{kl}(\varepsilon)} \|\nu - \nu^*\|_2$, and

$$\Gamma_t^{kl} = \max_{\nu \in \bar{\Omega}^{kl}(\varepsilon)} [s_t(\nu_{kl}, d) - s_t(\nu, d)] \tag{10.2}$$

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

$$\begin{aligned}
\bar{s}_T(\nu, d) =& \frac{1}{T}\sum_{t=1}^T s_t(\nu, d) - \frac{1}{T}\sum_{t=1}^T s_t(\nu_{kl}, d) + \frac{1}{T}\sum_{t=1}^T s_t(\nu_{kl}, d) \\
\ge& \underbrace{\mathbb{E}s_t(\nu_{kl}, d) - \mathbb{E}\Gamma_t^{kl}}_{10.3.1} + \underbrace{-\frac{1}{T}\sum_{t=1}^T \Gamma_t^{kl} + \mathbb{E}\Gamma_t^{kl}}_{10.3.2} + \underbrace{\frac{1}{T}\sum_{t=1}^T s_t(\nu_{kl}, d) - \mathbb{E}s_t(\nu_{kl}, d)}_{10.3.3}
\end{aligned} \tag{10.3}$$

We study the lower bounds of these 3 terms in (10.3) respectively.

Lower bound of 10.3.1:

$$\begin{aligned}
\mathbb{E}\Gamma_t^{kl} =& \mathbb{E}\max_{\nu \in \bar{\Omega}^{kl}(\varepsilon)} \left[f_t^*(b_t^\top(\lambda_{kl} + \mu_{kl})) - f_t^*(b_t^\top(\lambda + \mu)) - \nabla f_t^*(\lambda^* + \mu^*)^\top b_t^\top(\lambda_{kl} + \mu_{kl} - \lambda - \mu)\right] \\
=& \mathbb{E}\max_{\nu \in \bar{\Omega}^{kl}(\varepsilon)} \left[\int_0^1 v_1^\top(\nu)\left[\nabla f_t^*(b_t^\top(\lambda + \mu) + v_1 \cdot z) - \nabla f_t^*(b_t^\top(\lambda^* + \mu^*))\right] dz\right] \\
\le& \bar{b} \max_{\nu \in \bar{\Omega}^{kl}(\varepsilon)} \|\nu - \nu_{kl}\|_2 \cdot \left[\int_0^1 \mathbb{E}\max_{\nu \in \bar{\Omega}^{kl}(\varepsilon)} \left\|\nabla f_t^*(b_t^\top(\nu) + v_1 \cdot z) - \nabla f_t^*(b_t^\top(\nu^*))\right\| dz\right] \\
\le& L_1 \bar{b}^2 (\max_{\nu \in \bar{\Omega}^{kl}(\varepsilon)} \|\nu_{kl} - \nu\|_2) \cdot \|\bar{\nu}_{kl} - \nu^*\|_2
\end{aligned} \tag{10.4}$$

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

According to Proposition 4, we have

$$\mathbb{E}s_t(\nu_{kl}, d) \geq \underline{\mathcal{L}}_s \left\|\nu_{kl} - \nu^*\right\|_2^2$$

So for the first term, it is clear that

$$-\mathbb{E}\Gamma_t^{kl} + \mathbb{E}s_t(\nu_{kl}, d) \geq \underline{\mathcal{L}}_s(\left\|\nu_{kl} - \nu^*\right\|_2^2 - L_1\bar{b}^2(\max_{\nu\in\bar{\Omega}^{kl}(\varepsilon)} \left\|\nu_{kl} - \nu\right\|_2) \cdot \left\|\bar{\nu}_{kl} - \nu^*\right\|_2 \tag{10.5}$$

Lower bound of 10.3.2: Since the gradients $\left\|\nabla f_t^*\right\|_\infty$ is bounded by D, by the integral form of Γ_{kl} in the second equality of 10.4, we also have:

$$\left\|\Gamma_t^{kl}\right\|_2 \leq 2\sqrt{n}\bar{b}D \max_{\nu\in\bar{\Omega}^{kl}(\varepsilon)} \left\|\nu - \nu_{kl}\right\|_2,$$

for any $t \in [T]$. Define event

$$\mathcal{E}_{kl,1}(\varepsilon_1) = \left\{ -\frac{1}{T}\sum_{t=1}^{T} \Gamma_t^{kl} + \mathbb{E}\Gamma_t^{kl} < -2\varepsilon_1\sqrt{n}\bar{b}D \max_{\nu\in\bar{\Omega}^{kl}(\varepsilon)} \left\|\nu - \nu_{kl}\right\|_2 \right\}. \tag{10.6}$$

Then according to Hoeffding's inequality, $\mathbb{P}(\mathcal{E}_{kl,1}(\varepsilon_1)) \leq \exp(-\frac{T\varepsilon_1^2}{2})$

Lower bound of 10.3.3: We calculate the norm of each $s_t(\nu_{kl}, d)$:

$$\begin{aligned} \left\|s_t(\nu_{kl}, d)\right\|_2 =& \left\| \left[\int_0^1 v_2^\top \left[\nabla f_t^*(b_t(\lambda^* + \mu^*) + v_2 \cdot z) - \nabla f_t^*(\lambda^* + \mu^*)\right] dz \right] dz \right\|_2 \\ \leq& 2\sqrt{n}\bar{b}D \left\|\bar{\nu}_{kl} - \nu^*\right\|_2, \end{aligned} \tag{10.7}$$

for any $t \in [T]$, where $v_2 = b_t^\top(\nu_{kl} - \nu^*)$ is the direction vector. Define event

$$\mathcal{E}_{kl,2}(\varepsilon_2) = \left\{ \frac{1}{T}\sum_{t=1}^{T} s_t(\nu_{kl}, d) - \mathbb{E}s_t(\nu_{kl}, d) < -2\varepsilon_2\sqrt{n}\bar{b}D \left\|\bar{\nu}_{kl} - \nu^*\right\|_2 \right\}. \tag{10.8}$$

Then we have $\mathbb{P}(\mathcal{E}_{kl,2}) \leq \exp(-\frac{T\varepsilon_2^2}{2})$ 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

$$\max_{\nu\in\bar{\Omega}^{kl}(\varepsilon)} \left\|\nu - \nu_{kl}\right\|_2 = \sqrt{m}(1-q)q^{k-1}4H\varepsilon,$$
$$\left\|\nu^* - \nu_{kl}\right\|_2 \geq q^k 4H\varepsilon.$$

And also

$$\left\|\nu^* - \bar{\nu}_{kl}\right\|_2 \leq \left\|\nu^* - \nu_{kl}\right\|_2 + \max_{\bar{\Omega}^{kl}(\varepsilon)} \left\|\nu - \nu_{kl}\right\|_2$$

$$\max_{\nu\in\bar{\Omega}^{kl}(\varepsilon)} \left\|\nu - \nu_{kl}\right\|_2 \leq \frac{\sqrt{m}(1-q)}{q} \left\|\nu^* - \nu_{kl}\right\|_2 \leq \frac{\sqrt{m}(1-q)}{q} \left\|\nu^* - \bar{\nu}_{kl}\right\|_2.$$

Thus we have the following result for the 10.3.1 term in (10.5).

$$
\begin{aligned}
-\mathbb{E}\Gamma_t^{kl}+\mathbb{E}s_t(\nu_{kl},d) \geq & \underline{\mathcal{L}}_s(\|\nu_{kl}-\nu^*\|_2^2 - L_1\bar{b}^2(\max_{\nu\in\bar{\Omega}^{kl}(\varepsilon)}\|\nu_{kl}-\nu\|_2)\cdot\|\bar{\nu}_{kl}-\nu^*\|_2 \\
\geq & \frac{\underline{\mathcal{L}}_s}{\left(1+\frac{\sqrt{m}(1-q)}{q}\right)^2}\|\bar{\nu}_{kl}-\nu^*\|_2^2 - \frac{\sqrt{m}(1-q)}{q}\cdot L_1\bar{b}^2\|\bar{\nu}_{kl}-\nu^*\|_2^2
\end{aligned}
$$

So there exists $\underline{q}=\frac{\sqrt{m}}{\sqrt{m}+1\wedge\frac{\underline{\mathcal{L}}_s}{4L_1\bar{b}^2}}$ such that when $q\geq\underline{q}$, $\frac{\sqrt{m}(1-q)}{q}\leq 1\wedge\frac{\underline{\mathcal{L}}_s}{4L_1\bar{b}^2}$, and

$$
\frac{\underline{\mathcal{L}}_s}{(1+\frac{\sqrt{m}(1-q)}{q})^2} - \frac{\sqrt{m}(1-q)}{q}\cdot L_1\bar{b}^2 \geq \underline{\mathcal{L}}_s/2.
$$

Choose $q=\underline{q}\vee\frac{1}{2}$. Then for the 10.3.1 term in (10.5) we have

$$
-\mathbb{E}\Gamma_t^{kl}+\mathbb{E}s_t(\nu_{kl},d) \geq \frac{\underline{\mathcal{L}}_s}{2}\|\bar{\nu}_{kl}-\nu^*\|_2^2. \tag{10.9}
$$

Let $\varepsilon_1=\varepsilon_2=\sqrt{m\log m}\varepsilon$. For 10.3.2, under event $\mathcal{E}_{kl,1}^c(\varepsilon_1)$ in (10.6) we have

$$
\begin{aligned}
-\frac{1}{T}\sum_{t=1}^T\Gamma_t^{kl}+\mathbb{E}\Gamma_t^{kl} \geq & -2\varepsilon\sqrt{nm\log m}\bar{b}D(\max_{\nu\in\bar{\Omega}^{kl}(\varepsilon)}\|\nu-\nu^*\|_2) \\
\geq & -2\varepsilon\sqrt{nm\log m}\bar{b}D\frac{\sqrt{m}(1-q)}{q}\|\bar{\nu}_{kl}-\nu^*\|_2.
\end{aligned} \tag{10.10}
$$

For 10.3.3, under event $\mathcal{E}_{kl,2}^c(\varepsilon)$ in (10.8) we have

$$
\frac{1}{T}\sum_{t=1}^T s_t(\nu_{kl},d)-\mathbb{E}s_t(\nu_{kl},d) \geq -2\varepsilon\sqrt{nm\log m}\bar{b}D\|\bar{\nu}_{kl}-\nu^*\|_2. \tag{10.11}
$$

Now we combine second order lower bounds in (10.9), (10.10), (10.11) together under the desired good event

$$
\mathcal{E}(\varepsilon)=\cap_{k=1}^N\cap_l(\mathcal{E}_{kl,1}^c(\varepsilon)\cap\mathcal{E}_{kl,2}^c(\varepsilon)),
$$

where we choose N by setting the radius of $\bar{\Omega}^{N1}(\varepsilon)$: $\sqrt{m}q^N4H\varepsilon\leq 2H\varepsilon$, i.e.,

$$
N=\lceil\log_q(\frac{1}{2\sqrt{m}})\rceil\leq\frac{4L_1\bar{b}^2}{\underline{\mathcal{L}}_s}\sqrt{m}\log\sqrt{m}.
$$

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

$$
\begin{aligned}
\bar{s}_T(\nu,d) \geq & \frac{\underline{\mathcal{L}}_s}{2}\|\bar{\nu}_{kl}-\nu^*\|_2^2 - 2\varepsilon\sqrt{n}\bar{b}D(1+\frac{\sqrt{m}(1-q)}{q})\|\bar{\nu}_{kl}-\nu^*\|_2 \\
\geq & \frac{\underline{\mathcal{L}}_s}{2}\|\bar{\nu}_{kl}-\nu^*\|_2^2 - 4\varepsilon\sqrt{nm\log m}\bar{b}D\|\bar{\nu}_{kl}-\nu^*\|_2
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}(\varepsilon)) \geq & 1-\sum_{0\leq k\leq N-1,\ l}(\mathbb{P}(\mathcal{E}_{kl,1}(\varepsilon))+\mathbb{P}(\mathcal{E}_{kl,2}(\varepsilon))) \\
\geq & 1-2(2\lceil\log_q(\frac{1}{2\sqrt{m}})\rceil)^m\exp(-\frac{m\log mT\varepsilon^2}{2}) \geq 1-2\exp(-\frac{m\log m(T\varepsilon^2-1)}{4})
\end{aligned}
$$

The following Lemma can show the concentration of first order term:

LEMMA 6. *Under Assumptions 1-3, the concentration of the gradient in the first order term* $\bar{\phi}_{T,\nu}(\nu^*, d)$ *satisfies*

$$\mathbb{P}(\left\|\bar{\phi}_T(\nu^*, d) - \nabla D_\nu(\nu^*, d)\right\|_2 > \epsilon) \leq 2m \exp(-\frac{T\epsilon^2}{2m\sqrt{n}\bar{b}D}), \tag{10.12}$$

for any $\epsilon > 0$.

