Fast Algorithms for Online Stochastic Convex Programming

Shipra Agrawal[*] Nikhil R. Devanur[†]

Abstract

We introduce the *online stochastic Convex Programming (CP)* problem, a very general version of stochastic online problems which allows arbitrary concave objectives and convex feasibility constraints. Many well-studied problems like online stochastic packing and covering, online stochastic matching with concave returns, etc. form a special case of online stochastic CP. We present fast algorithms for these problems, which achieve near-optimal regret guarantees *for both the i.i.d. and the random permutation models* of stochastic inputs. When applied to the special case online packing, our ideas yield a simpler and faster primal-dual algorithm for this well studied problem, which achieves the optimal *competitive ratio*. Our techniques make explicit the connection of primal-dual paradigm and online learning to online stochastic CP.

1 Introduction

The theory of online matching and its generalizations has been a great success story that has had a significant impact on practice. The problems considered in this area are largely motivated by online advertising, and the theory has influenced how real advertising systems are run. As an example, the algorithms given by Devanur et al. [18] are being used at Microsoft, by the "delivery engine" that decides which display ads are shown on its "properties" such as webpages, Skype, Xbox, etc.

In one of the most basic problem formulations in online advertising, an "impression" can be allocated to one of many given advertisers, assigning an impression i to advertiser a generates a value v_{ai}, and an advertiser a can be allocated at most G_a impressions. The goal is to maximize the value of the allocation. In another variant, advertisers pay per click and have budget constraints on their total payment, instead of the capacity constraints as above. More sophisticated formulations consider the option to show multiple ads on one webpage, which means you can pick among various configurations of ads. Each configuration still provides some value which is to be maximized, and advertisers have either capacity or budget constraints.

While the algorithm in Devanur et al. [18] (DJSW algorithm) is used in practice, the actual problem has some aspects that are not captured by the formulations considered there. For instance, the actual objective function is not just a linear function, such as the sum of the values. There is a penalty for "under-delivering" impressions to an advertiser that increases with the amount of under-delivery. This translates into an objective that is a concave function of the total number of impressions assigned to an advertiser. Another consideration is the diversity of the impressions assigned. An advertiser targeting a certain segment of the population expects a representative sample of the entire population [24]. In order to avoid deviating from this ideal too much, there are certain (convex) penalty functions in the objective that punish such deviations. The 'essentially linear' formulations of online matching or online packing/covering considered in the literature cannot handle these extensions. In this paper, we consider a very general online convex programming framework that can incorporate these extensions, and present optimal algorithms for it.

An important practical consideration in the design of online algorithms is that the time taken by the algorithm in a single step should be very small. For instance, the decision to allocate an impression must be made in "real-time", in a matter of milliseconds. The DJSW algorithm satisfies this requirement, but requires solving an LP ever so often, to estimate the value of an optimum solution. In this paper, we give an algorithm that only requires solving a single LP (for *online packing* problems), making it even faster than the DJSW algorithm. This improvement comes from the fact that in our algorithm the error in the estimate of the optimal solution only occurs in the second order error bounds and hence we can tolerate much bigger errors in such an estimate.

From a theoretical point of view, two closely related online stochastic input models have been studied, the random permutation and the i.i.d. model. In the random permutation model, an adversary picks the *set* of inputs, which are then presented to the algorithm in a random order. In the i.i.d. model, the adversary picks a distribution over inputs that is unknown to the algorithm, and the algorithm receives i.i.d. samples from this distribution. The random permutation model is stronger than the i.i.d. model, any algorithm that works for the random permutation model also works for the i.i.d model. The difference between these two

[*]Microsoft Research
[†]Microsoft Research

models is like the difference between sampling with and without replacement. This intuition says that the two models should be very similar to each other, but the DJSW algorithm was only known to work for the i.i.d model, not for the random permutation model. Earlier algorithms by Devanur and Hayes [15], Agrawal et al. [3], Feldman et al. [22] worked for the random permutation model but gave worse guarantees. Recently Kesselheim et al. [33] gave an algorithm that matched the optimal guarantee of Devanur et al. [18] for the random permutation model, but their algorithm has to solve an LP in every step, making it not practical. To summarize, the DJSW algorithm is fast and works for the i.i.d. model but not for the random permutation model. The algorithm by Kesselheim et al. [33] works for the random permutation model but is slow. We get the best of both worlds, our algorithm is fast, and works for the random permutation model. Moreover, our proof formalizes the intuition mentioned earlier that the difference between i.i.d and the random permutation models is like the difference between sampling with and without replacement.

In terms of techniques, the earlier algorithms used dual variables to guide the allocation, whereas the algorithm of Kesselheim et al. [33] uses a primal only approach, and their result seemed to suggest that primal-only algorithms were more powerful than primal-dual algorithms. Our algorithms are primal-dual, and our results show that primal-dual algorithms are equally powerful. In fact, even though the DJSW algorithm could be interpreted as a primal-dual algorithm, duality was never used in the analysis. Our algorithm is a true primal-dual algorithm in the sense that we explicitly make use of the duality. Also, starting from Mehta et al. [39], it was suspected that there is some relation between these problems and online learning or the "experts" problem, but no formal connection was known. We show such a formal connection, all of our algorithms actually use blackbox access to algorithms for solving online learning problems. We show how getting better guarantees for these problems boils down to getting better "low-regret" guarantees for certain online learning problems. This also gives much simpler proofs than earlier papers.

To summarize, our contributions are as follows.

1. We present algorithms with optimal guarantees for a very general online convex programming problem, in a stochastic setting.

2. Our algorithms are primal-dual algorithms that are fast and simple, and work for the random permutation model. Our proof techniques formalize the intuition that the random permutation and the i.i.d models are not that different.

3. We establish a formal connection between these problems and online learning.

1.1 Other Related Work The seminal paper of Mehta et al. [39] introduced the so called "Adwords" problem, motivated by the allocation of ad slots on search engines, and started a slew of research into generalizations of the online bipartite matching problem [32]. For the worst-case model, the optimal competitive ratio is $1-1/e$, which can be achieved for a fairly general setting [10, 2, 20, 19]. A special case of an objective with a concave function was considered in Devanur and Jain [16].

In order to circumvent the impossibility results in the traditional worst-case models, stochastic models such as the random permutation model and the i.i.d model were introduced [25, 15, 42, 18]. The dominant theme for these stochastic models has been asymptotic guarantees, that show that the competitive ratio tends to 1 as the "bid-to-budget" ratio tends to 0 (as was first shown by Devanur and Hayes [15]). The focus then is the *convergence rate*, the rate at which the competitive ratio tends to 1 as a function of the bid-to-budget ratio. Feldman et al. [22], Agrawal et al. [3] gave improved convergence rates for the random permutation model and generalized the result to an *online packing* problem. Recently, Chen and Wang [13] extended these ideas to the concave returns problem of Devanur and Jain [16]. Devanur et al. [18] gave the optimal convergence rate for the online packing problem in the closely related i.i.d. model. Kesselheim et al. [33] matched these bounds for the random permutation model, and further improved the bounds either when the bid-to-budget ratio is large, or when the instances are sparse. This line of research has also had significant impact on the practice of ad allocation with most of the big ad allocation platforms using algorithms influenced by these papers [23, 31, 14, 12, 11].

Some versions of these problems also appear in literature under the name of 'secretary problems'. However the dominant theme in research on secretary problems is to aim for a constant competitive ratio while not making any assumption about "bid-to-budget" ratio (a notable exception is [34]).

Another interesting line of research has been for the case of bipartite matching. Feldman et al. [21], Bahmani and Kapralov [8], Manshadi et al. [38] gave algorithms with competitive ratios better than $1 - 1/e$ for the known distribution case, and Karande et al. [30], Mahdian and Yan [36] did the same for the random permutation model. Other variations such as models for combining algorithms from worst-case and average case, and achieving simultaneous guarantees have also been studied [37, 40].

A closely related problem is called the "Bandits with Knapsacks" problem [7], which is similar to the online stochastic packing problem. The bandit aspect is different: the algorithm picks an "arm" of the bandit at each time, and makes observations (cost, reward, etc.), which are i.i.d samples that depend on the arm. There is persistence in the available set of choices across time as the arms are persistent. In the online packing problem, the set of options in one time step are unrelated to the other time steps. Due to this, the main aspect of the bandit problem, the explore-exploit trade off in estimating the expectations of the observations for all arms, is absent from the online packing problem.

In an earlier paper [4], we generalized Bandits with Knapsacks to include general convex constraints and concave rewards, which is analogous to our generalization of the online packing to online convex programming here. Our high level ideas of using Fenchel duality for 'linearization' and online learning algorithms for estimating the dual variables is inspired by the use of similar ideas in [4]. Consequently, we obtain algorithms that are very similar looking to those in [4]. There are some significant differences in the proof techniques, however, due to the differences in the two problems mentioned in the previous paragraph. Also, the analysis for the random permutation model, and our adaptations (for the online packing problem) to get competitive ratios instead of regret bounds, were entirely absent from [4].

The online packing problem is also closely related to the Blackwell approachability problem [9]. The use of online learning algorithms to solve the Blackwell approachability problem [1] is similar to our use of online learning algorithms.

Concurrently and independently, Gupta and Molinaro [26] found results for online linear programming that are similar to some of ours: they also show how to get competitive ratio bounds for the online packing problem in the random permutation model via a connection to the experts problem. For the guarantees that hold "in expectation", their bounds are the same as ours. For the guarantees that hold "with high probability", they show bounds without an extra $\sqrt{\log T}$ factor that we get. They do not consider the more general convex programming framework.

1.2 Organization: The Preliminaries section (Section 2) contains the problem and the input model definitions, the statement of the main result and some background material on online learning and Fenchel duality. Section 4 illustrates the basic ideas using a special case with only convex feasibility constraints. Section 5 gives the algorithm, results and proof techniques for the general online stochastc convex programming. Section 6 gives tighter bounds for the special case of the online packing problem.

2 Problem definition and main results

The following problem captures a very general setting of online optimization problems with global constraints and utility functions.

Definition 1. [Online Stochastic Convex Programming] *We receive an initial input of a concave function f over a bounded domain $\subseteq \mathbb{R}^d$, which we may assume is $[0,1]^d$ w.l.o.g, and a convex set $S \subseteq [0,1]^d$. Subsequently we proceed in steps, at every time step $t = 1, \dots, T$, we receive a set $A_t \subseteq [0,1]^d$ of d-dimensional vectors. We have to pick one vector $\boldsymbol{v}_t^\dagger \in A_t$ before proceeding to time step $t+1$, using only information until time t. Let $\boldsymbol{v}_{avg}^\dagger := \frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t^\dagger$. The goal is to*

$$\textit{maximize } f(\boldsymbol{v}_{avg}^\dagger) \textit{ subject to } \boldsymbol{v}_{avg}^\dagger \in S.$$

We assume that the instance is always feasible, *i.e., there is a choice of $\boldsymbol{v}_t \in A_t \ \forall\, t$ such that $\frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t \in S$.*

2.1 Stochastic Input Models: In the random permutation (RP) model, there are T sets $X_1, ..., X_T$ fixed in advance but unknown to the algorithm, and these come in a uniformly random order (given by a random permutation π) as the sequence $A_1 = X_{\pi(1)}, ..., A_T = X_{\pi(T)}$. The number of time steps T is given to the algorithm in advance. In the i.i.d, unknown distribution (IID) model, there is a distribution $\mathcal{D}$ over subsets of $[0,1]^d$, and for each t, A_t is an independent sample from $\mathcal{D}$. The distribution $\mathcal{D}$ is unknown to the algorithm.

It is known that the RP model is stronger than the IID model. The IID model can be thought of as a distribution over RP instances and therefore any guarantee for the RP model also carries over to the IID model. Henceforth, we will consider the RP model by default, unless otherwise mentioned.

2.2 Benchmarks. We measure the performance of an algorithm with respect to a benchmark. The bechmark for the RP model is the *optimal offline* solution, i.e. the choice $\boldsymbol{v}_t^* \in A_t$ that maximizes the function f of the average of these vectors while making sure that the average lies in S. We denote the value of this solution as the benchmark, OPT. This is a deterministic value since it does not depend on the randomness in the input, which is in the order of arrival. For the IID model, the offline optimal actually depends on the randomness in the input, and OPT denotes the expected value of the offline optimal solution.

2.3 Performance Measures. While the standard measure in competitive analysis of online algorithms is a multiplicative error w.r.t the benchmark, we mostly adopt a concept of additive error that is common in online learning, called the *regret*. Since we make no assumptions about f, it could even be negative, so an additive error is more appropriate. For certain special cases where multiplicative errors or competitive ratios are more natural or desirable, we discuss how our algorithms and analysis can be adapted to get such guarantees. We define the following two (average) regret measures, one for the objective and another for the constraint.[1] Let $d(\boldsymbol{v}, S)$ denote the distance of the vector $\boldsymbol{v}$ from the set S, w.r.t. a given norm $\|\cdot\|$.

$$\begin{aligned} \text{avg-regret}_1(T) &= \text{OPT} - f(\boldsymbol{v}^\dagger_{\text{avg}}), \text{ and} \\ \text{avg-regret}_2(T) &= d(\boldsymbol{v}^\dagger_{\text{avg}}, S). \end{aligned}$$

2.4 Main Results. We now state the most general result we prove in this paper.

THEOREM 2.1. *There is an algorithm (Algorithm 5.1) that achieves the following regret guarantees for the Online Stochastic Convex Programming problem, in the RP model.*

$$\begin{aligned} \mathbb{E}[\textit{avg-regret}_1(T)] &= (Z+L)\cdot O\left(\sqrt{\tfrac{C}{T}}\right) \\ \mathbb{E}[\textit{avg-regret}_2(T)] &= O\left(\sqrt{\tfrac{C}{T}}\right) \end{aligned}$$

where C depends on the norm $\|\cdot\|$ used for defining distance. For Euclidean norm, $C = d\log(d)$. For L_∞ norm, $C = \log(d)$. The parameter Z captures the tradeoff between objective and constraints for the problem, its value is problem-dependent and is discussed in detail later in the text. L is the Lipschitz constant for f w.r.t. the same norm $\|\cdot\|$ as used to measure the distance.

In the main text we provide more detailed result statements, which will also make clear the dependence of our regret bounds on the regret bounds available for online learning, and implications of using different norms. These regret bounds can also be converted to *high probability* results, with an additional $\sqrt{\log T}$ factor in the regret. This extra factor comes from simply taking a union bound over all time steps. A more careful analysis could possibly get rid of this extra factor, as was shown in Gupta and Molinaro [26] in case of online linear programming.

These bounds are optimal, and this follows easily from an easy modification of a lower bound given by Agrawal et al. [3] for the online packing problem.

We also consider the following interesting special cases.

Feasibility problem: In this case, there is no objective function f, and there is only the constraint given by the set S. The goal is to make sure that the average of the chosen vectors lies as close to S as possible, i.e., minimize $d(\boldsymbol{v}^\dagger_{\text{avg}}, S)$.

Linear objective: In this case, we assume that each vector $\boldsymbol{v} \in A_t$ has an associated reward $r \in [0, 1]$. The objective is to maximize the total reward while making sure that the average of the vectors lies in S. This can be thought of as the special case where the vector you get is $(\boldsymbol{v}, r)$, and the constraint is only on the subspace defined by all coordinates of this vector except the last, while the objective is just the sum (or linear function) of its last coordinates.

Online Packing/Covering LPs: This is a well studied special case of linear objective. The packing constraints $\sum_t \boldsymbol{v}^\dagger_t \le B\mathbf{1}$ are equivalent to using constraint set S of the form $\{\boldsymbol{v} : 0 \le \boldsymbol{v} \le \frac{B}{T}\mathbf{1}\}$, where $\mathbf{1}$ is the vector of all 1s and $B > 0$ is some scalar. In this case, we also assume that the sets A_t always contain the origin, which corresponds to the option of "doing nothing". The covering constraints are obtained when S is $\{\boldsymbol{v} : \boldsymbol{v} \ge \frac{B}{T}\mathbf{1}\}$.

For online packing, we provide the following tighter guarantee in terms of competitive ratio.

THEOREM 2.2. *For online stochastic packing problem, Algorithm 6.1 achieves a competitive ratio of $1 - O(\epsilon)$ in the RP model, given any $\epsilon > 0$ such that $\min\{B, TOPT\} \ge \log(d)/\epsilon^2$. Further, the algorithm has fast per-step updates, and needs to solve a sample LP at most once.*

3 Preliminaries

3.1 Fenchel duality. As mentioned earlier, our algorithms are primal-dual algorithms. For the online packing problem, the LP duality framework (which is very well understood) is sufficient but for general convex programs we need the stronger framework of Fenchel duality. Below we provide some background on this useful mathematical concept. Let h be a convex function defined on $[0, 1]^d$. We define h^* as Fenchel conjugate of h,

$$h^*(\boldsymbol{\theta}) := \max_{\boldsymbol{y}\in[0,1]^d}\{\boldsymbol{y}\cdot\boldsymbol{\theta} - h(\boldsymbol{y})\}$$

For a given norm $\|\cdot\|$, we denote by $\|\cdot\|_*$, the dual norm defined as:

$$\|\boldsymbol{y}\|_* = \max_{\boldsymbol{x}:\|\boldsymbol{x}\|\le 1} \boldsymbol{x}^T\boldsymbol{y}.$$

Suppose that at every point $\boldsymbol{x}$, every supergradient $\boldsymbol{g}_x$ of h has bounded dual norm $||\boldsymbol{g}_x||_* \le L$. Then, the following dual relationship is known between h and h^*.

LEMMA 3.1. $h(\boldsymbol{z}) = \max_{||\boldsymbol{\theta}||_*\le L}\{\boldsymbol{\theta}\cdot\boldsymbol{z} - h^*(\boldsymbol{\theta})\}$.

A special case is when $h(\boldsymbol{x}) = d(\boldsymbol{x}, S)$ for some convex set S. This function is 1-Lipschitz with respect to norm $||\cdot||$ used in the definition of distance. In

[1] In online learning, the objective value is the sum of reward in every step, which scales with T, and the regret typically scales with $\sqrt{T}$. But in our formulation, the objective $f(\frac{1}{T}\sum_t \boldsymbol{v}^\dagger_t)$ is defined over average observations, therefore, to be consistent with the popular terminology, we call our regret 'average regret'.

this case, $h^*(\boldsymbol{\theta}) = h_S(\boldsymbol{\theta}) := \max_{\boldsymbol{y}\in S} \boldsymbol{\theta}\cdot\boldsymbol{y}$, and Lemma 3.1 specializes to the following derivation which also appears in Abernethy et al. [1].

$$d(\boldsymbol{x}, S) = \max_{||\boldsymbol{\theta}||_*\le 1}\{\boldsymbol{\theta}\cdot\boldsymbol{x} - h_S(\boldsymbol{\theta})\}.$$

3.2 Strong convexity/Smoothness Duality. We first define strong convexity and smoothness.

Definition 2. *A function $h : \mathcal{X} \to \mathbb{R}$ is β-strongly convex w.r.t. a norm $||\cdot||$ if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}, \boldsymbol{z} \in \partial h(\boldsymbol{x})$,*

$$h(\boldsymbol{y}) - h(\boldsymbol{x}) \ge \boldsymbol{z}\cdot(\boldsymbol{y}-\boldsymbol{x}) + \frac{\beta}{2}||\boldsymbol{x}-\boldsymbol{y}||^2.$$

Equivalently for any $\boldsymbol{x}, \boldsymbol{y}$ in the interior of $\mathcal{X}$, and all $\alpha \in (0,1)$, we have that

$$\begin{aligned} h(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \quad &\ge \quad \alpha h(\boldsymbol{x}) + (1-\alpha)h(\boldsymbol{y}) \\ &\quad -\frac{\beta}{2}\alpha(1-\alpha)||\boldsymbol{x}-\boldsymbol{y}||^2. \end{aligned}$$

A function h is β-strongly concave if and only if $(-h)$ is β-strongly convex.

Definition 3. *A function $h : \mathcal{X} \to \mathbb{R}$ is β-strongly smooth w.r.t. a norm $||\cdot||$ if h is everywhere differentiable, and for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, we have*

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}, |h(\boldsymbol{y}) - h(\boldsymbol{x}) - \nabla h(\boldsymbol{x})\cdot(\boldsymbol{y}-\boldsymbol{x})| \le \frac{\beta}{2}||\boldsymbol{x}-\boldsymbol{y}||^2.$$

The following lemma can be derived from the proof of Theorem 6 in [29]. A proof is given in Appendix B.1 for completeness.

Lemma 3.2. *If h is convex and β-strongly smooth with respect to norm $\|\cdot\|$, then $h^*(\boldsymbol{\theta}) = \max_{\boldsymbol{x}\in[0,1]^d}\{\boldsymbol{\theta}\cdot\boldsymbol{x} - h(\boldsymbol{x})\}$ is $\frac{1}{\beta}$-strongly convex with respect to norm $||\cdot||_*$ on domain $\nabla_h = \{\nabla h(\boldsymbol{x}) : \boldsymbol{x} \in [0,1]^d\}$.*

3.3 Online Learning. A well studied problem in online learning, called the Online Convex Optimization (OCO) problem, considers a T round game played between a learner and an adversary (nature), where at round t, the player chooses a $\boldsymbol{\theta}_t \in W$, and then the adversary picks a concave function $g_t(\boldsymbol{\theta}_t) : W \to \mathbb{R}$. The player's choice $\boldsymbol{\theta}_t$ may only depend on the adversary's choices in the previous rounds. The goal of the player is to minimize regret defined as the difference between the player's objective value and the value of the best single choice in hindsight:

$$\mathcal{R}(T) := \max_{\boldsymbol{\theta}\in W}\sum_{t=1}^T g_t(\boldsymbol{\theta}) - \sum_{t=1}^T g_t(\boldsymbol{\theta}_t)$$

Some popular algorithms for OCO are online mirror descent (OMD) algorithm and online gradient descent, which have very fast per step update rules, and provide the following regret guarantees. More details about these algorithms and their regret guarantees are in Appendix B.2.