Proof: According to Hoeffding's inequality, we have

$$\mathbb{P}(\left|\left(\bar{\phi}_T(\nu^*, d)\right)_i - \left(\nabla D_\nu(\nu^*, d)\right)_i\right| > \epsilon/\sqrt{m}) \leq 2\exp(-\frac{T\epsilon^2}{2m\sqrt{n}\bar{b}D})$$

for $\forall i \in [m]$. Combining all m dimensions together we conclude that

$$\mathbb{P}(\left\|\frac{1}{T}\sum_{t=1}^{T}\phi_t(\boldsymbol{\lambda}^*, d) - \nabla D(\boldsymbol{\lambda}^*, d)\right\| > \varepsilon) \leq 2m\exp(-\frac{T\varepsilon^2}{2m\sqrt{n}\bar{b}D}).$$

For the first order term, denote event $\mathcal{E}_0(\varepsilon_0) = \{\left\|\bar{\phi}_T(\boldsymbol{\lambda}, d) - \nabla D(\boldsymbol{\lambda}^*, d)\right\|_2 > \varepsilon_0\}$. Take $\varepsilon_0 = \varepsilon\sqrt{nm\log m}\bar{b}D$ Then by Lemma 6, we have

$$\mathbb{P}(\mathcal{E}_0(\varepsilon_0)) \leq 2m\exp(-\frac{T\sqrt{n}\log m\bar{b}D\varepsilon^2}{2}).$$

Under event $\mathcal{E}_0^c(\varepsilon) \cap \mathcal{E}(\varepsilon)$, we have

$$\begin{aligned}
\bar{D}_T(\boldsymbol{\lambda}, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) &\geq \bar{s}_T(\nu, d) + \left\langle \bar{\phi}_{T,\nu}(\nu^*, d) - \nabla_\nu D(\boldsymbol{\lambda}^*, d), \nu - \nu^* \right\rangle \\
&\geq \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\left\|\nu' - \nu^*\right\|_2^2 - 5\varepsilon\sqrt{nm\log m}\bar{b}D\left\|\bar{\nu}_{kl} - \nu^*\right\|_2 \\
&= \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\left\|\nu' - \nu^*\right\|_2^2 - \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\cdot 2H\varepsilon\left\|\nu' - \nu^*\right\|_2,
\end{aligned} \tag{10.13}$$

where we define $H = 10\sqrt{nm\log m}\bar{b}D/(\sigma_{\min}\underline{\mathcal{L}}_f)$.

We now show how the first inequality leads to the probabilistic bound of $\left\|\nu - \nu^*\right\|_2$. By the definition of $\bar{s}_T(\nu, d)$, if the first inequality holds, an argument similar to (3.2) will lead to

$$\begin{aligned}
\bar{D}_T(\boldsymbol{\lambda}, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) &\geq \bar{s}_T(\nu, d) + \left\langle \bar{\phi}_T(\boldsymbol{\lambda}^*, d), \boldsymbol{\lambda} - \boldsymbol{\lambda}^* \right\rangle \\
&\geq \bar{s}_T(\nu, d) + \left\langle \bar{\phi}_{T,\nu}(\nu^*, d) - \nabla_\nu D(\boldsymbol{\lambda}^*, d), \nu - \nu^* \right\rangle \\
&\geq \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\left(\left\|\nu' - \nu^*\right\|_2^2 - 2H\varepsilon\left\|\nu' - \nu^*\right\|_2\right)
\end{aligned} \tag{10.14}$$

with probability at least $1 - 2m\exp(-\frac{T\sqrt{n}\log m\bar{b}D\varepsilon^2}{2})$, where we use the optimality of $\boldsymbol{\lambda}^*$ and concentration of gradient. If ν_T^* is part of the optimal solution, then we claim that, the dual optimal solution ν_T^* must have $\left\|\nu_T^* - \nu^*\right\|_2 \leq 2H\varepsilon$. Otherwise:

1. If has $2H\varepsilon < \left\|\nu_T^* - \nu^*\right\|_2 \leq 4H\varepsilon$, then there will be a $\nu_T^{*\prime}$ such that $\left\|\nu_T^{*\prime} - \nu^*\right\|_2 \geq \left\|\nu_T^* - \nu^*\right\|_2 > 2H\varepsilon$, and

$$\bar{D}_T(\boldsymbol{\lambda}_T^*, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) \geq \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\left(\left\|\nu_T^{*\prime} - \nu^*\right\|_2^2 - 2H\varepsilon\left\|\nu_T^{*\prime} - \nu^*\right\|_2\right) > 0,$$

which contradicts the optimality of $\boldsymbol{\lambda}_T^*$.

2. If $\|\nu_T^* - \nu^*\|_2 > 4H\varepsilon$, since $\bar{D}_T(\boldsymbol{\lambda}^*, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) = 0$ and $\bar{D}_T(\boldsymbol{\lambda}_T^*, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) \le 0$, by the convexity of $\bar{D}_T$ we can always find a $\tilde{\boldsymbol{\lambda}}$ such that $2H\varepsilon < \|\tilde{\nu} - \nu^*\|_2 \le 4H\varepsilon$ and $\bar{D}_T(\tilde{\boldsymbol{\lambda}}, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) \le 0$. However, according to (10.14), we have

$$\bar{D}_T(\tilde{\boldsymbol{\lambda}}, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) \ge \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\left(\|\tilde{\nu}' - \nu^*\|_2^2 - 2H\varepsilon\|\tilde{\nu}' - \nu^*\|_2\right) > 0,$$

which also ends up with a contradiction.

To better present our idea to readers, we draw a figure here. This clearly shows that when our empirical function is lower bounded by a quadratic function with $H\varepsilon$ as the axis of symmetry, we essentially have $\|\nu_T^* - \nu^*\|_2 \le 2H\varepsilon$.

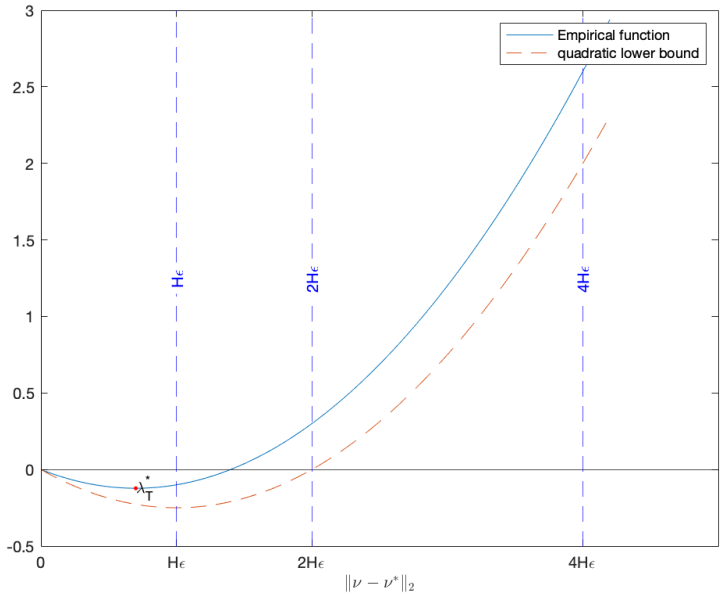


Figure 1 The value of $\bar{D}_T$ with respect to $\|\nu - \nu^*\|_2$. Since the quadratic lower bound has $H\varepsilon$ as the axis of symmetry, we have $\bar{D}_T(\boldsymbol{\lambda}, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) > 0$ for $\|\nu - \nu^*\|_2 > 2H\varepsilon$.

10.4. Proof of Theorem 1

By the tail expectation formula, for constant $H > 0$, we have

$$\mathbb{E}\|\nu_T^* - \nu^*\|_2^2 = 4H^2\int_0^\infty \mathbb{P}(\|\nu_T^* - \nu^*\|_2^2 > 4H^2 z)dz$$

According to the probabilistic bound in Proposition 2, for any $z > 0$,

$$\mathbb{P}(\|\nu_T^* - \nu^*\|_2^2 > 4H^2 z) \le 2m\exp\left(-\frac{T\sqrt{n}\log m\bar{b}D\varepsilon^2}{2}\right) \vee 1 + 2\exp\left(-\frac{m\log m(T\varepsilon^2 - 1)}{4}\right) \vee 1.$$

Then, calculating the integral, we get

$$\begin{aligned}
&\mathbb{E}(\|\nu_T^* - \nu^*\|_2^2) = H^2 \int_0^\infty \mathbb{P}(\|\nu_T^* - \nu^*\|_2^2 \geq 4H^2 z) dz \\
\leq &\frac{400nm\log m\bar{b}^2 D^2}{\sigma_{\min}^2 \underline{\mathcal{L}}_f^2} \int_{2/(T\sqrt{n}\bar{b}D)}^\infty \left[2\exp\left(-\frac{T\sqrt{n}\log m\bar{b}Dz}{2} + \log m\right) + \frac{2}{T\sqrt{n}\bar{b}D}\right] dz \\
&+ \frac{400nm\log m\bar{b}^2 D^2}{\sigma_{\min}^2 \underline{\mathcal{L}}_f^2} \int_{\frac{1}{T}}^\infty \left[2\exp\left(-\frac{m\log m(Tz-1)}{4}\right) + \frac{1}{T}\right] dz \\
\leq &C_1 \frac{\bar{b}^2 D^2}{\sigma_{\min}^2 \underline{\mathcal{L}}_f^2} \frac{nm\log m}{T}
\end{aligned}$$

For the optimality of the $O(T^{-1})$ order, let us consider a non-regularized case when $x \in [0,1]$ and $f_t(x) := f(x,\xi_t) := -(x-2\xi_t)^2/4 + \xi_t^2$, with the single constraint $d = 1/2$ and cost $b_t = 1$. The dual problem is

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

Let ξ_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 $\lambda \in [1/4,\ 1/2]$, $\bar{D}_T(\lambda) := \lambda^2 - 2(\bar{\xi_T} - 1/4)\lambda + \bar{\xi_T^2}$ with the optimal solution being $\lambda_T^* := \bar{\xi_T} - 1/4$. We have $\mathbb{E}(\lambda_T^* - \lambda^*)^2 \geq \mathrm{Var}(\bar{\xi_T}) = \sigma_\xi^2/T$. This shows that our $O(T^{-1})$ dual convergence rate is indeed optimal.

10.5. Proof of Theorem 2

Recall that, by the proof of Proposition 2, the convex function $\bar{D}_T$ is larger than a quadratic function in a neighborhood of ν^* with a high probability claimed there. Then, for any ϵ satisfying $\epsilon < 2H^2\varepsilon^2\sigma_{\min}\underline{\mathcal{L}}_f$, with the same high probability, the ϵ-optimal solution must belong to $\Omega_\nu(\varepsilon)$, because, for all the points in the border $\|\nu - \nu^*\|_2 = 4H\varepsilon$, we already have $\bar{D}_T(\boldsymbol{\lambda}, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) \geq 2H^2\varepsilon^2\sigma_{\min}\underline{\mathcal{L}}_f$. Then, with the same high probability, it follows that

$$\epsilon \geq \bar{D}_T(\boldsymbol{\lambda}_T^\epsilon, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) \geq \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4}\|\nu_T^{\epsilon\prime} - \nu^*\|_2^2 - \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{4} \cdot 2H\varepsilon\|\nu_T^{\epsilon\prime} - \nu^*\|_2,$$

which suggests that $\|\nu_T^\epsilon - \nu^*\|_2 \leq \|\nu_T^{\epsilon\prime} - \nu^*\|_2 \leq H\varepsilon + \left(H^2\varepsilon^2 + 4\epsilon/(\sigma_{\min}\underline{\mathcal{L}}_f)\right)^{1/2}$. Still, applying the tail expectation formula, we get

$$\begin{aligned}
\mathbb{E}(\|\nu_T^\epsilon - \nu^*\|_2^2) &= 4H^2 \int_0^{\frac{2\epsilon}{H^2\sigma_{\min}\underline{\mathcal{L}}_f}} \mathbb{P}(\|\nu_T^\epsilon - \nu^*\|_2 \geq 2H\sqrt{z}) dz \\
&\quad + 4H^2 \int_{\frac{2\epsilon}{H^2\sigma_{\min}\underline{\mathcal{L}}_f}}^\infty \mathbb{P}(\|\nu_T^\epsilon - \nu^*\|_2 \geq 2H\sqrt{z}) dz \\
&\leq \frac{8\epsilon}{\sigma_{\min}\underline{\mathcal{L}}_f} + 4H^2 \int_{\frac{\epsilon}{H^2\underline{\mathcal{L}}_D}}^\infty \mathbb{P}(\|\nu_T^\epsilon - \nu^*\|_2 \geq 2H\sqrt{z}) dz.
\end{aligned}$$

Let $2H\sqrt{z} = H\varepsilon + \sqrt{H^2\varepsilon^2 + \frac{4\epsilon}{\sigma_{\min}\underline{\mathcal{L}}_f}}$. When $z > \frac{\epsilon}{H^2\sigma_{\min}\underline{\mathcal{L}}_f}$, we have $\epsilon < 2H^2\varepsilon^2\sigma_{\min}\underline{\mathcal{L}}_f$, thus $\mathbb{P}(\|\nu_T^\epsilon - \nu^*\|_2 \geq 2H\sqrt{z})$ can be bounded by Proposition 2. Also when $2H\sqrt{z} = H\varepsilon + \sqrt{H^2\varepsilon^2 + \frac{2\epsilon}{\sigma_{\min}\underline{\mathcal{L}}_f}}$, we have $\varepsilon^2 \geq z - \frac{2\epsilon}{H^2\sigma_{\min}\underline{\mathcal{L}}_f}$. By the integral of z, we get the second part of the bound.

10.6. Proof of Corollary 1

Recall the proof of Theorem 2 that when ε satisfying $\epsilon < 2H^2\varepsilon^2\sigma_{\min}\underline{\mathcal{L}}_f$, with high probability the deterministic ϵ-optimal solution must be in $\Omega_\nu(\varepsilon)$. Similarly, for the stochastic ϵ-optimal solution, we try to confine it in a larger region so that with high probability $\mathbb{E}\left[\|\nu_T^\epsilon - \nu_T^*\|_2^2 \middle| \bar{D}_T\right]$ can still be bounded by ε. Notice that, although our Proposition 2 only focus on $\Omega_\nu(\varepsilon)$, it also bring us information outside $\Omega_\nu(\varepsilon)$. For any ε and ϵ, under the event when Proposition 2 holds, and any $\bar{D}_T$ we have:

1. If $\bar{D}_T(\boldsymbol{\lambda}_T^\epsilon, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) \leq 2H^2\varepsilon^2\sigma_{\min}\underline{\mathcal{L}}_f$, then $\|\nu_T^\epsilon - \nu_T^*\|_2 \leq 4H\varepsilon$.
2. If $\bar{D}_T(\boldsymbol{\lambda}_T^\epsilon, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) > 2H^2\varepsilon^2\sigma_{\min}\underline{\mathcal{L}}_f$, then we have $\|\nu_T^\epsilon - \nu_T^*\|_2 \leq \frac{2}{H\varepsilon\sigma_{\min}\underline{\mathcal{L}}_f}(\bar{D}_T(\boldsymbol{\lambda}_T^\epsilon, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d))$. Because the convex function $\bar{D}_T(\boldsymbol{\lambda}, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) = 0$ when $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, and when $\|\nu - \nu^*\|_2 = 4H\varepsilon$, $\bar{D}_T(\boldsymbol{\lambda}, d) - \bar{D}_T(\boldsymbol{\lambda}^*, d) \geq 2H^2\varepsilon^2\sigma_{\min}\underline{\mathcal{L}}_f$.