Lemma 3.3. *[41] There is an algorithm for the OCO problem that achieves regret*

$$\mathcal{R}(T) = O(G\sqrt{DT}),$$

where D is the diameter of W and G is an upper bound on the norm of gradient of $g_t(\boldsymbol{\theta})$ for all t. The value of these parameters are problem specific.

In particular, following corollary can be derived, which will be useful for our purpose. Details are in Appendix B.2.

Corollary 3.1. *For $g_t(\boldsymbol{\theta})$ of form $g_t(\boldsymbol{\theta}) = \boldsymbol{\theta}\cdot\boldsymbol{z} - h^*(\boldsymbol{\theta})$ and $W = \{\boldsymbol{\theta} : ||\boldsymbol{\theta}||_* \le L\}$, where h is an L-Lipschitz function, OCO algorithms achieve regret bounds of $\mathcal{R}(T) \le O(L\sqrt{dT})$ for Eucledian norm, and $O(L\sqrt{\log(d)T})$ for L_∞.*

For optimization over a simplex, the *multiplicative weight update* algorithm is very fast and efficient: the step t update of this algorithm takes the following form, given that $0 \le g_t(\boldsymbol{\theta}_t) \le M$ and a parameter $\epsilon > 0$,

(3.1)
$$\boldsymbol{\theta}_{t+1,j} = \frac{w_{t,j}}{\sum_j w_{t,j}}, \text{ where } w_{t,j} = w_{t-1,j}(1+\epsilon)^{g_t(\mathbf{e}_j)/M}.$$

The algorithm has the following stronger guarantees.

Lemma 3.4. *[5] For domain $W = \{||\boldsymbol{\theta}||_1 = 1, \boldsymbol{\theta} \ge 0\}$, given that $0 \le g_t(\boldsymbol{\theta}_t) \le M$, and for all $\epsilon > 0$, using the multiplicative weight update algorithm we obtain that for any $\boldsymbol{\theta} \in W$,*

$$\sum_{t=1}^T g_t(\boldsymbol{\theta}_t) \ge (1-\epsilon)\left(\sum_{t=1}^T g_t(\boldsymbol{\theta})\right) - \frac{M\ln(d+1)}{\epsilon},$$

For strongly concave functions, even stronger *logarithmic* regret bounds can be achieved.

Lemma 3.5. *[27] Suppose that g_t is H-strongly concave for all t, and $G \ge 0$ is an upper bound on the norm of the gradient, i.e. $\|\nabla g_t(\boldsymbol{\theta})\| \le G$, for all t. Then the online gradient descent algorithm achieves the following guarantees for OCO: for all $T \ge 1$,*

$$\mathcal{R}(T) \le \frac{G^2}{H}\log(T).$$

4 Feasibility Problem

It will be useful to first illustrate our algorithm and proof techniques for the special case of the feasibility problem. In this special case of online stochastic CP, there is no objective function f, and the aim of the algorithm is to have $\boldsymbol{v}^\dagger_{\text{avg}}$ be in the set S. The performance of the algorithm is measured by the distance from the set S, i.e., $d(\boldsymbol{v}^\dagger_{\text{avg}}, S)$. We assume that the instance is *always feasible*, i.e., there exist $\boldsymbol{v}^*_t \in A_t\ \forall\ t$ such that $\frac{1}{T}\sum_{t=1}^T \boldsymbol{v}^*_t \in S$.

The basic idea behind our algorithm is as follows. Suppose that instead of minimizing a convex function such as $d(\boldsymbol{v}^\dagger_{\text{avg}}, S)$ we had to minimize a linear function such as $\boldsymbol{\theta} \cdot \boldsymbol{v}^\dagger_{\text{avg}}$. This would be extremely easy since the problem then separates into small subproblems where at each time step we can simply solve $\min_{\boldsymbol{v}^\dagger_t \in A_t} \boldsymbol{\theta} \cdot \boldsymbol{v}^\dagger_t$. In fact, convex programming duality guarantees exactly this – that there is a $\boldsymbol{\theta}^*$, such that an optimal (i.e., feasible) solution is $\boldsymbol{v}^*_t = \arg\min_{\boldsymbol{v} \in A_t} \boldsymbol{\theta}^* \cdot \boldsymbol{v}$, however, we don't know $\boldsymbol{\theta}^*$. This is where online learning comes into play. Online learning algorithms can provide a $\boldsymbol{\theta}_t$ at every time t using only the observations before time t, which together provide a good approximation to the best $\boldsymbol{\theta}$ in hindsight.

Algorithm 4.1. (Feasibility problem)

Initialize $\boldsymbol{\theta}_1$.
for all $t = 1, ..., T$ **do**
  Set $\boldsymbol{v}^\dagger_t = \arg\min_{\boldsymbol{v} \in A_t} \boldsymbol{\theta}_t \cdot \boldsymbol{v}$
  Choose $\boldsymbol{\theta}_{t+1}$ by doing an OCO update with $g_t(\boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \boldsymbol{v}^\dagger_t - h_S(\boldsymbol{\theta})$, and domain $W = \{||\boldsymbol{\theta}||_* \le 1\}$.
end for

Here $||\cdot||_*$ is the dual norm of $||\cdot||$, the norm used in the distance function. The updates required for selecting $\boldsymbol{\theta}_{t+1}$, given $\boldsymbol{\theta}_t$ and $g_t(\cdot)$, are given as Equation B.3 and Equation 3.1 for OMD and multiplicative weight update algorithm, respectively. As discussed there, these updates are simple and fast, and do not require solving any complex optimization problems.

Theorem 4.1. *Algorithm 4.1 achieves the following regret bound for the* Feasibility Problem *in the RP model of stochastic inputs:*

$$\begin{aligned}\mathbb{E}[\textit{avg-regret}_2(T)] &:= \mathbb{E}[d(\boldsymbol{v}^\dagger_{avg}, S)] \\ &\le O\left(\frac{\mathcal{R}(T)}{T} + ||\mathbf{1}_d||\sqrt{\frac{s\log(d)}{T}}\right).\end{aligned}$$

Here $\mathcal{R}(T)$ denotes the regret for OCO with functions $g_t(\theta)$ and domain W, as defined in Section 3.3. And, $s \le 1$ is the coordinate-wise largest value a vector in S can take. This parameter can be used to obtain tighter problem-specific bounds.

Proof. From Fenchel duality, and by OCO guarantees,

$$\begin{aligned}d(\boldsymbol{v}^\dagger_{\text{avg}}, S) &= \max_{||\boldsymbol{\theta}||_* \le 1} \boldsymbol{\theta} \cdot \boldsymbol{v}^\dagger_{\text{avg}} - h_S(\boldsymbol{\theta}) \\ &= \max_{||\boldsymbol{\theta}||_* \le 1} \frac{1}{T}\sum_t g_t(\boldsymbol{\theta}) \\ &\le \frac{1}{T}\sum_t g_t(\boldsymbol{\theta}_t) + \frac{1}{T}\mathcal{R}(T).\end{aligned}$$

In Lemma 4.1, we upper bound $\mathbb{E}[\frac{1}{T}\sum_t g_t(\theta_t)]$ to obtain the statement of the theorem.

Lemma 4.1. $\mathbb{E}[\sum_t g_t(\theta_t)] \le O(||\mathbf{1}_d||\sqrt{s\log(d)T})$, *where* $s = \max_{\boldsymbol{v} \in S}\max_j v_j \le 1$, *and* $||\cdot||$ *is the norm used in the distance function.*

Proof. Let $\mathcal{F}_{t-1}$ denote the observations and decisions until time $t-1$. Note that $\boldsymbol{\theta}_t$ is completely determined by $\mathcal{F}_{t-1}$. Let $\boldsymbol{v}_{X_t}$ denote the option chosen to satisfy request X_t by the offline optimal (feasible) solution, and let $\boldsymbol{v}^*_t = \boldsymbol{v}_{A_t}$. Then, since $A_t = X_s$, for $s = 1, \dots, T$ with equal probability, we have that $\mathbb{E}[\boldsymbol{v}^*_t] = \frac{1}{T}(\boldsymbol{v}_{X_1} + \dots + \boldsymbol{v}_{X_T}) \in S$. Therefore, due to the manner in which $\boldsymbol{v}^\dagger_t$ was chosen by the algorithm, we have that

$$\begin{aligned}\mathbb{E}[g_t(\boldsymbol{\theta}_t)|\mathcal{F}_{t-1}] &= \mathbb{E}[\boldsymbol{\theta}_t \cdot \boldsymbol{v}^\dagger_t - h_S(\boldsymbol{\theta}_t)|\mathcal{F}_{t-1}] \\ &\le \mathbb{E}[\boldsymbol{\theta}_t \cdot \boldsymbol{v}^*_t - h_S(\boldsymbol{\theta}_t)|\mathcal{F}_{t-1}] \\ &= \boldsymbol{\theta}_t \cdot \mathbb{E}[\boldsymbol{v}^*_t] - h_S(\boldsymbol{\theta}_t) \\ &\quad + \boldsymbol{\theta}_t \cdot (\mathbb{E}[\boldsymbol{v}^*_t|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}^*_t])\end{aligned}$$

Now, by the Fenchel dual representation of distance, for any $\boldsymbol{v}, \boldsymbol{\theta}'$ such that $\|\boldsymbol{\theta}'\|_* \le 1$, $d(\boldsymbol{v}, S) = \max_{||\boldsymbol{\theta}||_* \le 1} \boldsymbol{\theta} \cdot \boldsymbol{v} - h_S(\boldsymbol{\theta}) \ge \boldsymbol{\theta}' \cdot \boldsymbol{v} - h_S(\boldsymbol{\theta}')$. Using this observation along with $\mathbb{E}[\boldsymbol{v}^*_t] \in S$, we obtain from above,

$$\begin{aligned}\mathbb{E}[g_t(\boldsymbol{\theta}_t)|\mathcal{F}_{t-1}] &\le d(\mathbb{E}[\boldsymbol{v}^*_t], S) + \boldsymbol{\theta}_t \cdot (\mathbb{E}[\boldsymbol{v}^*_t|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}^*_t]) \\ &= 0 + \boldsymbol{\theta}_t \cdot (\mathbb{E}[\boldsymbol{v}^*_t|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}^*_t]) \\ &\le \|\mathbb{E}[\boldsymbol{v}^*_t|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}^*_t]\|, \qquad (4.2)\end{aligned}$$

where the last inequality used the condition $\|\boldsymbol{\theta}_t\|_* \le 1$.

Note that under independence assumption (IID model), we would have $\mathbb{E}[\boldsymbol{v}^*_t|\mathcal{F}_{t-1}] = \mathbb{E}[\boldsymbol{v}^*_t]$, so that the above inequality would suffice to give the required bound. However, in random permutation (RP) model, the observations till time $t-1$ restrict the set of possible permutations. Conditional on realization $A_1 = X_{\pi(1)}, \dots, A_{t-1} = X_{\pi(t-1)}$ until time $t-1$, for a given ordering π, we have that A_t is one of the *remaining sets* with equal probability. So, $\mathbb{E}[\boldsymbol{v}^*_t|\mathcal{F}_{t-1}] = \frac{1}{T-t+1}(\boldsymbol{v}_{X_{\pi(t)}} + \dots + \boldsymbol{v}_{X_{\pi(T)}})$, for any ordering π that agrees with $\mathcal{F}_{t-1}$ on the first $t-1$ indices.

Next, we bound the gap $\|\mathbb{E}[\boldsymbol{v}^*_t|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}^*_t]\|$ under random permutation assumption. For any given ordering π, define $\mathbf{w}_{t,\pi} = \frac{\boldsymbol{v}_{X_{\pi(1)}} + \dots + \boldsymbol{v}_{X_{\pi(t)}}}{t}$. Also, for given ordering π, define π' as the reverse ordering. Then, $\mathbb{E}[\boldsymbol{v}^*_t|\mathcal{F}_{t-1}] = \mathbf{w}_{T-t+1,\pi'}$, for any ordering π that agrees with $\mathcal{F}_{t-1}$ on the first $t-1$ indices. Now, the input ordering π observed by the algorithm agrees with all the filtrations $\mathcal{F}_1, \dots, \mathcal{F}_{T-1}$, and therefore taking π' as the reverse of this ordering, we have that

$$\begin{aligned}\sum_{t=1}^{T} \|\mathbb{E}[\boldsymbol{v}^*_t|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}^*_t]\| &= \sum_{t=1}^{T} \|\mathbf{w}_{T-t+1,\pi'} - \mathbb{E}[\boldsymbol{v}^*_t]\| \\ &= \sum_{t=1}^{T} \|\mathbf{w}_{t,\pi'} - \mathbb{E}[\boldsymbol{v}^*_t]\|\end{aligned}$$

Due to the random permutation assumption, the input ordering π, and hence the reverse ordering π' in above, is a uniformly random permutation. Also, taking expectation over uniformly random permutations σ, $\mathbb{E}[\mathbf{w}_{t,\sigma}] = \frac{(\boldsymbol{v}_{X_1}+\ldots+\boldsymbol{v}_{X_T})}{T} = \mathbb{E}[\boldsymbol{v}_t^*]$. And, therefore,

$$\sum_{t=1}^{T} \|\mathbb{E}\left[\boldsymbol{v}_t^* \mid \mathcal{F}_{t-1}\right] - \mathbb{E}[\boldsymbol{v}_t^*]\| = \sum_{t=1}^{T} \|\mathbf{w}_{t,\pi} - \mathbb{E}[\mathbf{w}_{t,\sigma}]\| \tag{4.3}$$

where π is a uniformly random permutation. Taking outer expectations, and using (4.2), this implies,

$$\begin{aligned} \mathbb{E}[\sum_t g_t(\boldsymbol{\theta}_t)] &\leq \mathbb{E}\left[\sum_t \|\mathbb{E}\left[\boldsymbol{v}_t^* \mid \mathcal{F}_{t-1}\right] - \mathbb{E}[\boldsymbol{v}_t^*]\|\right] \\ &= \mathbb{E}\left[\sum_t \|\mathbf{w}_{t,\pi} - \mathbb{E}[\mathbf{w}_{t,\sigma}]\|\right]. \end{aligned}$$

Observe that for uniformly random permutation π, $\mathbf{w}_{t,\pi}$ can be viewed as the average of t vectors sampled uniformly *without replacement* from the ground set $\{\boldsymbol{v}_{X_1}, \ldots, \boldsymbol{v}_{X_T}\}$ of T vectors. We use Chernoff-Hoeffding type concentration bounds for sampling without replacement (refer to Appendix C for details), to obtain,

$$\mathbb{E}[||\mathbf{w}_{t,\pi} - \mathbb{E}[\mathbf{w}_{t,\sigma}]||] \leq O(||\mathbf{1}_d||\sqrt{\frac{s\log(d)}{t}}). \tag{4.4}$$

The lemma statement then follows by summing up these bounds over all t.

Remark 1. [RP vs. IID] *For the IID model, since* $\mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}] = \mathbb{E}[\boldsymbol{v}_t^*]$, *we would get* $\sum_t \mathbb{E}[\boldsymbol{g}_t(\boldsymbol{\theta}_t)] \leq 0$ *directly from Equation* (4.2). *Thus, the quantity* $\mathbb{E}[\sum_t \|\mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}]-\mathbb{E}[\boldsymbol{v}_t^*]\|] \leq O(\|\mathbf{1}_d\|\sqrt{sT\log(d)})$ *characterizes the gap between IID and RP models.*

Remark 2. [High probability bounds] *The above analysis can be extended to bound the sum of* conditional expectations $\sum_t \mathbb{E}[\boldsymbol{g}_t(\boldsymbol{\theta}_t)|\mathcal{F}_{t-1}] \leq \sum_t \|\mathbf{w}_{t,\pi} - \mathbb{E}[w_{t,\sigma}]\|$ *by* $O(\|\mathbf{1}_d\|\sqrt{T\log(dT/\rho)})$ *with high probability* $1-\rho$. *As a result, we obtain a high probability regret bound of* $O(\|\mathbf{1}_d\|\sqrt{\frac{\log(Td)}{T}})$. *Details are in Appendix C. For the IID model, this sum of conditional expectations is bounded by* 0, *so the resulting high probability bounds are slightly stronger, with no extra* $\sqrt{\log(T)}$ *factor.*

5 Online stochastic convex programming

In this section, we extend the algorithm from previous section to the general online stochastic Convex Programming (CP) problem, as defined in Section 2. Recall that the aim here is to maximize $f(\boldsymbol{v}_{\text{avg}}^{\dagger})$ while ensuring $\boldsymbol{v}_{\text{avg}}^{\dagger} \in S$.

A direct way to extend the algorithm from the previous section would be to reduce the convex program to the feasibilty problem with constraint set $S' = \{\boldsymbol{v} : f(\boldsymbol{v}) \geq \text{OPT}, \boldsymbol{v} \in S\}$. However, this requires the knowledge of OPT. If OPT is estimated, the errors in the estimation of OPT at all time steps t would add up to the regret, thus this approach would tolerate very small $\tilde{O}(\frac{1}{\sqrt{t}})$ per step estimation errors. In this section, we propose an alternate approach of combining objective value and distance from constraints using a parameter Z, which will capture the tradeoff between the two quantities. We may still need to estimate this parameter Z, however, Z will appear only in the second order regret terms, so that a constant factor approximation of Z will suffice to obtain optimal order of regret bounds. This makes the estimation task relatively easy and enable us to get better problem specific bounds. As a specific example, for the online packing problem, we can use $Z = \frac{\text{OPT}}{(B/T)}$ so this approach requires only a constant factor approximation of OPT and the resulting algorithm obtains the optimal competitive ratio. (See Section 6 for more details.)

To illustrate the main ideas in our algorithm, let us start with the following assumption.

Assumption 1. *Let* OPT^{δ} *denote the optimal value of the offline problem that maximizes* $f(\frac{1}{T}\sum_t \boldsymbol{v}_t)$ *with feasibility constraint relaxed to* $d(\frac{1}{T}\sum_t \boldsymbol{v}_t, S) \leq \delta$. *We are given a* $Z \geq 0$ *such that that for all* $\delta \geq 0$,

$$OPT^{\delta} \leq OPT + Z\delta. \tag{5.5}$$

In fact, such a Z always exists, as shown by the following lemma.

Lemma 5.1. OPT^{δ} *is a non-decreasing concave function of the constraint violation* δ, *and its gradient at* $\delta = 0$ *is the minimum value of* Z *that satisfies the property* (5.5). *This gradient is also equal to the value of the optimal dual variable corresponding to the distance constraint.*

The proof of this lemma is provided in Appendix D. This fact is known for linear programs.

Below, we present an algorithm (Algorithm 5.1) for online stochastic CP assuming we are given parameter Z as in Assumption 1. This algorithm is based on the same basic ideas as the algorithm for the feasibility problem in the previous section. Here, we linearize both objective and constraints using Fenchel duality, and estimate the corresponding dual variables using online learning as blackbox. And, we use parameter Z to combine objective with constraints. The resulting algorithm has very efficient per-step updates and does not require solving a (sample) CP in any step, and we prove that it achieves the regret bound stated in Theorem 2.1.

The regret of this algorithm (as stated in Theorem 2.1) scales with the value of Z, and it is desirable to use as small a value of Z as possible. If such a Z is not known, in Appendix F we demonstrate how we can approximate the optimal value of Z up to a constant factor by solving a logarithmic number of sample CPs overall.

ALGORITHM 5.1. (ONLINE CONVEX PROGRAMMING)
Initialize $\boldsymbol{\theta}_1, \boldsymbol{\phi}_1$.
for all $t = 1, ..., T$ **do**
Choose option
$\boldsymbol{v}_t^\dagger = \arg\max_{\boldsymbol{v}\in A_t} -\boldsymbol{\phi}_t \cdot \boldsymbol{v} - 2(Z+L)\boldsymbol{\theta}_t \cdot \boldsymbol{v}$.
Choose $\boldsymbol{\theta}_{t+1}$ by doing an OCO update for $g_t(\boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \boldsymbol{v}_t^\dagger - h_S(\boldsymbol{\theta})$ over domain $W = \{\|\boldsymbol{\theta}\|_* \le 1\}$.
Choose $\boldsymbol{\phi}_{t+1}$ by doing an OCO update for $\psi_t(\boldsymbol{\phi}) = \boldsymbol{\phi} \cdot \boldsymbol{v}_t^\dagger - (-f)^*(\boldsymbol{\phi})$ over domain $U = \{\|\boldsymbol{\phi}\|_* \le L\}$.
end for

A complete proof of Theorem 2.1, along with a more detailed theorem statement, is provided in Appendix E. Here, we provide the proof for the simpler case of *linear objective* discussed in Section 2. In this setting, each option in A_t is associated with a reward r in addition to the vector $\boldsymbol{v}$. And, at every time step t, the player chooses $(r_t^\dagger, \boldsymbol{v}_t^\dagger)$, in order to maximize $\frac{1}{T}\sum_t r_t^\dagger$ while ensuring $\boldsymbol{v}_{\text{avg}}^\dagger \in S$. (We will use $r_{\text{avg}}^\dagger$ to denote $\frac{1}{T}\sum_t r_t^\dagger$.) The proof for this special case will illustrate the main ideas required for proving regret bounds for the online CP problems with 'objective plus constraints', over and above the techniques used in the previous section for the case of 'only constraints'.