We conclude that under the event when Proposition 2 holds, for any $\epsilon < 2H^2\varepsilon^2\sigma_{\min}\underline{\mathcal{L}}_f$,

$$\mathbb{E}_{\mathcal{B}}\left[\|\nu_T^\epsilon - \nu^*\|_2^2 \middle| \bar{D}_T\right] \leq 16H^2\varepsilon^2 + \frac{4\sqrt{m\left(2\frac{\bar{f}+\bar{r}}{\underline{d}} + G\right)}}{H\varepsilon\sigma_{\min}\underline{\mathcal{L}}_f} \cdot \epsilon$$

because $\|\nu_T^\epsilon - \nu^*\|_2 \leq 4\sqrt{m\left(2\frac{\bar{f}+\bar{r}}{\underline{d}} + G\right)}$. The RHS term has a minimum value

$$z_0 = 3 \cdot 8\epsilon^{\frac{2}{3}}\left(m\left(2\frac{\bar{f}+\bar{r}}{\underline{d}} + G\right)\right)^{\frac{1}{3}} / (\sigma_{\min}\underline{\mathcal{L}}_f)^{\frac{2}{3}}$$

when $\varepsilon_0 = \epsilon^{\frac{1}{3}}\left(m\left(2\frac{\bar{f}+\bar{r}}{\underline{d}} + G\right)\right)^{\frac{1}{6}} / (2H(\sigma_{\min}\underline{\mathcal{L}}_f)^{\frac{1}{3}})$. When the RHS term is larger than is minimum value, we can always take the corresponding ε at the right side where $\varepsilon > \varepsilon_0$ and it follows that

$$z = 16H^2\varepsilon^2 + \frac{4\sqrt{m\left(2\frac{\bar{f}+\bar{r}}{\underline{d}} + G\right)}}{H\varepsilon\sigma_{\min}\underline{\mathcal{L}}_f} \cdot \epsilon \leq 48H^2\varepsilon^2.$$

Then by the tail expectation formula we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{B},\mathcal{P}}\|\nu_T^\epsilon - \nu^*\|_2^2 &= \int_0^{z_0} \mathbb{P}(\mathbb{E}_{\mathcal{B}}\left[\|\nu_T^\epsilon - \nu^*\|_2^2 \middle| \bar{D}_T\right] \geq z)dz + \int_{z_0}^{\infty} \mathbb{P}(\mathbb{E}_{\mathcal{B}}\left[\|\nu_T^\epsilon - \nu^*\|_2^2 \middle| \bar{D}_T\right] \geq z)dz \\
&\leq z_0 + \int_{z_0}^{\infty}\left[2m\exp\left(-\frac{T\sqrt{n}\log m\bar{b}Dz/(48H^2)}{2}\right) \vee 1\right]dz \\
&\quad + \int_{z_0}^{\infty}\left[2\exp\left(-\frac{m\log m(Tz/(48H^2) - 1)}{4}\right) \vee 1\right]dz. \\
&\leq z_0 + C_2\frac{\bar{b}^2D^2}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2}\frac{nm\log m}{T}.
\end{aligned}$$

10.7. Proof of Lemma 2

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

$$\begin{aligned}
R^*(\mathcal{P}) :=& \mathbb{E}_{\mathcal{P}}\left[\max_{x_t\in\mathcal{X}} \sum_{t=1}^{T} f_t(x_t) + T\cdot r(\frac{\sum_{t=1}^{T} b_t x_t}{T}),\ s.t.\ \sum_{t=1}^{T} b_t x_t \preceq dT\right] \\
\le& \mathbb{E}\sum_{t=1}^{T}\left[f_t(\tilde{x}_t(\nu^*)) + r(\tilde{a}(\mu^*)) + (\tilde{a}(\mu^*) - b_t\tilde{x}_t(\nu^*))^\top\mu^* + (d - b_t\tilde{x}_t(\nu^*))^\top\lambda^*\right] \\
=& T\cdot g(\nu^*)
\end{aligned}$$

10.8. Proof of Proposition 3

Since r is proper, by Fenchel conjugate, the definition of $\hat{\mu}_T$ implies

$$\begin{aligned}
r(\frac{\sum_{t=1}^{T} b_t x_t}{T}) + \hat{\mu}_T^\top\frac{\sum_{t=1}^{T} b_t x_t}{T} =& r^*(-\hat{\mu}_T) - \hat{\mu}_T^\top\mathbb{E}b_t\tilde{x}_t(\nu^*) + \hat{\mu}_T^\top\mathbb{E}b_t\tilde{x}_t(\nu^*) \\
\ge& r(\tilde{a}(\mu^*)) + \hat{\mu}_T^\top\tilde{a}(\mu^*)
\end{aligned}$$

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

$$R(A|\mathcal{P}) \ge \mathbb{E}\left[\sum_{t=1}^{T} f_t(x_t) + Tr(\tilde{a}(\mu^*)) + T\hat{\mu}_T^\top\tilde{a}(\mu^*) - T\hat{\mu}_T^\top\frac{\sum_{t=1}^{T} b_t x_t}{T}\right].$$

The Assumption 2 suggests that

$$\|\hat{\mu}_T\|_2 \le \sqrt{m}G,\ \text{and}\ \|a\|_2 = \|\nabla r^*(-\mu)\|_2 \le \sqrt{n}D\bar{b}.$$

Thus

$$\begin{aligned}
R(A|\mathcal{P}) \ge& \mathbb{E}\left[\sum_{t=1}^{T}[f_t(\tilde{x}_t(\nu_{t-1})) + r(\tilde{a}(\mu^*))] + \left\langle\hat{\mu}_T, \sum_{t=1}^{T}(\tilde{a}(\mu^*) - b_t x_t)\right\rangle\right] \\
=& \mathbb{E}\left[\sum_{t=1}^{T} g(\nu_{t-1}) + \left\langle\hat{\mu}_T - \mu^*, \sum_{t=1}^{T}(\tilde{a}(\mu^*) - b_t x_t)\right\rangle - \left\langle\lambda^*, \sum_{t=1}^{T}(d - b_t x_t)\right\rangle\right]
\end{aligned}$$

Combined with (5.2), we can show that

$$R^*(\mathcal{P}) - R(A|\mathcal{P}) \le \mathbb{E}\left[\sum_{t=1}^{T} g(\nu^*) - g(\nu_{t-1}) + \left\langle\lambda^*, \sum_{t=1}^{T}(d - b_t x_t)\right\rangle\right],$$

or, equivalent, in a two-phase form:

$$R^*(\mathcal{P}) - R(A|\mathcal{P}) \le \mathbb{E}\left[\sum_{t=1}^{\tau} g(\nu^*) - g(\nu_{t-1}) + \left\langle\lambda^*, \sum_{t=1}^{\tau}(d - b_t x_t)\right\rangle\right] + 2(\bar{f} + \bar{r} + \sqrt{mn}GD\bar{b})(T - \tau).$$

We conclude the proof.

10.9. Proof of Lemma 3

For technical convenience, we assume that, for each non-binding dimensions $i \in I_{\mathsf{NB}}$, 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. We start with a lemma on the continuity of dual optimal solution to prove Lemma 3.

LEMMA 7 **(Continuity of dual optimal solution)**. *Under Assumption 1, 2, 3, for the stochasitc program* $\min_{\mu,\lambda\succeq 0} D(\boldsymbol{\lambda}, d') = \mathbb{E} f_t^*(b_t^\top(\mu+\lambda)) + r^*(-\mu) + {d'}^\top \lambda$, *let* d' *be* $d'_1, d'_2 \in \Omega_d$ *separately, then the corresponding optimal solution* $\nu^*(d'_1), \nu^*(d'_2)$ *satisfies*

$$\|\nu^*(d'_1) - \nu^*(d'_2)\|_2^2 \le \frac{1}{\sigma_{\min}^2 \underline{\mathcal{L}}_f^2} \|d'_1 - d'_2\|_2^2 .$$

If further d'_1, d'_2 *identify the same binding/non-binding dimensions, then*

$$\|\nu^*(d'_1) - \nu^*(d'_2)\|_2^2 \le \frac{1}{\sigma_{\min}^2 \underline{\mathcal{L}}_f^2} \sum_{i\in I_B} (d'_{1i} - d'_{2i})^2,$$

where the binding dimension I_B *is with respect to* d'_1 *and* d'_2.

Proof of Lemma 7 By Proposition 4 and the uniform assumption on d, we have

$$D(\nu^*(d'_2), \mu^*(d'_1), d'_1) - D(\boldsymbol{\lambda}^*(d'_1), d'_1) \ge \frac{1}{2}\sigma_{\min}\underline{\mathcal{L}}_f \|\nu^*(d'_2) - \nu^*(d'_1)\|_2^2$$
$$D(\nu^*(d'_1), \mu^*(d'_2), d'_2) - D(\boldsymbol{\lambda}^*(d'_2), d'_2) \ge \frac{1}{2}\sigma_{\min}\underline{\mathcal{L}}_f \|\nu^*(d'_1) - \nu^*(d'_2)\|_2^2 .$$

Summing up two inequality we have

$$(d'_1 - d'_2)^\top (\nu^*(d'_2) - \nu^*(d'_1)) \ge \sigma_{\min}\underline{\mathcal{L}}_f \|\nu^*(d'_2) - \nu^*(d'_1)\|_2^2 , \tag{10.15}$$

or equivalently, $\sum_{i\in I_{\mathsf{B}}} (d'_{1i} - d'_{2i})(\nu_i^*(d'_2) - \nu_i^*(d'_1)) \ge \sigma_{\min}\underline{\mathcal{L}}_f \|\nu^*(d'_2) - \nu^*(d'_1)\|_2^2$ if further d'_1, d'_2 share the same binding/non-binding dimensions. From (10.15) we can show that

$$\sigma_{\min}\underline{\mathcal{L}}_f \|\nu^*(d'_2) - \nu^*(d'_1)\|_2^2 \le (d'_1 - d'_2)^\top (\nu^*(d'_2) - \nu^*(d'_1)) \le \|d'_1 - d'_2\|_2 \|\nu^*(d'_2) - \nu^*(d'_1)\|_2$$
$$\|\nu^*(d'_2) - \nu^*(d'_1)\|_2 \le \frac{1}{\sigma_{\min}\underline{\mathcal{L}}_f} \|d'_1 - d'_2\|_2 .$$

Thus we get the first statement. For the second statement, we focus on the binding dimensions:

$$\begin{aligned}
\sigma_{\min}\underline{\mathcal{L}}_f \|\nu^*(d'_2) - \nu^*(d'_1)\|_2^2 &\le \sum_{i\in I_{\mathsf{B}}} (d'_{1i} - d'_{2i})(\nu_i^*(d'_2) - \nu_i^*(d'_1)) \\
&\le \sqrt{\sum_{i\in I_{\mathsf{B}}} (d'_{1i} - d'_{2i})^2} \sqrt{\sum_{i\in I_{\mathsf{B}}} (\nu_i^*(d'_2) - \nu_i^*(d'_1))^2} \\
&\le \sqrt{\sum_{i\in I_{\mathsf{B}}} (d'_{1i} - d'_{2i})^2} \|\nu^*(d'_2) - \nu^*(d'_1)\|_2 ,
\end{aligned}$$

which completes the proof of Lemma 7.