For this special case, Algorithm 5.1 reduces to the following:

ALGORITHM 5.2. (LINEAR OBJECTIVES)
Initialize $\boldsymbol{\theta}_1$.
for all $t = 1, ..., T$ **do**
Choose option
$(r_t^\dagger, \boldsymbol{v}_t^\dagger) = \arg\max_{(r,\boldsymbol{v})\in A_t} r - 2Z\boldsymbol{\theta}_t \cdot \boldsymbol{v}$.
Choose $\boldsymbol{\theta}_{t+1}$ by doing OCO update with $g_t(\boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \boldsymbol{v}_t^\dagger - h_S(\boldsymbol{\theta})$, and domain $W = \{\|\boldsymbol{\theta}\|_* \le 1\}$.
end for

THEOREM 5.1. *Given Z that satisfies Assumption 1, Algorithm 5.2 achieves the following regret bounds for online stochastic CP with linear objective, in RP model:*

$$\mathbb{E}[\textit{avg-regret}_1(T)] \le \frac{Z}{T} \cdot O(\mathcal{R}(T) + \mathcal{Q}(T)) \textit{ and}$$
$$\mathbb{E}[\textit{avg-regret}_2(T)] \le \frac{1}{T} \cdot O(\mathcal{R}(T) + \mathcal{Q}(T)).$$

Here, $\mathcal{Q}(T) = O(\|\mathbf{1}_d\|\sqrt{sT\log(d)})$, $s = \max_{\boldsymbol{v}\in S}\max_j v_j$, and $\mathcal{R}(T)$ denotes the OCO regret for $g_t(\cdot)$ over domain W.

Proof. Denote by $(r_t^*, \boldsymbol{v}_t^*)$ the choice made by the offline optimal solution to satisfy request A_t. Then,

$$\mathbb{E}[r_t^*] = \text{OPT}, \text{ and } \mathbb{E}[\boldsymbol{v}_t^*] \in S,$$

where expectation is over A_t drawn uniformly at random from $X_1, \dots, X_T$.

Lemma 5.2 upper bounds $\sum_t \mathbb{E}[2Zg_t(\boldsymbol{\theta}_t) - r_t^\dagger + r_t^*]$ by $2Z\mathcal{Q}(T) = 2ZO(||\mathbf{1}_d||\sqrt{s\log(d)T})$, using exactly the same line of argument as the proof of Lemma 4.1. Therefore, using $\mathbb{E}[r_t^*] = \text{OPT}$, the expected average reward obtained by the algorithm can be lower bounded as

$$\mathbb{E}[r_{\text{avg}}^\dagger] \ge \text{OPT} + \frac{2Z}{T}\sum_t \mathbb{E}[g_t(\boldsymbol{\theta}_t)] - \frac{2Z}{T}\mathcal{Q}(T).$$

As in the proof of Theorem 4.1, using Fenchel duality and OCO guarantees, it follows that $d(\boldsymbol{v}_{\text{avg}}^\dagger, S) \le \frac{1}{T}\sum_t g_t(\boldsymbol{\theta}_t) + \frac{1}{T}\mathcal{R}(T)$, which gives,

$$\mathbb{E}[r_{\text{avg}}^\dagger] \ge \text{OPT} + (2Z)\mathbb{E}[d(\boldsymbol{v}_{\text{avg}}^\dagger, S)] - \tfrac{2Z}{T}\mathcal{R}(T) - \tfrac{2Z}{T}\mathcal{Q}(T) \tag{5.6}$$

Now, we use Assumption 1 to upper bound the reward obtained by the algorithm in terms of OPT and distance from set S. In particular, for $\delta := \mathbb{E}[d(\boldsymbol{v}_{\text{avg}}^\dagger, S)]$, since $d(\mathbb{E}[\boldsymbol{v}_{\text{avg}}^\dagger], S) \le \mathbb{E}[d(\boldsymbol{v}_{\text{avg}}^\dagger, S)] = \delta$,

$$\mathbb{E}[r_{\text{avg}}^\dagger] \le \text{OPT}^\delta \le \text{OPT} + Z\delta = \text{OPT} + Z \cdot \mathbb{E}[d(\boldsymbol{v}_{\text{avg}}^\dagger, S)]. \tag{5.7}$$

Combining inequalities (5.6) and (5.7), we obtain

$$\mathbb{E}[d(\boldsymbol{v}_{\text{avg}}^\dagger, S)] \le \tfrac{2}{T}\mathcal{R}(T) + \tfrac{2}{T}\mathcal{Q}(T),$$

and from (5.6), using the fact that $\mathbb{E}[d(\boldsymbol{v}_{\text{avg}}^\dagger, S)] \ge 0$, we get that

$$\mathbb{E}[r_{\text{avg}}^\dagger] \ge \text{OPT} - \tfrac{2Z}{T} \cdot (\mathcal{R}(T) + \mathcal{Q}(T)).$$

This gives the theorem statement.

LEMMA 5.2. $\mathbb{E}[\sum_t 2Zg_t(\theta_t) - r_t^\dagger + r_t^*] \le O(Z\|\mathbf{1}_d\|\sqrt{sT\log(d)})$.

The proof of the above lemma follows exactly the same line of argument as the proof of Lemma 4.1. We omit it for brevity.

6 Online stochastic packing

Recall that the *online stochastic packing* problem is a special case of the online stochastic CP with linear objectives, with $S = \{\boldsymbol{y} : \boldsymbol{y} \le \frac{B}{T}\mathbf{1}\}$. However, the performance of an algorithm for online stochastic packing is typically measured by *competitive ratio*, which is the ratio of total expected reward obtained by the online algorithm to the optimal solution or benchmark. The benchmarks in online packing are defined as sum of rewards, where as we defined OPT as the average reward. Therefore, in our notation, the competitive ratio for the online packing problem is given by $\frac{\mathbb{E}[\sum_t r_t^\dagger]}{TOPT} = \frac{\mathbb{E}[\frac{1}{T}\sum_t r_t^\dagger]}{\mathrm{OPT}}$. The competitive ratio we obtain is $1 - O(\epsilon)$, for any $\epsilon > 0$ such that $\min\{B, T\mathrm{OPT}\} \geq \log(d)/\epsilon^2$.

Another important difference is that for online packing the budget is not allowed to be violated at all, while online CP allows a small violation of the constraint. A simple fix to make sure that budgets are not violated is to simply stop whenever a budget constraint is breached.[2] Another change we make to the algorithm is that we use a slightly different function in the OCO algorithm. We will use

$$g_t(\boldsymbol{\theta}) = (\boldsymbol{v}_t^\dagger - \frac{B}{T}\mathbf{1}) \cdot \boldsymbol{\theta}$$

over the domain $||\boldsymbol{\theta}||_1 \leq 1, \boldsymbol{\theta} \geq \mathbf{0}$. This domain is the convex hull of all the basis vectors and the origin, therefore we can use the multiplicative weight update algorithm as our OCO algorithm, which provides strong guarantees (refer to Lemma 3.4, here $M = 1$).

Finally, as with the previous algorithms, we state the algorithm assuming we are given the parameter Z. We then show how to estimate Z to desired accuracy using only an $O(\epsilon^2 \log(1/\epsilon))$ fraction of samples and solving an LP *only once* (in Lemma 6.4), assuming that $\min\{B, T\mathrm{OPT}\} \geq \frac{\log(d)}{\epsilon^2}$.

We now state the algorithm below for the online stochastic packing problem:

Algorithm 6.1. (Online Packing)

Initialize $\boldsymbol{\theta}_1 = \frac{1}{d+1}\mathbf{1}$.
Initialize Z such that $\frac{TOPT}{B} \leq Z \leq O(1)\frac{TOPT}{B}$.
for all $t = 1, ..., T$ **do**
Choose option
$(r_t^\dagger, \boldsymbol{v}_t^\dagger) = \arg\max_{(r,\boldsymbol{v})\in A_t} \{r - Z\boldsymbol{\theta}_t \cdot \boldsymbol{v}\}$.
If for some $j = 1..d$, $\sum_{t' \leq t} \boldsymbol{v}_{t'}^\dagger \cdot \mathbf{e}_j \geq B$ then EXIT.
Update $\boldsymbol{\theta}_{t+1}$ using multiplicative weight update:
$\forall\ j = 1..d, w_{t,j} = w_{t-1,j}(1+\epsilon)^{\boldsymbol{v}_t^\dagger \cdot \mathbf{e}_j - B/T}$
and
$\forall\ j = 1..d, \boldsymbol{\theta}_{t+1,j} = \frac{w_{t,j}}{1+\sum_{j'=1}^d w_{t,j'}}$,
end for

Strictly speaking, if we use the first few requests as samples to estimate Z, then we need to ignore these requests, and bound the error due to this. However, since the number of samples required is only $O(\epsilon^2 \log(1/\epsilon))$ fraction of all requests, this error is quite small relative to the guarantee we obtain, which is a competitive ratio of $1 - O(\epsilon)$. We therefore ignore this error for the ease of presentation.

[2] Note that such a stopping rule does not make sense for a general S. If S is downwards closed, then one can consider similar stopping rules in those cases as well.

Let τ be the stopping time of the algorithm. Denote by $(r_t^*, \boldsymbol{v}_t^*)$ the choice made by the offline optimal solution to satisfy request A_t. We begin with the following lemma which is similar to Lemma 4.1.

Lemma 6.1.

$$\sum_{t=1}^{\tau} \mathbb{E}[r_t^\dagger|\mathcal{F}_{t-1}] \geq \tau OPT + Z\sum_{t=1}^{\tau} \boldsymbol{\theta}_t \cdot \mathbb{E}[\boldsymbol{v}_t^\dagger - \mathbf{1}\frac{B}{T}|\mathcal{F}_{t-1}] - \sum_{t=1}^{\tau} Q(t)$$

where $Q(t) = Z||\mathbb{E}[\boldsymbol{v}_t^*] - \mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}]|| + |\mathbb{E}[r_t^*] - \mathbb{E}[r_t^*|\mathcal{F}_{t-1}]|$.

Proof. If A_t is drawn uniformly at random from $X_1, \ldots, X_T$, then $\mathbb{E}[r_t^*] = \mathrm{OPT}$, and $\mathbb{E}[\boldsymbol{v}_t^*] \leq \frac{B}{T}\mathbf{1}$. The algorithm chooses $(r_t^\dagger, \boldsymbol{v}_t^\dagger) = \arg\max_{(r,\boldsymbol{v})\in A_t} r - Z(\boldsymbol{\theta}_t \cdot \boldsymbol{v})$. By the choice made by the algorithm

$$r_t^\dagger - Z(\boldsymbol{\theta}_t \cdot \boldsymbol{v}_t^\dagger) \geq r_t^* - Z(\boldsymbol{\theta}_t \cdot \boldsymbol{v}_t^*)$$

$$\begin{aligned} \mathbb{E}[r_t^\dagger - Z(\boldsymbol{\theta}_t \cdot \boldsymbol{v}_t^\dagger)|\mathcal{F}_{t-1}] &\geq \mathbb{E}[r_t^*|\mathcal{F}_{t-1}] \\ &\quad -Z(\boldsymbol{\theta}_t \cdot \mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}]) \\ &\geq \mathbb{E}[r_t^*] - Z(\boldsymbol{\theta}_t \cdot \mathbb{E}[\boldsymbol{v}_t^*]) \\ &\quad -Q(t) \\ &\geq \mathrm{OPT} - Z\boldsymbol{\theta}_t \cdot \frac{B\mathbf{1}}{T} - Q(t) \end{aligned}$$

Summing above inequality for $t = 1$ to τ gives the lemma statement.