Then, we return to Lemma 3 and consider the original constraints d and the its binding/non-binding dimensions: $I_{\mathsf{B}} = \{i | d_i - \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)_i = 0\}$, and $I_{\mathsf{NB}} = \{i | d_i - \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)_i > 0\}$. Here we write the corresponding optimal solution to $\min_{\mu,\lambda \succeq 0} D(\boldsymbol{\lambda}, d)$ as $\boldsymbol{\lambda}^*$, and write $\boldsymbol{\lambda}^*(d')$ if we change d to d'. Then if $i \in I_{\mathsf{B}}$ and i changes to non-binding dimensions for d', by Lemma 7 and Assumption 5.2, for any $\|d' - d\|_2 \le \delta_0 \wedge \frac{\underline{\lambda}}{2\bar{\mathcal{L}}_r}$ we have

$$\|d - d'\|_2 \ge \sigma_{\min}\underline{\mathcal{L}}_f \|\nu^*(d') - \nu^*\|_2 \ge \sigma_{\min}\underline{\mathcal{L}}_f \left(|\lambda_i^* - 0| - |\mu_i - \mu_i(d')|\right) \ge \frac{1}{2}\sigma_{\min}\underline{\mathcal{L}}_f \underline{\lambda}, \tag{10.16}$$

where $\underline{\lambda} = \min\{\lambda_i^* | i \in I_{\mathsf{B}}\}$. If on the other hand, $i \in I_{\mathsf{NB}}$ and i changes to binding dimensions for d', by Assumption 4, we have

$$\begin{aligned}
\mathbb{E}\|\nu^*(d') - \nu^*\|_2 &\ge \frac{1}{2\bar{b}^2 L_1} \left|\mathbb{E}\left(b_t \tilde{x}_t(\nu^*(d'))\right)_i - \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)_i\right| = \frac{1}{2\bar{b}^2 L_1} \left|d_i' - \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)_i\right| \\
&\ge \frac{1}{2\bar{b}^2 L_1} \left(\left|d_i - \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)_i\right| - |d_i' - d_i|\right).
\end{aligned}$$

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

$$\gamma = \min_{i \in I_{\mathsf{NB}}} \left\{d_i - \mathbb{E}\left(b_t \tilde{x}_t(\nu^*)\right)_i\right\}.$$

By Lemma 7 we have

$$\|d - d'\|_2 \ge \sigma_{\min}\underline{\mathcal{L}}_f \mathbb{E}\|\nu^*(d') - \nu^*\|_2 \ge \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{2\bar{b}^2 L_1} \left(\gamma - |d_i' - d_i|\right) \ge \frac{\sigma_{\min}\underline{\mathcal{L}}_f}{2\bar{b}^2 L_1} \left(\gamma - \|d - d'\|_2\right),$$

i.e., $\|d - d'\|_2 \ge \frac{\gamma \sigma_{\min}\underline{\mathcal{L}}_f}{\sigma_{\min}\underline{\mathcal{L}}_f + 2\bar{b}^2 L_1}$. Combined with (10.16), taking

$$\delta_d = \frac{1}{\sqrt{m}} \cdot \left(\frac{\gamma \sigma_{\min}\underline{\mathcal{L}}_f}{\sigma_{\min}\underline{\mathcal{L}}_f + 2\bar{b}^2 L_1}\right) \wedge \left(\frac{1}{2}\sigma_{\min}\underline{\mathcal{L}}_f \underline{\lambda}\right) \wedge \delta_0 \wedge \frac{\underline{\lambda}}{2\bar{\mathcal{L}}_r},$$

we can conclude that when $|d_i - d_i'| \le \delta_d$, the binding/non-binding dimensions will never change. Moreover, enlarging the constraint in a non-binding dimension will never change this constraint to the binding dimension. So, for the non-binding dimensions, $d_i' - d_i$ can be any large. This finishes the proof.

10.10. Proof of lemma 4

Proof of this lemma under frequent resolving is similar in spirit as that in Li and Ye (2021), but with different induction and dual convergence rate. Here we focus on the infrequent re-solving scheme in Algorithm 4. All the proof for the infrequent re-solving scheme is valid for the frequent resolving case. we define $d_t = d_{T_j}$ if $T_j \le t < T_{j+1}$. Then d_t will only update when $t = T_j$ for some j.

Without loss of generality, we assume $\frac{1}{\rho}$ is a integer and $T=\left(\frac{1}{\rho}\right)^K$ for some integer K. Denote the ratio $C_\rho=\rho/(1-\rho)$. Also assume $C_4=O(mn\log m)$ in Condition 1. Decomposing the perturbation of dual variables will lead to

$$\mathbb{E}\left[\sum_{t=1}^{\tau}\|\nu_{t-1}-\nu^*\|_2^2\right]\le 2\mathbb{E}\left[\sum_{t=1}^{\tau}\|\nu_{t-1}-\nu^*(d_{t-1})\|_2^2+\|\nu^*(d_{t-1})-\nu^*\|_2^2\right]$$

By the definition of stopping time τ and Condition 1, the first term in the RHS has

$$\mathbb{E}\left[\sum_{t=1}^{\tau}\|\nu_{t-1}-\nu^*(d_{t-1})\|_2^2\right]\le\sum_{j=1}^{J}2C_4\frac{1}{T_j}\cdot(T_{j+1}-T_j)\text{ or }C_4\frac{1}{T-T_j}\cdot(T_{j+1}-T_j)\le 2C_4(\log T+1).$$

Notice that, we need a good initialization: $\mathbb{E}\left\|\nu_0-\nu^*\right\|^2=O(1/T)$ to reach this condition for the first $(1-\rho)T$ terms in infrequent resolving case. For the second term, we apply lemma 7 to it.

$$2\mathbb{E}\left[\sum_{t=1}^{\tau}\|\nu^*(d_{t-1})-\nu^*\|_2^2\right]\le\frac{2}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2}\mathbb{E}\left[\sum_{t=1}^{\tau}\sum_{i\in I_{\mathsf{B}}}(d_{it}-d_i)^2\right].$$

Thus we transform the perturbation of $\nu^*(d_t)$ into the deviation of d_t in the binding dimensions.

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

$$d'_t=\begin{cases}d_t, & \text{if } t\le\tau\\ d_{t-1}, & \text{if } t>\tau\end{cases}$$

which shares the same stopping time with d_t and define $\tau_i=\min_t\{T-\left\lceil\frac{m\bar{b}}{\underline{d}}\right\rceil\}\cup\{t|d'_{it}\notin\mathcal{D}_i\}$ for $i\in[m]$ as the stopping time on each dimension with $\tau=\min\{\tau_1,...,\tau_m\}$. We follow a similar approach as Li and Ye (2021) to bound the stopping time. The distinction is that our analysis is more refined and we use a martingale argument that is different from Li and Ye (2021) to study the infrequent resolving scheme.

We first consider the binding dimensions. For any $i\in I_{\mathsf{B}}$, we derive:

$$\begin{aligned}
d'_{i,T_{j+1}}&=d'_{i,T_j}+\frac{\sum_{k=T_j+1}^{T_{j+1}}\left[d'_{i,T_j}-\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i\right]}{T-T_{j+1}}\mathbb{I}(\tau>T_j)\\
\mathbb{E}\left(d'_{i,T_{j+1}}-d_i\right)^2&=\mathbb{E}\left(d'_{i,T_j}-d_i\right)^2+\underbrace{\mathbb{E}\frac{\left(\sum_{k=T_j+1}^{T_{j+1}}\left[d'_{i,T_j}-\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i\right]\right)^2}{(T-T_{j+1})^2}\mathbb{I}(\tau>T_j)}_{A'}\\
&\quad+\underbrace{2\mathbb{E}\frac{\left(d'_{i,T_j}-d_i\right)\left(\sum_{k=T_j+1}^{T_{j+1}}\left[d'_{i,T_j}-\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i\right]\right)}{T-T_{j+1}}\mathbb{I}(\tau>T_j)}_{B'}\\
&\quad+\underbrace{2\mathbb{E}\frac{\left(d'_{i,T_j}-d_i\right)\left(\sum_{k=T_j+1}^{T_{j+1}}\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i-\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i\right)}{T-T_{j+1}}\mathbb{I}(\tau>T_j)}_{C'}
\end{aligned}\tag{10.17}$$

For the term A' we have

$$\begin{aligned}
A' &= \mathbb{E}\frac{\left(\sum_{k=T_j+1}^{T_{j+1}}\left[d'_{i,T_j}-\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i|\mathcal{H}_{T_j}\right]+\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i|\mathcal{H}_{T_j}\right]-\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i\right]\right)^2}{(T-T_{j+1})^2}\mathbb{I}(\tau>T_j)\\
&\le 2\mathbb{E}\frac{\left(\sum_{k=T_j+1}^{T_{j+1}}d'_{i,T_j}-\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i|\mathcal{H}_{T_j}\right]\right)^2}{(T-T_{j+1})^2}\\
&\quad+2\mathbb{E}\frac{\left(\sum_{k=T_j+1}^{T_{j+1}}\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i|\mathcal{H}_{T_j}\right]-\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i\right)^2}{(T-T_{j+1})^2}\\
&\le 2\mathbb{E}\frac{\sum_{k=T_j+1}^{T_{j+1}}\left(\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i-\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i\middle|\mathcal{H}_{T_j}\right]\right)^2}{T-T_{j+1}}\\
&\le 2\bar{b}^4L_1^2\mathbb{E}\left\|\nu^*(d_{T_j})-\nu_{T_j}\right\|_2^2\\
&\le \frac{C_4\bar{b}^4L_1^2+4nD^2\bar{b}^2}{T-T_j}+\frac{C_4\bar{b}^4L_1^2}{T_j}
\end{aligned}$$

For the second inequality, since $i\in I_{\mathsf{B}}$ and $d_t\in\sigma(\mathcal{H}_t)$, conditioned on past history $\mathcal{H}_{T_j}$, we always have

$$d'_{i,T_j}-\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i\middle|\mathcal{H}_{T_j}\right]=0,\ \text{for any } k\ge T_j+1.$$

This also indicates that $B'=0$. For the third inequality, we use Assumption 4. For the term C', we apply Assumption 4 and Condition 1:

$$\begin{aligned}
C' &= 2\mathbb{E}\left[\mathbb{E}\frac{\left(d'_{i,T_j}-d_i\right)\left(\sum_{k=T_j+1}^{T_{j+1}}\left(b_k\tilde{x}_k(\nu_{T_j})\right)_i-\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i\right)}{T-T_{j+1}}\mathbb{I}(\tau>T_j)\middle|\mathcal{H}_{T_j}\right]\\
&\le 2\bar{b}^2L_1\mathbb{E}\frac{\left|d'_{i,T_j}-d_i\right|\sum_{k=T_j+1}^{T_{j+1}}\left\|\nu_{T_j}-\nu^*(d_{T_j})\right\|}{T-T_{j+1}}\\
&\le 2\sqrt{C_4}\bar{b}^2L_1\sqrt{\mathbb{E}\left(d'_{i,T_j}-d_i\right)^2}\sqrt{\frac{1}{T-T_j}+\frac{1}{T_j}}.
\end{aligned}$$

Here the first inequality is because of Assumption 4, and the second inequality is from Condition 1 and Cauchy inequality. Here in the derivation, we can treat $\{\nu_t\}$ as a new sequence generated by $\{d'_t\}$, which has the same value with the original one when $t\le\tau$, and takes $\nu_t=\mathcal{B}_t(\mathcal{H}_t,d'_t)$ when $t>\tau$. We then get the recurrence relation of $d'_{i,T_j}-d_i$:

$$\begin{aligned}
\mathbb{E}\left(d'_{i,T_{j+1}}-d_i\right)^2 \le& \mathbb{E}\left(d'_{i,T_j}-d_i\right)^2+\frac{C_4\bar{b}^4L_1^2+4nD^2\bar{b}^2}{T-T_j}+\frac{C_4\bar{b}^4L_1^2}{T_j}\\
&+2\sqrt{C_4}\bar{b}^2L_1\sqrt{\mathbb{E}\left(d'_{it}-d_i\right)^2}\sqrt{\frac{1}{T-T_j}+\frac{1}{T_j}}.
\end{aligned}$$

Since $d_0=d$, assigning $C_4=O(mn\log m)$ and by induction we have

$$\mathbb{E}\left(d'_{i,T_j}-d_i\right)^2\le C_5C_\rho mn\log mD^2\bar{b}^4L_1^2\frac{\rho^{-j}-1}{T}$$

So, we have

$$\begin{aligned}2\mathbb{E}\left[\sum_{t=1}^{\tau}\|\nu^*(d_{t-1})-\nu^*\|_2^2\right]\leq&2\mathbb{E}\left[\sum_{t=1}^{\tau}\sum_{i\in I_{\mathsf{B}}}(d_{i,t-1}-d_i)^2\right]\\ \leq&2m\mathbb{E}\sum_{j=1}^{J}(T_j-T_{j-1})\left[\left(d'_{i,T_j}-d_i\right)^2\right]\leq C_5C_\rho m^2n\log mD^2\bar{b}^4L_1^2\log T+C,\\ \text{and }\mathbb{E}\left[\sum_{t=1}^{\tau}\|\nu_{t-1}-\nu^*\|_2^2\right]\leq&\left(2C_2+\frac{mC_3}{\underline{\mathcal{L}}_D^2}\right)\log T+2C_2,\end{aligned}$$

Notice that, by Condition 1, C_4 can take as small as $O(mn\log m/(\sigma_{\min}^2\underline{\mathcal{L}}_f^2))$, which completes the proof that

$$\mathbb{E}\left[\sum_{t=1}^{\tau}\|\nu_t-\nu^*\|_2^2\right]\leq C\cdot C_\rho\frac{m^2n\log mD^2\bar{b}^4L_1^2\log T}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2}+C$$

In the following discussion, we will treat C_ρ as a constant and thus can be omitted. In the case of frequent resolving, we only need to substitute the recurrence relation of $d'_{it}-d_i$ with:

$$\mathbb{E}\left(d'_{i,t+1}-d_i\right)^2\leq\mathbb{E}\left(d'_{it}-d_i\right)^2+\frac{\left(\bar{d}+\sqrt{n}D\bar{b}\right)^2}{(T-t-1)^2}+\frac{2\sqrt{2C_2L_2}\bar{b}^2\sqrt{\frac{1}{t+1}+\frac{1}{T-t}}\sqrt{\mathbb{E}\left(d'_{it}-d_i\right)^2}}{T-t-1},$$

which can be derived by the same analysis as the infrequent resolving case. 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 $C_3=\left(2\cdot\left(\bar{d}+\sqrt{n}D\bar{b}\right)^2\vee\left(2\sqrt{2C_2L_2}\bar{b}^2\right)+1\right)^2$. The lemma still holds.