Lemma 6.2.

$$\sum_{t=1}^{\tau} \boldsymbol{\theta}_t \cdot (\boldsymbol{v}_t^\dagger - \frac{B}{T}\mathbf{1}) \geq (1-\epsilon)(B - \frac{\tau B}{T}) - \frac{\log(d+1)}{\epsilon}.$$

Proof. Recall that $g_t(\boldsymbol{\theta}_t) = \boldsymbol{\theta}_t \cdot \left(\boldsymbol{v}_t^\dagger - \frac{B}{T}\mathbf{1}\right)$, therefore the LHS in the required inequality is $\sum_{t=1}^{\tau} g_t(\boldsymbol{\theta}_t)$. Let $\boldsymbol{\theta}^* := \arg\max_{||\boldsymbol{\theta}||_1 \leq 1, \boldsymbol{\theta} \geq 0} \sum_{t=1}^{\tau} g_t(\boldsymbol{\theta})$. We use the regret bounds for the multiplicative weight update algorithm given in Lemma 3.4, to get that $\sum_{t=1}^{\tau} g_t(\boldsymbol{\theta}_t) \geq (1-\epsilon)\sum_{t=1}^{\tau} g_t(\boldsymbol{\theta}^*) - \frac{\log(d+1)}{\epsilon}$.

Now either $\sum_{t=1}^{\tau}(\boldsymbol{v}_t^\dagger \cdot \mathbf{e}_j) \geq B$ for some j at the stopping time τ, so that $\sum_{t=1}^{\tau} g_t(\boldsymbol{\theta}^*) \geq \sum_{t=1}^{\tau} g_t(\mathbf{e}_j) \geq B - \frac{\tau B}{T}$. Or, $\tau = T, \sum_{t=1}^{\tau}(\boldsymbol{v}_t^\dagger)_j < B$ for all j, in which case, the maximizer is $\boldsymbol{\theta}^* = \mathbf{0}$. Therefore we have that $\sum_{t=1}^{\tau} g_t(\boldsymbol{\theta}^*) \geq B - \frac{\tau B}{T}$, which completes the proof of the lemma.

Now, we are ready to prove Theorem 2.2, which states that Algorithm 6.1 achieves a competitive ratio of $1 - O(\epsilon)$, given $\min\{B, T\mathrm{OPT}\} \geq \frac{\log(d)}{\epsilon^2}$ for the online stochastic packing problem in RP model.

Proof of Theorem 2.2. Substituting the inequality from Lemma 6.2 in Lemma 6.1, we get

$$\begin{aligned}\sum_{t=1}^{\tau}\mathbb{E}[r_t^\dagger|\mathcal{F}_{t-1}] &\geq \tau\mathrm{OPT} + (1-\epsilon)ZB\left(1-\frac{\tau}{T}\right)\\ &\quad -Z\frac{\log(d+1)}{\epsilon} - \sum_{t=1}^{\tau}Q(t)\end{aligned}$$

Now, using $Z \leq O(1)\frac{T\mathrm{OPT}}{B}$ and $B \geq \frac{\log(d)}{\epsilon^2}$, we get

$$Z\frac{\log(d+1)}{\epsilon} \leq O(1)\frac{T\mathrm{OPT}}{B}\frac{\log(d+1)}{\epsilon} = O(\epsilon)T\mathrm{OPT}.$$

Also, $Z \geq \frac{T\mathrm{OPT}}{B}$. Substituting in above,

$$\begin{aligned}\sum_{t=1}^{\tau}\mathbb{E}[r_t^\dagger|\mathcal{F}_{t-1}] &\geq (1-\epsilon)\tau\mathrm{OPT} + (1-\epsilon)\mathrm{OPT}(T-\tau)\\ &\quad -O(\epsilon)T\mathrm{OPT} - \sum_{t=1}^{\tau}Q(t)\\ &\geq (1-O(\epsilon))T\mathrm{OPT} - \sum_{t=1}^{\tau}Q(t)\end{aligned}$$

Then, taking expectation on both sides, $\mathbb{E}[\sum_{t=1}^{\tau} r_t^\dagger] \geq (1-O(\epsilon))T\mathrm{OPT} - \mathbb{E}[\sum_{t=1}^{\tau}Q(t)]$.

Just like in the proof of Lemma 4.1, we can bound $\mathbb{E}[\sum_{t=1}^{\tau}Q(t)] \leq Z||\mathbf{1}_{d+1}||_\infty\sqrt{sT\log(d+1)}$ which is $O(\epsilon)T\mathrm{OPT}$, using the fact that for $S = \{\boldsymbol{y} : \boldsymbol{y} \leq \frac{B}{T}\mathbf{1}\}$, the parameter $s = \max_{j,\boldsymbol{y}\in S} y_j = \frac{B}{T}$, $||\mathbf{1}_{d+1}||_\infty = 1$, and that $Z \leq O(1)\frac{T\mathrm{OPT}}{B}, \epsilon \geq \sqrt{\frac{\log(d)}{B}}$. This completes the proof.

We now show how to compute a Z as required using the first $O(\epsilon^2\log(1/\epsilon))$ requests as samples. For convenience, let $\mathrm{OPT}_{\text{SUM}} := T\mathrm{OPT}$ denote the optimum for the sum. We first state a lemma that relates the optimum value of an offline packing instance to the optimum value on a sample of the requests. The proof of this is along the lines of a similar lemma (Lemma 14) in [17], and we present the proof in Appendix G for the sake of completeness.

LEMMA 6.3. *For all $\rho \in (0,1]$, there exists $\eta = O\left(\sqrt{\log(\frac{d}{\rho})}\right)$ such that for all $\delta \in (0,1]$, given a random sample of δT requests, one can compute a quantity $\hat{OPT}$ such that with probability $1-\rho$,*

1. $\hat{OPT} \geq OPT_{\text{SUM}} - \eta\sqrt{OPT_{\text{SUM}}/\delta}$.
2. $\frac{\hat{OPT}}{1+\eta/\sqrt{\delta B}} \leq OPT_{\text{SUM}} + \eta\sqrt{OPT_{\text{SUM}}/\delta}$.

LEMMA 6.4. *Given a random sample of $O(\epsilon^2\log(1/\epsilon))$ fraction of requests, one can compute a quantity Z such that with probability at least $1-\epsilon^2$,*

$$\frac{OPT_{\text{SUM}}}{B} \leq Z \leq \frac{9}{2}\frac{OPT_{\text{SUM}}}{B}.$$

Proof. We use Lemma 6.3 with $\rho = \epsilon^2$ and $\delta = 4\eta^2\epsilon^2/\log(d)$. Then, from the assumption that $\min\{B, \mathrm{OPT}_{\text{SUM}}\} \geq \log(d)/\epsilon^2$, we have that $\delta \geq 4\eta^2/\mathrm{OPT}_{\text{SUM}}$, and $\delta \geq 4\eta^2/B$. Therefore, we get that with probability at least $1-\epsilon^2$,

$$\begin{aligned}\hat{\mathrm{OPT}} &\geq \mathrm{OPT}_{\text{SUM}} - \eta\sqrt{\mathrm{OPT}_{\text{SUM}}/\delta}\\ &\geq \mathrm{OPT}_{\text{SUM}} - \mathrm{OPT}_{\text{SUM}}/2 = \mathrm{OPT}_{\text{SUM}}/2.\end{aligned}$$

Also,

$$\begin{aligned}\hat{\mathrm{OPT}} &\leq (1+\eta/\sqrt{\delta B})(\mathrm{OPT}_{\text{SUM}} + \eta\sqrt{\mathrm{OPT}_{\text{SUM}}/\delta})\\ &\leq \frac{3}{2}(\mathrm{OPT}_{\text{SUM}} + \frac{1}{2}\mathrm{OPT}_{\text{SUM}})\\ &\leq \frac{9}{4}\mathrm{OPT}_{\text{SUM}}.\end{aligned}$$

Therefore $Z := 2\hat{\mathrm{OPT}}/B$ satisfies the conclusion of the lemma. Finally, note that $\delta = 4\eta^2\epsilon^2/\log(d) = O\left(\epsilon^2\log(\frac{d}{\epsilon})/\log(d)\right) = O(\epsilon^2\log(1/\epsilon))$.

7 Stronger bounds for smooth functions

We show that when f is a strongly smooth function, and, instead of distance function a strongly smooth function is used to measure regret in constraint violation, then stronger regret bounds of $\tilde{O}(\frac{\log T}{T})$ can be achieved in IID case. Intuitively, this is because as discussed in Section 2, the dual of strongly smooth functions is strongly convex, and for strongly convex/concave functions, stronger logarithmic regret guarantees are provided by online learning algorithms.

More precisely, consider the following *smooth* version of Online Convex Programming problem.

DEFINITION 4. [ONLINE STOCHASTIC SMOOTH CONVEX PROGRAMMING] *Let f be a β-smooth concave function. And, let h be a β-smooth convex function At time t, the algorithm needs to choose $\boldsymbol{v}_t^\dagger \in A_t$ to minimize regret defined as*

$$\begin{aligned}\textit{avg-regret}_1(T) &:= f(\boldsymbol{v}_{avg}^*) - f(\boldsymbol{v}_{avg}^\dagger),\\ \textit{avg-regret}_2(T) &:= h(\boldsymbol{v}_{avg}^\dagger).\end{aligned}$$

Here, $\boldsymbol{v}_{avg}^* = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{v}_t^*, \boldsymbol{v}_{avg}^\dagger = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{v}_t^\dagger$. Also, assume that there exist $\boldsymbol{v}_t \in A_t$ for all t, such that $h(\frac{1}{T}\sum_t\boldsymbol{v}_t) = 0$.

Note that we do not require Lipschitz condition for f or h. We make an additional assumption.

ASSUMPTION 2. *Let ∇_f and ∇_g denote the set of gradients of functions f and g, respectively, on domain $[0,1]^d$, i.e.,*

$$\begin{aligned}\nabla_f &= \{\nabla f(\boldsymbol{x}) : \boldsymbol{x} \in [0,1]^d\}, \text{ and },\\ \nabla_g &= \{\nabla g(\boldsymbol{x}) : \boldsymbol{x} \in [0,1]^d\}.\end{aligned}$$

Assume that the sets $cl(\nabla_f)$ and $cl(\nabla_g)$ are convex and easy to project upon. Here $cl(S)$ denotes the closure of set S.

This assumption is true for many natural concave utility and convex risk functions, in particular, for all separable smooth functions. Now, an algorithm similar to Algorithm 5.1 can be used for this problem. One change we make is that we perform online learning for g_t and ψ_t on domain ∇_g and ∇_f, respectively, which is possible because from Assumption 2, these domains are convex and easy to project upon.