10.11. Proof of lemma 5

We prove this Lemma under an infrequent resolving setting. The frequent resolving can be handled analogously. Since $\tau=\min\{\tau_1,...,\tau_m\}$, we only need to show $\mathbb{E}(T-\tau_i)\leq C\log T$ for any i in binding dimensions and non-binding dimensions. By the definition of ρ, and τ, the algorithm will only hit the stopping time after the first update of d_t, i.e, $\tau>T_1$. For the binding dimensions, applying Chebyshev's inequality, we have

$$\begin{aligned}\mathbb{E}(T-\tau_i)\leq&\sum_{i=1}^{T}\mathbb{P}\left(\tau_i\leq t\right)\leq1+\frac{\sqrt{n}D\bar{b}}{\underline{d}}+\sum_{j=2}^{J}\mathbb{P}\left(\tau_i\leq T_j\right)\left(T_j-T_{j-1}\right)\\ \leq&1+\frac{\sqrt{n}D\bar{b}}{\underline{d}}+\sum_{j=2}^{J}\mathbb{P}\left(\left|d_{i,T_j}-d_i\right|\geq\delta_d\right)\left(T_j-T_{j-1}\right)\\ \leq&1+\frac{\sqrt{n}D\bar{b}}{\underline{d}}+\sum_{j=2}^{J}\left(\frac{\mathbb{E}\left(d'_{i,T_{j+1}}-d_i\right)^2}{\delta_d^2}\right)\left(\rho^{j-1}(1-\rho)T\right)\\ \leq&C+\frac{\sqrt{n}D\bar{b}}{\underline{d}}+\frac{C_5m^2nD^2\bar{b}^4L_1^2}{\delta_d^2}\log T\end{aligned}\tag{10.18}$$

For the non-binding dimensions, $\mathcal{D}$ ensures that binding/non-binding dimensions remain unchanged when $d' \in \mathcal{D}$. Then for $d' \in \mathcal{D}$, we define

$$\tilde{d}'_i = \begin{cases} d'_i, & \text{if } i \in I_{\mathsf{B}} \\ d_i - \delta_d, & \text{if } i \in I_{\mathsf{NB}} \end{cases}$$

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

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

Recall that $\tilde{x}_k(\cdot) \perp\!\!\!\perp \mathcal{H}_t$, thus $\mathbb{E}\left[\left(b_k \tilde{x}_k(\nu^*(d_t))\right)_i | \mathcal{H}_t\right] < d_i - \delta_d$ for any $k \geq t+1$, $i \in I_{\mathsf{NB}}$ and $d_t \in \mathcal{D}$. This implies that

$$\begin{aligned}
\mathbb{P}\left(\tau_i \leq T_j\right) \leq & \mathbb{P}\left(\sum_{t=1}^{t'} \left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i \geq t'\left(d_i - \delta_d\right) + T\delta_d \text{ for some } 1 \leq t' \leq T_j\right) \\
\leq & \mathbb{P}\left(\sum_{t=1}^{t'} \left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i - \mathbb{E}\left[\left(b_t \tilde{x}_t(\nu^*(d_{t-1}))_i | \mathcal{H}_{t-1}\right]\right] \geq T\delta_d \text{ for some } 1 \leq t' \leq T_j\right) \\
\leq & \mathbb{P}\left(\sum_{t=1}^{t'} \left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i - \mathbb{E}\left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i | \mathcal{H}_{t-1}\right]\right] + \right. \\
& \left. \sum_{t=1}^{t'} \left|\mathbb{E}\left[\left(b_t \tilde{x}_t(\nu^*(d_{t-1}))_i | \mathcal{H}_{t-1}\right] - \mathbb{E}\left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i | \mathcal{H}_{t-1}\right]\right| \geq T\delta_d \text{ for some } 1 \leq t' \leq T_j\right) \\
\leq & \mathbb{P}\left(\sum_{t=1}^{t'} \left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i - \mathbb{E}\left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i | \mathcal{H}_{t-1}\right]\right] \geq \frac{T\delta_d}{2} \text{ for some } 1 \leq t' \leq T_j\right) \\
& + \mathbb{P}\left(\sum_{t=1}^{t'} \left|\mathbb{E}\left[\left(b_t \tilde{x}_t(\nu^*(d_{t-1}))_i | \mathcal{H}_{t-1}\right] - \mathbb{E}\left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i | \mathcal{H}_{t-1}\right]\right| \geq \frac{T\delta_d}{2} \text{ for some } 1 \leq t' \leq T_j\right)
\end{aligned}$$

Since sequences in the last two lines are martingales/sub-martingales, we use Doob's martingale inequality and get the following derivation:

$$\begin{aligned}
\mathbb{P}\left(\tau_i \leq T_j\right) \leq & \frac{4}{T^2\delta_d^2} \sum_{t=1}^{T_j} \mathbb{E}\left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i - \mathbb{E}\left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i | \mathcal{H}_{t-1}\right]\right]^2 \\
& + \frac{4}{T^2\delta_d^2} \mathbb{E} \sum_{t=1}^{T_j} \left[\left|\mathbb{E}\left[\left(b_t \tilde{x}_t(\nu^*(d_{t-1}))_i | \mathcal{H}_{t-1}\right] - \mathbb{E}\left[\left(b_t \tilde{x}_t\left(\nu_{t-1}\right)\right)_i | \mathcal{H}_{t-1}\right]\right|\right]^2
\end{aligned} \tag{10.19}$$

By Assumption 4, we have

$$\begin{aligned}
\mathbb{P}\left(\tau_i \leq T_j\right) \leq & \frac{16n\bar{b}^2 D^2 T_j}{T^2\delta_d^2} + \frac{8L_1^2\bar{b}^4}{T^2\delta_d^2} \sum_{t=1}^{T_j} \mathbb{E}\left\|\nu_{t-1} - \nu^*(d_{t-1})\right\|_2^2 \\
\leq & \frac{16n\bar{b}^2 D^2 T_j}{T^2\delta_d^2} + \frac{C_5 m^2 n \log m D^2 \bar{b}^4 L_1^2 \log T}{T^2\delta_d^2},
\end{aligned}$$

where for $j=1$, we use the assumption on initialization. We now go back to calculate the $\mathbb{E}(T-\tau_i)$:

$$\begin{aligned}\mathbb{E}(T-\tau_i) \leq & 2+\frac{\sqrt{n}D\bar{b}}{\underline{d}}+\sum_{j=2}^{J}\mathbb{P}\left(\tau_i \leq T_j\right)\left(T_j-T_{j-1}\right) \\ \leq & C+\frac{\sqrt{n}D\bar{b}}{\underline{d}}+\sum_{j=2}^{J}\frac{16n\bar{b}^2D^2T_j(T_j-T_{j-1})}{T^2\delta_d^2}+\frac{C_5(T_j-T_{j-1})}{T^2\delta_d^2}(m^2n\log mD^2\bar{b}^4L_1^2\log T) \\ \leq & C+\frac{\sqrt{n}D\bar{b}}{\underline{d}}+\frac{C_5n\bar{b}^2D^2\log T}{\delta_d^2}+\frac{C_5m^2n\log m\bar{b}^4D^2L_1^2\log T}{\delta_d^2}\end{aligned} \tag{10.20}$$

Notice that, by Condition 1, C_4 can take as small as $O(m\log mn/(\sigma_{\min}^2\underline{\mathcal{L}}_f^2))$. Putting together (10.18) and (10.20) we conclude the proof of lemma 5:

$$\mathbb{E}(T-\tau) \leq \frac{Cm^2n\log m\bar{b}^4D^2L_1^2\log T}{\delta_d^2\sigma_{\min}^2\underline{\mathcal{L}}_f^2}$$

10.12. Proof of Theorem 3

Equipped with Lemma 4 and 5, we now continue sketching the proof of Theorem 3. By Proposition 3, it suffices to bound the two 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 $g(\nu)$ by

$$\begin{aligned}g(\nu) =& \mathbb{E}\left[f_t(\tilde{x}_t(\nu))+r(\tilde{a}(\mu))+(\tilde{a}(\mu)-b_t\tilde{x}_t(\nu))^\top\mu^*+(d-b_t\tilde{x}_t(\nu))^\top\lambda^*\right] \\ =& \mathbb{E}\left[f_t^*(b_t^\top(\lambda+\mu))+r^*(-\mu)\right]+\mathbb{E}(\mu^*-\mu)^\top(\tilde{a}(\mu)-b_t\tilde{x}_t(\nu))+\mathbb{E}(\lambda^*-\lambda)^\top(d-b_t\tilde{x}_t(\nu)) \\ =& \mathbb{E}\left[f_t^*(b_t^\top(\lambda+\mu))+r^*(-\mu)\right]-\mathbb{E}\left[\nabla f_t^*(b_t^\top(\lambda+\mu))^\top b_t^\top(\lambda+\mu-\lambda^*-\mu^*)\right. \\ & \left.-\nabla r^*(-\mu)^\top(\mu-\mu^*)+d^\top(\lambda-\lambda^*)\right] \\ =& s(\nu,d)-\langle\nabla_\nu D(\nu,\mu^*,d)-\nabla_\nu D(\nu^*,\mu^*,d),(\nu-\nu^*)\rangle\end{aligned}$$

By Assumption 2 and 3, we get

$$\begin{aligned}g(\nu^*)-g(\nu) =& s(\nu^*,d)-s(\nu,d)+\langle\nabla_\nu D(\nu,\mu^*,d)-\nabla_\nu D(\nu^*,\mu^*,d),(\nu-\nu^*)\rangle \\ & +\langle\nabla s(\nu,d)-\nabla s(\nu^*,d),\nu-\nu^*\rangle \leq \left(2\bar{\mathcal{L}}_s-\underline{\mathcal{L}}_s\right)\|\nu-\nu^*\|_2^2 \\ & =(\bar{b}^2\bar{\mathcal{L}}_f-\frac{1}{2}\sigma_{\min}\underline{\mathcal{L}}_f)\|\nu-\nu^*\|_2^2\end{aligned}$$

Then Lemma 4 gives rise to the following bound.

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

Step 2: bounding R.2. This term can be controlled by the definition of stopping time and Lemma 5.

$$\begin{aligned}
&\mathbb{E}\left[2(\bar{f}+\bar{r}+C_3)(T-\tau)+\left\langle \lambda^*, \sum_{t=1}^{\tau}(d-b_t x_t)\right\rangle\right]\\
&=\mathbb{E}\left[2(\bar{f}+\bar{r}+C_3)(T-\tau)+\langle \lambda^*, d_\tau(T-\tau)-d(T-\tau)\rangle\right]\\
&\le \mathbb{E}\left[2(\bar{f}+2\bar{r}+C_3)(T-\tau)+\sum_{i\in I_{\mathsf{B}}}\lambda_i^*(d_i+\delta_d)(T-\tau)\right]\\
&\le (2\bar{f}+2\bar{r}+2C_3+(\|d\|+\sqrt{m}\delta_d)\frac{2(\bar{f}+\bar{r})}{\underline{d}})\mathbb{E}(T-\tau)=O(\log T).
\end{aligned}$$

Thus we finish the proof.

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 8. *Suppose $i\in I_{NB}$. Under Assumptions 1-5, Algorithm 1 with selected dual optimizer $\{\mathcal{B}_t\}_{t\ge 1}$ satisfying Condition 1 and stopping time (5.5) ensures*

$$\mathbb{E}\left\|\hat{\mu}_{I_{NB},T}-\mu^*_{I_{NB}}\right\|_2^2\le O(\frac{\log T}{T}),\ \textit{and}\ \ \mathbb{E}\left\|\sum_{t=1}^{\tau}(\tilde{a}(\mu^*)-b_t x_t)\right\|_2^2\le O(T\log T).$$

The proof Lemma 8 exploits the local smoothness of r and $\tilde{x}_t$ with the help of the optimality of μ^*, i.e., $\tilde{a}(\mu^*)=\mathbb{E}b_t\tilde{x}_t(\nu^*)$. We first check the binding dimensions: for all the binding dimensions $i\in I_{\mathsf{B}}$, since $\mathbb{E}(b_t\tilde{x}_t(\nu^*))_i=d_i$, by the definition of stopping time, it can be shown that

$$\begin{aligned}
\mathbb{E}\left\langle \hat{\mu}_{I_{\mathsf{B}},T}-\mu^*_{I_{\mathsf{B}}}, \sum_{t=1}^{\tau}\tilde{a}(\mu^*)_{I_{\mathsf{B}}}-(b_t x_t)_{I_{\mathsf{B}}}\right\rangle &= \mathbb{E}\left\langle \hat{\mu}_{I_{\mathsf{B}},T}-\mu^*_{I_{\mathsf{B}}}, \sum_{t=1}^{\tau} d_{I_{\mathsf{B}}}-(b_t x_t)_{I_{\mathsf{B}}}\right\rangle\\
&\le \mathbb{E}\sqrt{mn}GD\bar{b}(d+\delta_d)(T-\tau)\\
&= O(\log T)
\end{aligned}\tag{10.21}$$

We then go back to the non-binding dimension $i\in I_{\mathsf{NB}}$ with the help of Lemma 8. By Cauchy–Schwarz inequality, we get

$$\mathbb{E}\left[\left\langle \hat{\mu}_{I_{\mathsf{NB}},T}-\mu^*_{I_{\mathsf{NB}}}, \sum_{t=1}^{\tau}\tilde{a}(\mu^*)_{I_{\mathsf{NB}}}-(b_t x_t)_{I_{\mathsf{NB}}}\right\rangle\right]\le\left(\mathbb{E}\left\|\hat{\mu}_{I_{\mathsf{NB}},T}-\mu^*_{I_{\mathsf{NB}}}\right\|_2^2\mathbb{E}\left\|\sum_{t=1}^{\tau}(\mathbb{E}b_t\tilde{x}_t(\nu^*)-b_t x_t)_{I_{\mathsf{NB}}}\right\|_2^2\right)^{1/2}.\tag{10.22}$$

Thus R.3 can be controlled by $\log T$. The proof is concluded. The constant factor is as large as

$$\mathring{C}\lesssim\frac{m^2 n\log m\bar{b}^6D^4L_1^2}{\delta_d^2\sigma_{\min}^2\underline{\mathcal{L}}_f^2}\cdot\left((\bar{f}+\bar{r}+\sqrt{mn}GD\bar{b})\vee\frac{n\bar{b}^2G^2+m\bar{\mathcal{L}}_r^2\delta_0^2}{\delta_0^2}\right)$$

from the proof of Lemma 8. Generally, taking $m\lesssim n$ we have the order of $\mathring{C}=O(m^2n^2\log m)$