ALGORITHM 7.1. (ONLINE SMOOTH CP)
Initialize $\boldsymbol{\theta}_1, \boldsymbol{\phi}_1$.
for all $t = 1, \ldots, T$ **do**
Choose vector
$$\boldsymbol{v}_t^\dagger = \arg\max_{\boldsymbol{v} \in A_t} -\boldsymbol{\phi}_t \cdot \boldsymbol{v} - 2Z\boldsymbol{\theta}_t \cdot \boldsymbol{v}.$$
Choose $\boldsymbol{\theta}_{t+1}$ by doing an OCO update for $g_t(\boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \boldsymbol{v}_t^\dagger - h_S(\boldsymbol{\theta})$ over domain ∇_g.
Choose $\boldsymbol{\phi}_{t+1}$ by doing an OCO update for $\psi_t(\boldsymbol{\phi}) = \boldsymbol{\phi} \cdot \boldsymbol{v}_t^\dagger - (-f)^*(\boldsymbol{\phi})$ over domain ∇_f.
end for

THEOREM 7.1. *Under Assumption 2, and given Z that satisfies Assumption 1, Algorithm 7.1 achieves the following regret for the Online* Smooth *Convex Programming problem, in the stochastic IID input model.*

$$\begin{aligned}\mathbb{E}[\textit{avg-regret}_1(T)] &= Z \cdot O\left(\frac{C\log(T)}{T}\right),\\ \mathbb{E}[\textit{avg-regret}_2(T)] &= O\left(\frac{C\log(T)}{T}\right),\end{aligned}$$

where $C = \beta||\mathbf{1}_d||^2$.

Proof. The proof follows from the proof of Theorem 2.1 on observing that stronger OCO regret bounds of $O(\log(T))$ are available for strongly convex functions. More precisely, in case of IID inputs, the proof of Theorem 2.1 can be followed as it is to achieve the following regret bounds. (These are same as in the detailed statement of Theorem 2.1, provided in Appendix E, but with $\mathcal{Q}(T) = 0$ due to IID assumption.)

$$\mathbb{E}[\text{avg-regret}_1(T)] \le \frac{Z}{T} \cdot O(\mathcal{R}(T)) + O(\frac{\mathcal{R}'(T)}{T}),$$

$$\mathbb{E}[\text{avg-regret}_2(T)] \le \frac{1}{T} \cdot O\left(\mathcal{R}(T)\right) + \frac{1}{Z} O(\frac{\mathcal{R}'(T)}{T}),$$

Here $\mathcal{R}(T)$ is *OCO* regret for the problem of maximizing concave function $g_t(\boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \boldsymbol{v}_t - h^*(\boldsymbol{\theta})$, $\mathcal{R}'(T)$ is *OCO* regret for the problem of maximizing concave function $\psi_t(\boldsymbol{\phi}) = \boldsymbol{\phi} \cdot \boldsymbol{v}_t - (-f)^*(\boldsymbol{\phi})$. Now, using Lemma 3.2, given that h and f are β-strongly smooth, g_t and ψ_t are $\frac{1}{\beta}$-strongly concave over domain ∇_g and ∇_f respectively. Also, the gradient of these functions is some $\boldsymbol{v} \in [0,1]^d$, so that the norms of gradients are bounded by $\|\mathbf{1}_d\|$.

Therefore, using online learning guarantees for smooth functions from Lemma 3.5, along with $G = \|\mathbf{1}_d\|, H = 1/\beta$, we get $\mathcal{R}(T) = O(\|\mathbf{1}_d\|^2 \beta \log T)$, and $\mathcal{R}'(T) = O(\|\mathbf{1}_d\|^2 \beta \log T)$. The theorem statement is obtained by substituting these OCO regret bounds in above.

In above, observe that Assumption 2 was required because Lemma 3.2 provided strong convexity of $g_t(\cdot)$ and $\psi_t(\cdot)$ only on the domains ∇_g and ∇_f, respectively. We conjecture that it is possible to remove this assumption to get similar regret guarantees for the smooth case.

References

[1] Jacob Abernethy, Peter L. Bartlett, and Elad Hazan. Blackwell approachability and low-regret learning are equivalent. In *COLT*, 2011.

[2] Gagan Aggarwal, Gagan Goel, Chinmay Karande, and Aranyak Mehta. Online vertex-weighted bipartite matching and single-bid budgeted allocations. In *SODA*, 2011.

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

[4] Shipra Agrawal and Nikhil R. Devanur. Bandits with concave rewards and convex knapsacks. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, 2014.

[5] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(6):121–164, 2012.

[6] Moshe Babaioff, Shaddin Dughmi, Robert Kleinberg, and Aleksandrs Slivkins. Dynamic pricing with limited supply. In *EC*, 2012.

[7] Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. In *FOCS*, pages 207–216, 2013.

[8] Bahman Bahmani and Michael Kapralov. Improved bounds for online stochastic matching. In *ESA*, pages 170–181, 2010.

[9] David Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

[10] Niv Buchbinder, Kamal Jain, and Joseph Seffi Naor. Online primal-dual algorithms for maximizing ad-auctions revenue. In *Proceedings of the 15th Annual European Conference on Algorithms*, ESA'07, 2007.

[11] Deepayan Chakrabarti and Erik Vee. Traffic shaping to optimize ad delivery. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, 2012.

[12] Peiji Chen, Wenjing Ma, Srinath Mandalapu, Chandrashekhar Nagarjan, Jayavel Shanmugasundaram, Sergei Vassilvitskii, Erik Vee, Manfai Yu, and Jason Zien. Ad serving using a compact allocation plan. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, 2012.

[13] Xiao Chen and Zizhuo Wang. A near-optimal dynamic learning algorithm for online matching problems with concave returns. http://arxiv.org/abs/1307.5934, 2013.

[14] Ye Chen, Pavel Berkhin, Bo Anderson, and Nikhil R. Devanur. Real-time bidding algorithms for performance-based display ad allocation. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, 2011.

[15] Nikhil R. Devanur and Thomas P. Hayes. The adwords problem: online keyword matching with budgeted bidders under random permutations. In *EC*, 2009.

[16] Nikhil R. Devanur and Kamal Jain. Online matching with concave returns. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, 2012.

[17] Nikhil R. Devanur, Kamal Jain, Balasubramanian Sivan, and Christopher A. Wilkens. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. Full version, accessible from http://research.microsoft.com/en-us/um/people/bsivan/, 2011.

[18] Nikhil R. Devanur, Kamal Jain, Balasubramanian Sivan, and Christopher A. Wilkens. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. In *EC*, 2011.

[19] Nikhil R. Devanur, Zhiyi Huang, Nitish Korula, Vahab S. Mirrokni, and Qiqi Yan. Whole-page optimization and submodular welfare maximization with online bidders. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, 2013.

[20] J. Feldman, N. Korula, V. Mirrokni, S. Muthukrishnan, and M. Pal. Online ad assignment with free disposal. In *WINE*, 2009.

[21] Jon Feldman, Aranyak Mehta, Vahab Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *FOCS '09: Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science*, 2009.

[22] Jon Feldman, Monika Henzinger, Nitish Korula, Vahab S. Mirrokni, and Cliff Stein. Online stochastic packing applied to display ad allocation. In *Proceedings of the 18th Annual European Conference on Algorithms: Part I*, ESA'10, 2010.

[23] Jon Feldman, Monika Henzinger, Nitish Korula, Vahab S. Mirrokni, and Clifford Stein. Online stochastic ad allocation: Efficiency and fairness. *CoRR*, abs/1001.5076, 2010.

[24] Arpita Ghosh, Randolph Preston McAfee, Kishore Papineni, and Sergei Vassilvitskii. Bidding for representative allocations for display advertising. In *WINE*, 2009.

[25] Gagan Goel and Aranyak Mehta. Online budgeted matching in random input models with applications to adwords. In *SODA '08: Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, 2008.

[26] Anupam Gupta and Marco Molinaro. How the Experts Algorithm Can Help Solve LPs Online. *Algorithms - ESA 2014, Lecture Notes in Computer Science*, 8737:517–529, 2014.

[27] Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Mach. Learn.*, 69(2-3), December 2007.

[28] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[29] Sham M. Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization. Technical report, Toyota Technological Institute - Chicago, USA, 2009. http://ttic.uchicago.edu/~shai/papers/KakadeShalevTewari09.pdf.

[30] Chinmay Karande, Aranyak Mehta, and Pushkar Tripathi. Online bipartite matching with unknown distributions. In *STOC*, 2011.

[31] Chinmay Karande, Aranyak Mehta, and Ramakrishnan Srikant. Optimizing budget constrained spend in search advertising. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, 2013.

[32] R. M. Karp, U. V. Vazirani, and V. V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the Twenty-second Annual ACM Symposium on Theory of Computing*, STOC '90, 1990.

[33] Thomas Kesselheim, Andreas Tönnis, Klaus Radke, and Berthold Vöcking. Primal beats dual on online packing LPs in the random-order model. In *STOC*, 2014.

[34] R. Kleinberg. A multiple-choice secretary algorithm with applications to online auctions. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete algorithms*, pages 630–631, January 2005.

[35] Robert Kleinberg, Alex Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *STOC*, 2008.

[36] Mohammad Mahdian and Qiqi Yan. Online bipartite matching with random arrivals: an approach based on strongly factor-revealing LPs. In *STOC*, 2011.

[37] Mohammad Mahdian, Hamid Nazerzadeh, and Amin Saberi. Online optimization with uncertain information. *ACM Trans. Algorithms*, 8(1), January 2012.

[38] Vahideh Manshadi, Shayan Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. In *SODA*, 2011.

[39] Aranyak Mehta, Amin Saberi, Umesh V. Vazirani, and Vijay V. Vazirani. Adwords and generalized online matching. *J. ACM*, 54(5), 2007.

[40] Vahab S. Mirrokni, Shayan Oveis Gharan, and Morteza Zadimoghaddam. Simultaneous approximations for adversarial and stochastic online budgeted allocation. In *Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, 2012.

[41] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2012.

[42] Erik Vee, Sergei Vassilvitskii, and Jayavel Shanmugasundaram. Optimal online assignment with forecasts. In *EC '10: Proceedings of the 11th ACM conference on Electronic commerce*, 2010.

A Concentration Inequalities

LEMMA A.1. *[28], Theorem 4 Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a finite population of N real points, $X_1, \ldots, X_n$ denote a random sample without replacement from $\mathcal{X}$, and $Y_1, \ldots, Y_n$ denote a random sample with replacement from $\mathcal{X}$. If $\ell : \mathbb{R} \to \mathbb{R}$ is continuous and convex, then*

$$\mathbb{E}[\ell(\sum_{t=1}^{n} X_t)] \leq \mathbb{E}[\ell(\sum_{t=1}^{n} Y_t)].$$

LEMMA A.2. *[28] Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a finite population of N real points, $X_1, \ldots, X_n$ denote a random sample without replacement from $\mathcal{X}$. Let $a = \min_{1\leq i\leq N} x_i$, $b = \max_{1\leq i\leq N} x_i$ and $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$. Then, for all $\epsilon > 0$,*

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq \epsilon\right) \leq \exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right).$$

LEMMA A.3. *(Multiplicative version) Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a finite population of N real points, and $X_1, \ldots, X_n$ denote a random sample without replacement from $\mathcal{X}$. Let $a = \min_{1\leq i\leq N} x_i$, $b = \max_{1\leq i\leq N} x_i$ and $\mu = \frac{n}{N}\sum_{i=1}^{N} X_i$. Then, for all $\epsilon > 0$,*

$$\Pr\left(|\sum_{i=1}^{n} X_i - \mu| \geq \epsilon\mu\right) \leq \exp\left(-\frac{\mu\epsilon^2}{3(b-a)^2}\right).$$