10.13. Proof of Lemma 8

For the $\mathbb{E}\left\|\hat{\mu}_{I_{\mathsf{NB}},T}-\mu^*_{I_{\mathsf{NB}}}\right\|_2^2$, $i\in I_{\mathsf{NB}}$, the optimality of μ^* implies $\tilde{a}(\mu^*)=\mathbb{E}b_t\tilde{x}_t(\nu^*)$, thus by conjugate we have

$$\begin{aligned}
\mathbb{E}\left\|\hat{\mu}_{I_{\mathsf{NB}},T}-\mu^*_{I_{\mathsf{NB}}}\right\|_2^2 &= \mathbb{E}\left\|\nabla_{I_{\mathsf{NB}}} r(\tilde{a}(\mu^*)) - \nabla_{I_{\mathsf{NB}}} r(\frac{\sum_{t=1}^T b_t x_t}{T})\right\|_2^2 \\
&\le n\bar{b}^2G^2\,\mathbb{P}\left(\left\|\mathbb{E}b_t\tilde{x}_t(\nu^*)-\frac{\sum_{t=1}^T b_t x_t}{T}\right\|\ge\delta_0\right)+\mathbb{E}m\bar{\mathcal{L}}_r^2\left\|\tilde{a}(\mu^*)-\frac{\sum_{t=1}^T b_t x_t}{T}\right\|_2^2 \\
&\le \frac{n\bar{b}^2G^2+m\bar{\mathcal{L}}_r^2\delta_0^2}{\delta_0^2}\mathbb{E}\left\|\frac{\sum_{t=1}^T b_t x_t-\mathbb{E}b_t\tilde{x}_t(\nu^*)}{T}-\frac{\sum_{t=1}^\tau \mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})|\mathcal{H}_{t-1}\right]}{T}+\frac{\sum_{t=1}^\tau \mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})|\mathcal{H}_{t-1}\right]}{T}\right\|_2^2 \\
&\le 3C_0\,\mathbb{E}\left\|\frac{\sum_{t=1}^\tau b_t\tilde{x}_t(\nu_{t-1})-\mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})|\mathcal{H}_{t-1}\right]}{T}\right\|_2^2 \text{ (part 10.13.1)} \\
&+3C_0\,\mathbb{E}\left\|\frac{\sum_{t=1}^\tau \mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})|\mathcal{H}_{t-1}\right]-\mathbb{E}b_t\tilde{x}_t(\nu^*)}{T}\right\|_2^2 \text{ (part 10.13.2)} \\
&+3C_0\,\mathbb{E}\left\|\frac{\sum_{t=\tau+1}^T b_t x_t-\mathbb{E}b_t\tilde{x}_t(\nu^*)}{T}\right\|_2^2 \text{ (part 10.13.3).}
\end{aligned}$$

For the part 10.13.1, notice that this is a martingale series. Thus we have

$$\begin{aligned}
\mathbb{E}\left\|\frac{\sum_{t=1}^\tau b_t\tilde{x}_t(\nu_{t-1})-\mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})|\mathcal{H}_{t-1}\right]}{T}\right\|_2^2 &\le \frac{\sum_{t=1}^T \mathrm{Var}(b_t\tilde{x}_t(\nu_{t-1})-\mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})|\mathcal{H}_{t-1}\right])}{T^2} \\
&\le \frac{nD^2\bar{b}^2}{T}
\end{aligned}$$

For the part 10.13.2, applying Assumption 4 we can yield

$$\begin{aligned}
\mathbb{E}\left\|\frac{\sum_{t=1}^\tau \mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})|\mathcal{H}_{t-1}\right]-b_t\tilde{x}_t(\nu^*)}{T}\right\|_2^2 &\le \mathbb{E}\frac{\tau\sum_{t=1}^\tau\left\|\mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})|\mathcal{H}_{t-1}\right]-b_t\tilde{x}_t(\nu^*)\right\|_2^2}{T^2} \\
&\le \frac{\mathbb{E}\sum_{t=1}^T\left\|\mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})|\mathcal{H}_{t-1}\right]-b_t\tilde{x}_t(\nu^*)\right\|_2^2\mathbb{I}(t\le\tau)}{T} \\
&= \frac{\sum_{t=1}^T\mathbb{E}\left\|\left[\mathbb{E}\left[b_t\tilde{x}_t(\nu_{t-1})-b_t\tilde{x}_t(\nu^*)|\mathcal{H}_{t-1},b_t\right]\mathbb{I}(t\le\tau)\right]\right\|_2^2}{T} \\
&\overset{\text{(a)}}{\le} L_1^2\bar{b}^2\frac{\sum_{t=1}^T\mathbb{E}\left[\left\|\nu_{t-1}-\nu^*\right\|_2^2\mathbb{I}(t\le\tau)\right]}{T}=L_1^2\bar{b}^2\frac{\mathbb{E}\left[\sum_{t=1}^\tau\left\|\nu_{t-1}-\nu^*\right\|_2^2\right]}{T} \\
&\overset{\text{(b)}}{\le} \frac{L_1^2\bar{b}^2m^2n\log mD^2\bar{b}^4L_1^2\log T}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2T}
\end{aligned}$$

(a) is by Assumption 4 and the fact $\{t\le\tau\}\in\sigma(\mathcal{H}_{t-1})$, and $\tilde{x}_t(\cdot)\perp\!\!\!\perp\nu_{t-1}$. (b) is by Lemma 4.

For the part 10.13.3, since $\left\|b_tx_t-\mathbb{E}b_t\tilde{x}_t(\nu^*)\right\|_2^2\le nD^2\bar{b}^2$, by Lemma 5, we have

$$\begin{aligned}
\mathbb{E}\left\|\frac{\sum_{t=\tau+1}^T b_tx_t-\mathbb{E}b_t\tilde{x}_t(\nu^*)}{T}\right\|_2^2 &\le \frac{\mathbb{E}\left[(T-\tau)\sum_{t=\tau+1}^T\left\|b_tx_t-\mathbb{E}b_t\tilde{x}_t(\nu^*)\right\|^2\right]}{T^2} \\
&\le nD^2\bar{b}^2\frac{\mathbb{E}(T-\tau)}{T}\le\frac{Cm^2n^2\log m\bar{b}^6D^4L_1^2}{\delta_d^2\sigma_{\min}^2\underline{\mathcal{L}}_f^2}\frac{\log T}{T}
\end{aligned}$$

We then go back to control the next term $\mathbb{E}\left\|\sum_{t=1}^{\tau}(\tilde{a}(\mu^*)-b_t x_t)\right\|_2^2$. This can be bounded exactly by part 10.13.1 and 10.13.2. From the argument above, we show that $\mathbb{E}\left\|\sum_{t=1}^{\tau}(\tilde{a}(\mu^*)-b_t x_t)\right\|_2^2$ is controlled by $O(T\log T)$. Thus we finish the proof.

10.14. Proof of Theorem 4 and 5

We specify a non-regularized case where $f_t(x)=-\frac{1}{4}(x-2\xi_t)^2+\xi_t^2$, with fixed cost $b_t=1$, average resource capacity $d=\frac{1}{2}D$, and ξ_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

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

Suppose λ^* is the optimal solution to the deterministic problem $\min_{\lambda\ge0}D(\lambda)=\mathbb{E}D_t(\lambda)$. Without loss of generality, we assume that our dual variable λ is taken within $\left[\frac{1}{4}D,\frac{1}{2}D\right]$ since we know that $\lambda^*=\mathbb{E}\xi_t-\frac{1}{4}D=\frac{3}{8}D\in\left[\frac{1}{4}D,\frac{1}{2}D\right]$.

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

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:

$$\begin{aligned}\text{Regret}(A)=&R^*(\mathcal{P})-R(A|\mathcal{P})\\ =&\mathbb{E}\left[\max_{x_t\in[0,D]}\left\{\sum_{t=1}^T f_t(x_t)\text{ s.t. }\sum_{t=1}^T x_t\le\frac{1}{2}DT\right\}\right]-\mathbb{E}\left[\sum_{t=1}^T f_t(x_t)\right]\\ =&\mathbb{E}\left[\min_{\lambda\ge0}\left\{\sum_{t=1}^T D_t(\lambda)\right\}\right]-\mathbb{E}\left[\sum_{t=1}^T f_t(x_t)\right]\\ =&\mathbb{E}\left[\sum_{t=1}^T D_t(\lambda_T^*)\right]-\mathbb{E}\left[\sum_{t=1}^T f_t(x_t)\right]\end{aligned}$$

Define the corresponding

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

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

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

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 $\tau_0 = \min\left\{t \in [T] \middle| \sum_{i=1}^{t} x_t \geq \frac{1}{2}DT - D\right\} \cup \{T\}$. 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

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} f_t(x_t)\right] &\leq \mathbb{E}\left[\sum_{t=1}^{\tau_0} f_t(\tilde{x}_t(\lambda_{t-1})) + \langle \frac{1}{2}D - \tilde{x}_t(\lambda_{t-1}), \lambda^* \rangle\right] + \mathbb{E}\left[\sum_{t=\tau_0+1}^{T} f_t(x_t) + \langle \frac{1}{2}D - x_t(\lambda), \lambda^* \rangle\right] \\
&\leq \mathbb{E}\sum_{t=1}^{\tau} g(\lambda_{t-1}) + \mathbb{E}\left[\sum_{t=\tau_0+1}^{T} \frac{3}{4}Dx_t + \frac{1}{2}D\lambda^*\right] \\
&\leq \mathbb{E}\sum_{t=1}^{\tau} g(\lambda_{t-1}) + \frac{3}{16}D^2\mathbb{E}[T - \tau_0] + \frac{3}{4}D^2.
\end{aligned}$$

The first inequality is because of the resource constraint, and the second one is because $f_t(x) \leq f_t'(0)(x-0) \leq \frac{3}{4}Dx$. 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

$$\begin{aligned}
Tg(\lambda^*) - \mathbb{E}\left[\sum_{t=1}^{T} f_t(x_t)\right] &\geq \mathbb{E}\left[\sum_{t=1}^{\tau_0} g(\lambda^*) - g(\lambda_{t-1})\right] + \mathbb{E}(g(\lambda^*) - \frac{3}{16}D^2)\mathbb{E}[T - \tau_0] - \frac{3}{4}D^2 \\
&= \mathbb{E}\left[\sum_{t=1}^{\tau_0} (\lambda^* - \lambda_{t-1})^2\right] + \frac{5}{64}D^2\mathbb{E}[T - \tau_0] - \frac{3}{4}D^2,
\end{aligned} \tag{10.23}$$

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 λ^* (Li and Ye 2021), 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 $\mathbb{E}\tau_0 \leq (1-c)T + c(T - \sqrt{T}) \leq T - c\sqrt{T}$.

$$\begin{aligned}\mathbb{P}(\tau_0 \le t') &= \mathbb{P}\left(\sum_{t=1}^{t'} 2(\xi_t - \lambda_{t-1}) \ge \frac{DT}{2} - D\right)\\ &\ge \mathbb{P}\left(\left\{\sum_{t=1}^{t'} 2(\xi_t - \lambda^*) \ge \frac{DT}{2} - D + \varepsilon D\sqrt{t'}\right\} \cap \left\{\sum_{t=1}^{t'} |\lambda_{t-1} - \lambda^*| < \varepsilon D\sqrt{t'}\right\}\right)\\ &\ge \mathbb{P}\left(\left\{\sum_{t=1}^{t'} 2(\xi_t - \lambda^*) \ge \frac{DT}{2} - D + \varepsilon D\sqrt{t'}\right\}\right) - \mathbb{P}\left(\sum_{t=1}^{t'} |\lambda_{t-1} - \lambda^*| \ge \varepsilon D\sqrt{t'}\right)\end{aligned}$$

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

$$\begin{aligned}\mathbb{P}(\tau_0 \le t') &\ge \mathbb{P}\left(\left\{\sum_{t=1}^{t'} 2(\xi_t - \lambda^*) \ge \frac{DT}{2} - D + \varepsilon D\sqrt{t'}\right\}\right) - \frac{2c_2}{\varepsilon}\\ &= \mathbb{P}\left(\left\{\sum_{t=1}^{t'} \frac{4}{D}(\xi_t - \frac{D}{2}) \ge \frac{t'}{2} + (T - t') - 2 + 2\varepsilon\sqrt{t'}\right\}\right) - \frac{2c_2}{\varepsilon}\\ &\ge \mathbb{P}\left(\left\{\sum_{t=1}^{t'} \frac{4}{D}(\xi_t - \frac{D}{2}) \ge \frac{t'}{2} + (1 + 2\varepsilon)\sqrt{t'}\right\}\right) - \frac{2c_2}{\varepsilon},\end{aligned}$$

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 $T - t' \le \sqrt{T} + 1$ and $\sqrt{T} - \sqrt{t'} \le \sqrt{T} - \sqrt{T - \sqrt{T}} = \frac{\sqrt{T}}{\sqrt{T - \sqrt{T}} + \sqrt{T}} \le 1$. For the binomial distribution, $\mathbb{P}(X \ge \mu + x\sigma)$ 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, $\mathbb{P}\left(\left\{\sum_{t=1}^{t'} \frac{4}{D}(\xi_t - \frac{D}{2}) \ge \frac{t'}{2} + (1 + 2\varepsilon_0)\sqrt{t'}\right\}\right) \ge \frac{3c_2}{\varepsilon_0}$, which indicates that $\mathbb{P}(\tau_0 \le t') \ge \frac{c_2}{\varepsilon_0}$. This makes our proof complete.

10.15. Proof of Theorem 6

Theorem 6 can be proved following the same path as in the proof of Theorem 3. The key is that, with a good initialization, Lemma 4 and 5 still hold. This has actually been mentioned in the proof of Lemma 4 and 5 and thus omitted here.

10.16. Proof of Theorem 7

To prove the $O(\log^2 T)$ bound, we revise the previous proof of Lemma 4 and 5 and re-compute some important rates. For online gradient descent, without loss of generality, suppose $d_t \in \Omega_d$ before the

stopping time τ. Then By Assumption 3 and the stochastic gradient descent approach in Rakhlin et al. (2012), for $t \geq T_j$,we have

$$\begin{aligned}\mathbb{E}\left(\left\|\nu_{t+1}-\nu^*(d_{T_j})\right\|_2^2+\left\|\mu_{t+1}-\mu^*(d_{T_j})\right\|_2^2\right) \leq \mathbb{E}\left\|\nu_t-\eta_{t+1}\nabla D_{t+1,\nu}(\nu_t,\mu_t,d_{T_j})-\nu^*(d_{T_j})\right\|_2^2 \\ +\mathbb{E}\left\|\mu_{t+1}-\mu^*(d_{T_j})-\eta_{t+1}\nabla D_{t+1,\mu}(\nu_t,\mu_t,d_{T_j})\right\|_2^2,\end{aligned}$$

and we also have

$$\begin{aligned}&\mathbb{E}\left\|\nu_{t+1}-\nu^*(d_{T_j})\right\|_2^2 \leq \mathbb{E}\left\|\nu_t-\eta_{t+1}\nabla D_{\nu}(\nu_t,\mu_t,d_{T_j})-\nu^*(d_{T_j})\right\|_2^2, \text{ by projection}\\ &\mathbb{E}\left\|\nu_{t+1}-\nu^*(d_{T_j})\right\|_2^2 \leq (1-2\eta_{t+1}\sigma_{\min}\underline{\mathcal{L}}_f)\mathbb{E}\left\|\nu_t-\nu^*(d_{T_j})\right\|_2^2+\eta_{t+1}^2 n\bar{b}^2D^2.\end{aligned}$$

And hence, by choosing $\eta_t = 1/(\sigma_{\min}\underline{\mathcal{L}}_f(t-T_j+1))$, we have

$$\mathbb{E}\left[\left\|\nu_t-\nu^*(d_{T_j})\right\|_2^2\middle|\mathcal{H}_{T_j}\right] \leq \frac{4n\bar{b}^2D^2}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2(t-T_j+1)}, \tag{10.24}$$

where $\nu^*(d_{T_j})$ is part of the optimal solution to $D(\boldsymbol{\lambda}, d_{T_j})$. See Rakhlin et al. (2012) for a detailed approach. By this convergence rate, we have the following rates

$$\mathbb{E}\left[\sum_{t=1}^{\tau}\left\|\nu_{t-1}-\nu^*(d_{t-1})\right\|_2^2\right] \leq \sum_{j=1}^{J}\frac{4n\bar{b}^2D^2(\log(T_{j+1}-T_j)+1)}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2} \leq \frac{Cn\bar{b}^2D^2(\log^2 T)}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2}$$

We now revisit the previous induction of d_t in (10.17). For the three parts A', B', C', we have the new rates:

$$\begin{aligned}A' &= \mathbb{E}\frac{\left(\sum_{k=T_j+1}^{T_{j+1}}\left[d'_{i,T_j}-\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i|\mathcal{H}_{T_j}\right]+\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i|\mathcal{H}_{T_j}\right]-(b_k\tilde{x}_k(\nu_{k-1}))_i\right]\right)^2}{(T-T_{j+1})^2}\mathbb{I}(\tau>T_j)\\ &\leq 2\mathbb{E}\frac{\sum_{k=T_j+1}^{T_{j+1}}\left(\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i-(b_k\tilde{x}_k(\nu_{k-1}))_i\middle|\mathcal{H}_{T_j}\right]\right)^2}{T-T_{j+1}}\\ &\leq \frac{2n\bar{b}^4L_1^2D^2}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2}\frac{\log(T_{j+1}-T_j)}{T-T_{j+1}}\\ &\leq \frac{C_4n\bar{b}^4L_1^2D^2\log T}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2\rho^jT}\end{aligned}$$

And for B', we also have:

$$d'_{i,T_j}-\mathbb{E}\left[\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i\middle|\mathcal{H}_{T_j}\right]=0, \text{ for any } k \geq T_j+1.$$

For part C', it follows that

$$\begin{aligned}
C' &= 2\mathbb{E}\left[\mathbb{E}\frac{\left(d'_{i,T_j}-d_i\right)\left(\sum_{k=T_j+1}^{T_{j+1}}\left(b_k\tilde{x}_k(\nu_{k-1})\right)_i-\left(b_k\tilde{x}_k(\nu^*(d_{T_j}))\right)_i\right)}{T-T_{j+1}}\mathbb{I}(\tau>T_j)\middle|\mathcal{H}_{T_j}\right]\\
&\le 2\bar{b}^2L_1\mathbb{E}\frac{\left|d'_{i,T_j}-d_i\right|\sum_{k=T_j+1}^{T_{j+1}}\left\|\nu_{k-1}-\nu^*(d_{T_j})\right\|}{T-T_{j+1}}\\
&\le \frac{C\sqrt{n}\bar{b}^3DL_1}{\sigma_{\min}\underline{\mathcal{L}}_f}\sqrt{\mathbb{E}\left(d'_{i,T_j}-d_i\right)^2}\sqrt{\frac{1}{T-T_j}}\\
&\le \frac{C\sqrt{n}\bar{b}^3DL_1}{\sigma_{\min}\underline{\mathcal{L}}_f}\sqrt{\mathbb{E}\left(d'_{i,T_j}-d_i\right)^2}\sqrt{\frac{1}{\rho^jT}}.
\end{aligned}$$

Thus, we then get the recurrence relation of $d'_{i,T_j}-d_i$:

$$\begin{aligned}
\mathbb{E}\left(d'_{i,T_{j+1}}-d_i\right)^2 \le& \mathbb{E}\left(d'_{i,T_j}-d_i\right)^2+\frac{C_4n\bar{b}^4L_1^2D^2\log T}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2\rho^jT}\\
&+\frac{C\sqrt{n}\bar{b}^3DL_1}{\sigma_{\min}\underline{\mathcal{L}}_f}\sqrt{\mathbb{E}\left(d'_{i,T_j}-d_i\right)^2}\sqrt{\frac{1}{\rho^jT}}.
\end{aligned}$$

Since $d_0=d$, by induction we have

$$\mathbb{E}\left(d'_{i,T_j}-d_i\right)^2\le C_5C_\rho nD^2\bar{b}^6L_1^2\frac{\log T}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2\rho^jT}$$

So, we have

$$\begin{aligned}
2\mathbb{E}\left[\sum_{t=1}^{\tau}\|\nu^*(d_{t-1})-\nu^*\|^2\right] \le& 2\mathbb{E}\left[\sum_{t=1}^{\tau}\sum_{i\in I_{\mathsf{B}}}(d_{i,t-1}-d_i)^2\right]\\
\le& 2m\mathbb{E}\sum_{j=1}^{J}(T_j-T_{j-1})\left[\left(d'_{i,T_j}-d_i\right)^2\right]\le C_5\frac{mnD^2\bar{b}^6L_1^2\log^2T}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2}+C,\\
\text{and }\mathbb{E}\left[\sum_{t=1}^{\tau}\|\nu_{t-1}-\nu^*\|_2^2\right] \le& \frac{Cmn\bar{b}^6D^2L_1^2(\log^2T)}{\sigma_{\min}^2\underline{\mathcal{L}}_f^2}+2C_2,
\end{aligned}$$

Extending this computation to the control of stopping time, we also have

$$\mathbb{E}(T-\tau)\le\frac{Cmn\bar{b}^6D^2L_1^2\log^2T}{\delta_d^2\sigma_{\min}^2\underline{\mathcal{L}}_f^2},$$

where the proof essentially follows Lemma 5. Similarly, following the proof of Lemma 8, we also have

$$\mathbb{E}\left\|\hat{\mu}_{I_{\mathsf{NB}},T}-\mu^*_{I_{\mathsf{NB}}}\right\|_2^2\le O(\frac{\log^2T}{T}),\text{ and }\ \mathbb{E}\left\|\sum_{t=1}^{\tau}(\tilde{a}(\mu^*)-b_tx_t)\right\|_2^2\le O(T\log^2T).$$

Combining these together, we have the regret upper bound

$$\text{Regret}(A)\le\widetilde{C}\log^2T,$$

for some constant $\widetilde{C}=O(mn^2)$ depending on the values in Assumptions 1-5. Here the additional n factor comes from the complementary slackness bound by the scale of our problem as is shown in Lemma 8.

11. Numerical Experiments

The implementation details on multiple input models are as follows: the dual updates are 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 parameter settings of different inputs. For each T, we randomly generate T observations from distribution, run each algorithm in an online fashion, and keep record of their output. The regret is calculated as the difference between the offline optimal (solved by *cvxpy*) and the online output. We report the average regret over 10 repetitions for all following graphs.

Input	$f_t(x)$	$r(x)$	b_{it}	d_i
I	$a_t^\top x$	$-\kappa \left\|x - d/2\right\|_2^2$	$\mathrm{U}(0,1)$	0.1
II	$a_t^\top x$	$-\kappa \left\|x - d/2\right\|_2^2$	$\mathrm{Bernoulli}(p_i)$	$\mathrm{U}(0.25, 0.75)$
III	$-\frac{1}{4}x^2 + \xi_t x$	0	1	0.5
IV	$-\frac{a_t}{4}x^2 + \frac{a_t x}{2}$	$\kappa \min_i x_i/d_i$	$\mathrm{U}(0, 0.5)$	0.3

Table 1 Parameter Settings of Inputs

Input model I: Online welfare maximization with costs, independent reward, and resource consumption. The reward functions are linear as $f_t(x) = a_t^\top x$. The regularization function is the ℓ_2 loss $r(x) = -\kappa \left\|x - d/2\right\|_2^2$, 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 vectors with dimension $m = 6$. Sepecifically, a_{it} is generated from the uniform distribution $U(0, 10)$, and b_{it} is generated from the uniform distribution $U(0, 1)$. κ is set to 0.001.

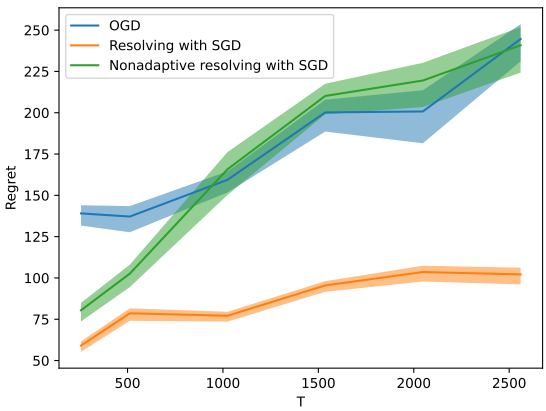
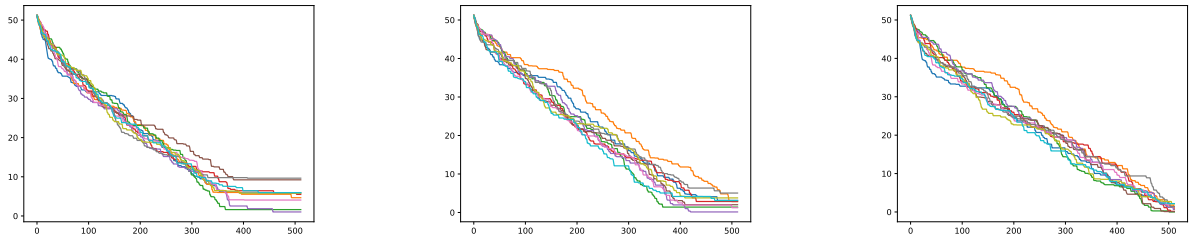


Figure 2 Regret versus horizon (T) under Input I. OGD stands for online gradient descent in Balseiro et al. (2020); resolving with SGD is our Algorithm 2; nonadaptive resolving with SGD is the nonadaptive version (i.e., without updating the constraints) of Algorithm 2.



(a) Online gradient descent (b) Nonadaptive version of Algorithm 2 (c) Algorithm 2

Figure 3 Remaining resource of one binding dimension versus time under input model I with $T = 512$. Ten curves are displayed, each of which corresponds to one simulation.

To illustrate how the regret scales with the time horizon T, we evaluate the algorithms with different T chosen from $\{256, 512, 1024, 1536, 2048, 2560\}$. 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 2. In Figure 3, we plot the dynamics 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 $a_t = \text{Proj}_{[0,10]}\{\theta_t^\top b_t + \delta_t \mathbf{1}\}$, where θ_t is generated from a multi-variate Gaussian distribution $N(0, \text{diag}(1))$, and δ_t is generated from the standard Gaussian distribution $N(0,1)$. The constraint coefficients b_{it}'s are generated from Bernoulli distribution with probability parameter p_i with $p_i = (1+\alpha)/2$, and α is generated from the beta distribution Beta$(1,3)$. The average resource constraints d_i's are generated from the uniform distribution U$(0.25, 0.75)$. κ is set to 0.001.

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 4 and Figure 5, respectively. Among the three algorithms (Algorithm 2, the nonadaptive Algorithm 2 and the OGD method in Balseiro et al. (2020)), Algorithm 2 achieves a logarithmic regret, the

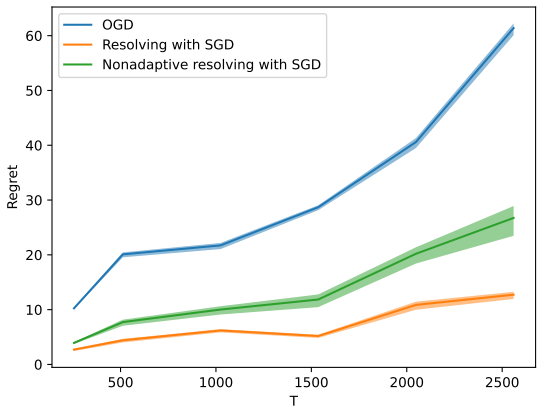
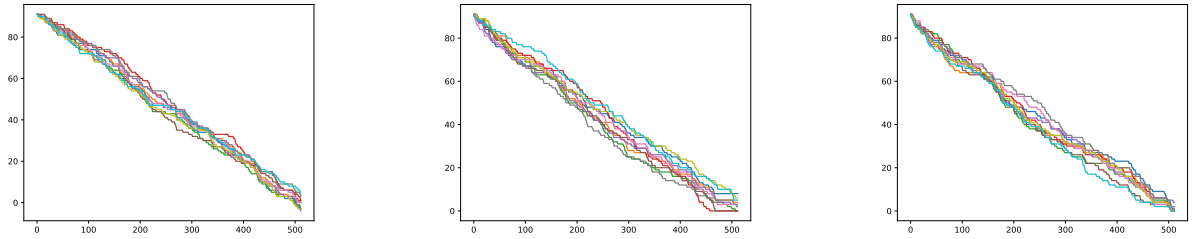


Figure 4 Regret versus horizon (T) under Input II. OGD stands for online gradient descent in Balseiro et al. (2020); resolving with SGD is Algorithm 2; nonadaptive resolving with SGD is the nonadaptive version of Algorithm 2.

nonadaptive Algorithm 2 suffers from a higher regret while the regret of OGD grows in a much faster speed.

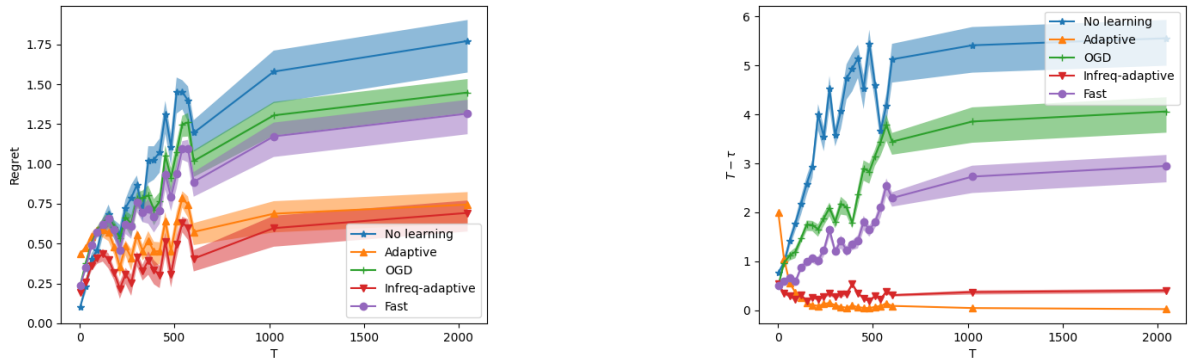


(a) Online gradient descent (b) Nonadaptive version of Algorithm 2 (c) Algorithm 2

Figure 5 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.

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 ξ_t follows a two-point distribution that takes value in $\{\frac{1}{2}, \frac{3}{4}\}$ with equal probability, i.e., $\mathbb{P}[\xi_t = \frac{1}{2}] = \mathbb{P}[\xi_t = \frac{3}{4}] = 0.5$. This special case is used in the proof of Theorem 5.

For input model III, the optimal solution to Problem (2.8) admits a closed-form due to the simple distribution, which also leads to a relatively small regret compared to other input models. We compare further with the "No learning" algorithm, which is the convex version of Algorithm 1 in Li and Ye (2021). It requires the computation of optimal dual solutions, while neither Adaptive



(a) Regret versus horizon (T) under input III.

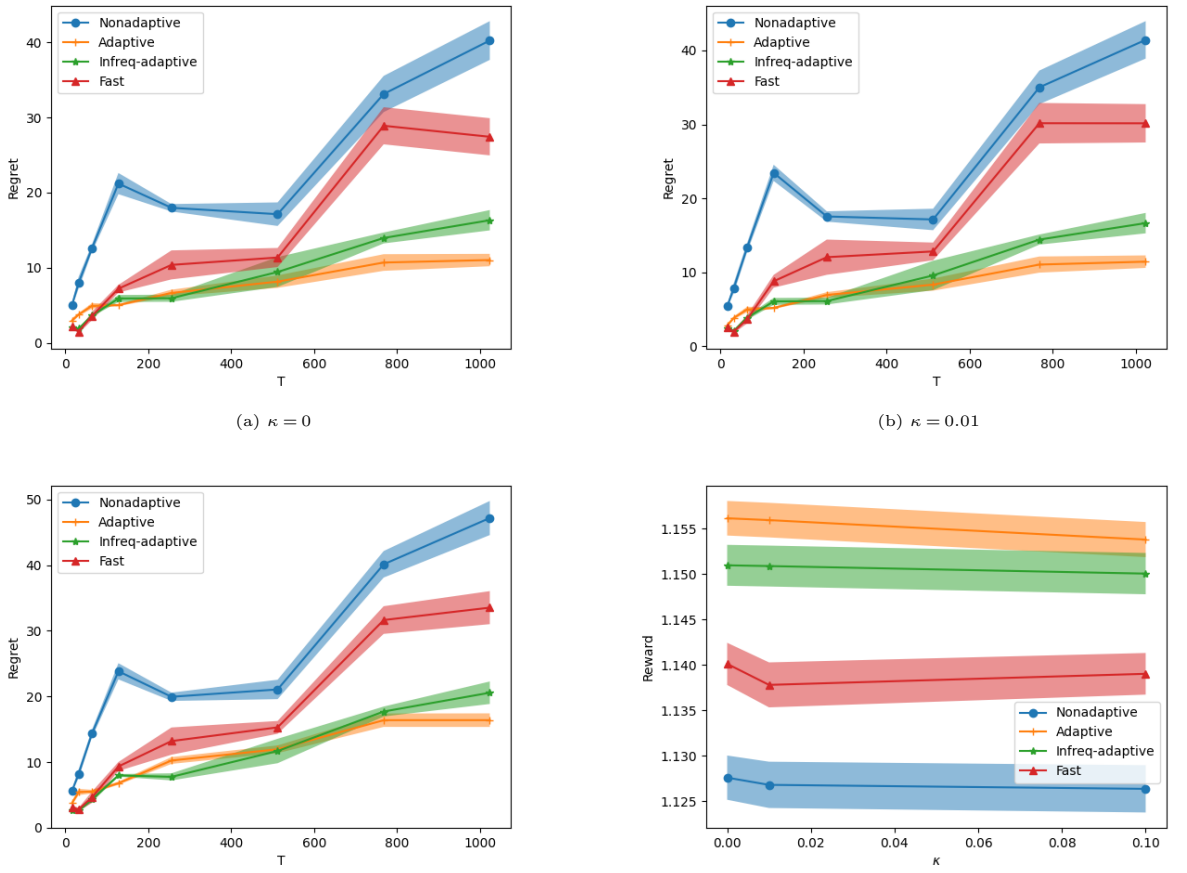
(b) Remaining time ($T-\tau$) versus horizon (T) under input III

Figure 6 Performance evaluation under input III. No learning is the convex version of Algorithm 1 in Li and Ye (2021); OGD is online gradient descent in Balseiro et al. (2020); Adaptive is Algorithm 2; infreq-adaptive is Algorithm 4; Fast is Algorithm 5.

(Algorithm 2) nor OGD needs this information. The regret comparison is shown in Figure 6a. The empirical performance corroborates the theoretical results in that the infrequent resolving saves computation significantly with a relatively small performance loss. We further explain the reason for the performance advantage by plotting the remaining time before stopping in Figure 6b. Benchmark algorithms without constraint update (No learning, OGD) stop allocating resource $O(\sqrt{T})$ steps earlier than Adaptive (Algorithm 2) and Infreq-adaptive (Algorithm 4), which leads to the terrible regret performance. In comparison, Fast (Algorithm 5) uses both the updated constraints and the original constraints, leading to a less accurate dual solution compared to Algorithm 4 and worse performance.

Input model IV: Online convex allocation with max-min regularizer. To demonstrate that our algorithm also works well for nonsmooth regularizers, we set up input model IV. The reward functions are $f_t(x) = -\frac{a_t}{4}x^2 + \frac{a_t}{2}x$, where $a_t \in \mathbb{R}$ is generated from the uniform distribution $U(0,10)$. Each dimension of the tth constraint coefficient, i.e., $b_{it}, i \in [m]$, is generated from the uniform distribution $U(0,0.5)$. We set $m = 12$ in this input model. The regularization function is $r(x) = \kappa \min_i x_i/d_i$ so that the minimum allocation is not too small. κ is set to $[0, 0.01, 0.1]$ to compare the performance of algorithms under different regularization levels.

In Figure 7, regret curves under different κs of Nonadaptive (Algorithm 3 with OGD), Adaptive (Algorithm 2), Infreq-adaptive (Algorithm 4) and Fast (Algorithm 5) are plotted. The regret of Algorithm 2 and Algorithm 4 grow slower than the other two, which shows the advantage of using updated constraint information. It is observed in Figure 7d that the per-step reward decreases as we regularize the solution by setting $\kappa > 0$, which is consistent with the intuition that regularization has a negative impact on the revenue. Also, it is observed that the reward decreases when the regularization level further increases, but we also observe a slight increase for Fast (Algorithm 5) when



(a) $\kappa = 0$

(b) $\kappa = 0.01$

(c) $\kappa = 0.1$

(d) Per-step reward $(\sum_t f_t(x_t)/T)$ versus κ when $T = 1024$.

Figure 7 Regret versus horizon (T) under input IV with different κs in Figure 7a. Nonadaptive is Algorithm 3 with OGD; Adaptive is Algorithm 2; Infreq-adaptive is Algorithm 4; Fast is Algorithm 5. Figure 7d shows the impact of different regularization levels on the per-step reward (with 95% confidence interval).

κ increases from 0.01 to 0.1, which indicates that a proper regularization level may not decrease the reward too much. It would be interesting to study the problem of choosing the appropriate regularization level in the future.

12. More Discussions

12.1. Computational cost

Specifically, for strongly convex dual objectives, Our algorithm of frequent resolving requires computing gradients for $O(T^2)$ times in total; for more general dual objectives, it requires $O(T^4)$ times of gradient computation in total. With a good initialization, we can unevenly divide time horizon T into $\log T$ epochs and only solve empirical dual optimization at the beginning of each epoch. This infrequent algorithm only takes gradient computations for $O(T \log T)$ times for strongly convex problems (which is nearly linear) and $O(T^3 \log T)$ for general problems, while delivering an optimal

regret bound. The fast algorithm we established has only linear computational cost $O(T)$, which is comparable with OGD but reaches sub-optimal regret $O(\log^2 T)$.

12.2. Fenchel conjugate of regularizers

Here we provide Fenchel conjugates for three commonly used regularizers.

1. ℓ_1**-loss**: $r(a) := -\kappa \|a\|_1$. Define $\mathcal{Z} := \{a | \|a\|_\infty \le L\}$, $h(a) = r(a) - \mu^\top a$ then $r^*(\mu) = \max_{a\in\mathcal{Z}}\{r(a) - \mu^\top a\} := \max_{a\in\mathcal{Z}}\{h(a)\}$. We have the subgradient: $\nabla h(a) = -\kappa\,\mathrm{sign}(a) - \mu$. Since $\|\mu\|_\infty \le G = \kappa$, we know that $h(a)$ takes its maximum only when $a = 0$. the conjugate r^* in Ω_μ is thus of the form $r^*(\mu) = 0$ when $\kappa \ge G = \kappa$.
2. **Max-min loss**: $r(a) := \kappa \min_i(a_i/d_i)$. Define $\mathcal{Z} := \{a | |a_i| \le d_i L,\ i \in [m]\}$, and $z_i = a_i/d_i$. Then we define the function to be maximized as $h(z) := r(a) - \mu^\top a = \kappa \min(z_i) - \sum \mu_i d_i z_i$, for the region $\|z\|_\infty \le L$. By computing the subgradient of each dimension, we know that the optimal z that maximizes the $h(z)$ must have:
$$z_i = \begin{cases} L & \text{if } \mu_i d_i < 0 \\ -L & \text{if } \mu_i d_i > \kappa \\ \min_i(z_i) & \text{if } \mu_i d_i \in [0, \kappa] \end{cases}.$$
Therefor, we have $z_i = \min_i(z_i)$ if $\mu_i \ge 0$. Moreover, whether $\min_i(z_i)$ takes L or $-L$ depends on the value $\kappa - \sum d_i (\mu_i)_+$. If $\kappa - \sum d_i (\mu_i)_+ > 0$, then we will have $\min_i(z_i) = L$, otherwise $\min_i(z_i) = -L$. Thus, the conjugate r^* in Ω_μ is of the form $r^*(\mu) = L \cdot \left(\left|\kappa - \sum d_i (\mu_i)_+\right| - \sum d_i (\mu_i)_-\right)$.
3. **Negative max loss**: $r(a) := -\kappa \max_i(a_i/d_i)$. Define $\mathcal{Z} := \{a | |a_i| \le d_i L,\ i \in [m]\}$, and $z_i = a_i/d_i$. Then we have $h(z) := r(a) - \mu^\top a = -\kappa \max(z_i) - \sum \mu_i d_i z_i$. Following an analogous argument as in the max-min loss, we have $r^*(\mu) = L \cdot \left(\left|\kappa + \sum d_i (\mu_i)_-\right| + \sum d_i (\mu_i)_+\right)$