COROLLARY A.1. *(to Lemma A.3) Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a finite population of N real points, and $X_1, \ldots, X_n$ denote a random sample without replacement from $\mathcal{X}$. Let $a = \min_{1\leq i\leq N} x_i$, $b = \max_{1\leq i\leq N} x_i$ and $\mu = \frac{n}{N}\sum_{i=1}^{N} X_i$. Then, for all $\rho > 0$, with probability at least $1-\rho$,*

$$|\sum_{i=1}^{n} X_i - \mu| \leq (b-a)\sqrt{3\mu\log(1/\rho)}$$

Proof. Given $\rho > 0$, use Lemma A.3 with

$$\epsilon = (b-a)\sqrt{\frac{3\log(1/\rho)}{\mu}},$$

to get that the probability of the event $|\sum_{i=1}^{n} X_i - \mu| > \epsilon\mu = (b-a)\sqrt{3\mu\log(1/\rho)}$ is at most

$$\exp\left(-\frac{\mu\epsilon^2}{3(b-a)^2}\right) = \exp\left(-\log(1/\rho)\right) = \rho.$$

LEMMA A.4. *[35, 6, 7] Consider a probability distribution with values in $[0,1]$, and expectation ν. Let $\hat{\nu}$ be the average of N independent samples from this distribution. Then, with probability at least $1 - e^{-\Omega(\gamma)}$, for all $\gamma > 0$,*

$$|\hat{\nu} - \nu| \leq \mathrm{rad}(\hat{\nu}, N) \leq 3\mathrm{rad}(\nu, N), \tag{A.1}$$

where $\mathrm{rad}(\nu, N) = \sqrt{\frac{\gamma\nu}{N}} + \frac{\gamma}{N}$. More generally this result holds if $X_1, \ldots, X_N \in [0,1]$ are random variables, $N\hat{\nu} = \sum_{t=1}^{N} X_t$, and $N\nu = \sum_{t=1}^{N} \mathbb{E}[X_t|X_1, \ldots, X_{t-1}]$.

B Preliminaries

B.1 Strong smoothness/Strong convexity duality.

Proof of Lemma 3.2 Given h is convex and β-strong smooth with respect to norm $||\cdot||$. We prove that h^*, defined as

$$h^*(\boldsymbol{\theta}) = \max_{\boldsymbol{y}\in[0,1]^d} \{\boldsymbol{y}\cdot\boldsymbol{\theta} - h(\boldsymbol{y})\},$$

is $\frac{1}{\beta}$-strongly convex with respect to norm $||\cdot||_*$ on domain $\nabla_h = \{\nabla h(\boldsymbol{x}) : \boldsymbol{x} \in [0,1]^d\}$.

For any $\boldsymbol{\theta}, \boldsymbol{\phi} \in \nabla_h$, $\boldsymbol{\theta} = \nabla h(\boldsymbol{z}), \boldsymbol{\phi} = \nabla h(\boldsymbol{x})$ for some $\boldsymbol{z}, \boldsymbol{x} \in [0,1]^d$. And, therefore,

$$\begin{aligned}
& h^*(\boldsymbol{\theta}) - h^*(\boldsymbol{\phi}) - \boldsymbol{x}\cdot(\boldsymbol{\theta}-\boldsymbol{\phi}) \\
= \ & h^*(\nabla h(\boldsymbol{z})) - h^*(\nabla h(\boldsymbol{x})) - \boldsymbol{x}\cdot(\nabla h(\boldsymbol{z}) - \nabla h(\boldsymbol{x})) \\
= \ & \boldsymbol{z}\cdot\nabla h(\boldsymbol{z}) - h(\boldsymbol{z}) \\
& -(\boldsymbol{x}\cdot\nabla h(\boldsymbol{x}) - h(\boldsymbol{x})) - \boldsymbol{x}\cdot(\nabla h(\boldsymbol{z}) - \nabla h(\boldsymbol{x})) \\
= \ & \boldsymbol{z}\cdot\nabla h(\boldsymbol{z}) - h(\boldsymbol{z}) + h(\boldsymbol{x}) - \boldsymbol{x}\cdot\nabla h(\boldsymbol{z}) \\
= \ & (\boldsymbol{z}-\boldsymbol{x})\cdot(\nabla h(\boldsymbol{z}) - \nabla h(\boldsymbol{x})) \\
& -(h(\boldsymbol{z}) - h(\boldsymbol{x}) - \nabla h(\boldsymbol{x})(\boldsymbol{z}-\boldsymbol{x})) \\
= \ & (\boldsymbol{z}-\boldsymbol{x})\cdot(\nabla h(\boldsymbol{z}) - \nabla h(\boldsymbol{x})) - g(\boldsymbol{z}-\boldsymbol{x}),
\end{aligned} \tag{B.2}$$

where we define

$$g(\boldsymbol{y}) := h(\boldsymbol{x}+\boldsymbol{y}) - h(\boldsymbol{x}) - (\nabla h(\boldsymbol{x}))\cdot\boldsymbol{y}.$$

Now, for any φ,

$$\begin{aligned}
g^*(\varphi) &:= \sup_{\boldsymbol{y}} \varphi\cdot\boldsymbol{y} - g(\boldsymbol{y}) \\
&= \varphi\cdot\boldsymbol{y}^* - g(\boldsymbol{y}^*)
\end{aligned}$$

where $\boldsymbol{y}^*$ is such that $\varphi = \nabla g(\boldsymbol{y}^*) = \nabla h(\boldsymbol{x}+\boldsymbol{y}^*) - \nabla h(\boldsymbol{x})$. Therefore, for $\varphi = \nabla h(\boldsymbol{z}) - \nabla h(\boldsymbol{x})$, $\boldsymbol{y}^* = \boldsymbol{z}-\boldsymbol{x}$, so that,

$$\begin{aligned}
g^*(\nabla h(\boldsymbol{z}) - \nabla h(\boldsymbol{x})) &= (\nabla h(\boldsymbol{z}) - \nabla h(\boldsymbol{x}))\cdot(\boldsymbol{z}-\boldsymbol{x}) \\
&\quad -g(\boldsymbol{z}-\boldsymbol{x}).
\end{aligned}$$

Substituting in (B.2), we get

$$\begin{aligned}
& h^*(\boldsymbol{\theta}) - h^*(\boldsymbol{\phi}) - \boldsymbol{x}\cdot(\boldsymbol{\theta}-\boldsymbol{\phi}) \\
= \ & g^*(\nabla h(\boldsymbol{z}) - \nabla h(\boldsymbol{x})) \\
= \ & g^*(\boldsymbol{\theta}-\boldsymbol{\phi})
\end{aligned}$$

By smoothness assumption, $g(\boldsymbol{y}) \le \frac{\beta}{2}||\boldsymbol{y}||^2$. This implies that $g^*(\boldsymbol{\theta}) \ge \frac{1}{2\beta}||\boldsymbol{\theta}||_*^2$ because the conjugate of β times half squared norm is $1/\beta$ times half squared of the dual norm. This gives

$$h^*(\boldsymbol{\theta}) - h^*(\boldsymbol{\phi}) - \boldsymbol{x}\cdot(\boldsymbol{\theta}-\boldsymbol{\phi}) \ge \frac{1}{2\beta}\|\boldsymbol{\theta}-\boldsymbol{\phi}\|_*^2.$$

This completes the proof.

B.2 Online learning. A popular algorithm for OCO is the online mirror descent (OMD) algorithm. The OMD algorithm with regularizer $R(\boldsymbol{\theta})$ uses the following fast update rule to select player's decision $\boldsymbol{\theta}_{t+1}$ for this problem:

$$\begin{aligned}
\boldsymbol{\theta}_{t+1} &= \arg\max_{\boldsymbol{\theta}\in W} \frac{1}{\eta}R(\boldsymbol{\theta}) - \boldsymbol{\theta}\cdot\boldsymbol{y}_{t+1}, \text{ where} \\
\boldsymbol{y}_{t+1} &= \boldsymbol{y}_t - z_t, \text{ and } z_t \in \partial g_t(\boldsymbol{\theta}_t)
\end{aligned} \tag{B.3}$$

The maximization problem in above is particularly simple when domain W is of form $||\boldsymbol{\theta}|| \le \gamma$, and this is the main use case of this algorithm in this paper. Further, for domain W of form $||\boldsymbol{\theta}||_2 \le L$, and $R(\boldsymbol{\theta}) = ||\boldsymbol{\theta}||_2^2$, this simply becomes online gradient descent. OMD has the following guarantees for this problem:

Lemma B.1. *[41]*

$$\mathcal{R}(T) \le \frac{D}{\eta} + \eta T G^2,$$

where $D = (\max_{\boldsymbol{\theta}''} R(\boldsymbol{\theta}'') - \min_{\boldsymbol{\theta}'\in W} R(\boldsymbol{\theta}'))$, $\frac{1}{T}\sum_{t=1}^T ||z_t||^2 \le G$ *for* $z_t \in \partial g_t(\boldsymbol{\theta}_t)$*, and* R *is a* 1*-strongly-convex function with respect to norm* $||\cdot||_*$.

Now, to derive Corollary 3.1, observe that for $W = \{||\boldsymbol{\theta}||_2 \le L\}$, Euclidean regularizer $R(\boldsymbol{\theta}) = ||\boldsymbol{\theta}||_2^2$ gives $\mathcal{R}(T) \le LG\sqrt{T}$, with $G^2 = d \ge \frac{1}{T}\sum_{t=1}^T ||z_t||_2^2$, when $z_t \in [0,1]^d$. And, for $W = \{||\boldsymbol{\theta}||_1 \le L, \boldsymbol{\theta} > 0\}$, entropic regularizer $R(\boldsymbol{\theta}) = \sum_i \boldsymbol{\theta}_i \log \boldsymbol{\theta}_i$ gives $\mathcal{R}(T) \le G\sqrt{LT\log(d)}$, where $G^2 = 1 \ge \frac{1}{T}\sum_{t=1}^T ||z_t||_\infty^2$, when $z_t \in [0,1]^d$.

C Sampling without replacement bounds for Section 4

Proof of Equation (4.4)**.** Let $\boldsymbol{\omega} = \mathbb{E}[\mathbf{w}_{t,\sigma}] = \mathbb{E}[\boldsymbol{v}_t^*]$. To bound the quantity $\mathbb{E}[||\mathbf{w}_{t,\pi} - \boldsymbol{\omega}||]$, note that $\mathbf{w}_{t,\pi}$ can be viewed as the average of t vectors sampled uniformly *without replacement* from the ground set $\{\boldsymbol{v}_{X_1}, \ldots, \boldsymbol{v}_{X_T}\}$ of T vectors.

Now, let $w_{t,\pi,j}$ denote the j^{th} component of vector $\mathbf{w}_{t,\pi}$. Then, by applying concentration bounds from Corollary A.1, we get that

$$|w_{t,\pi,j} - \omega_j| \le \sqrt{\frac{3\omega_j \log(d/\rho)}{t}},$$

with probability $1 - \frac{\rho}{d}$ for all $\rho \in (0,1)$. From the condition $\boldsymbol{\omega} = \mathbb{E}[\boldsymbol{v}_t^*] \in S$, we have $\omega_j \le \max_{\boldsymbol{v}\in S} v_j \le s$. Taking union bound over d, for every $\rho \in (0,1)$, we have that with probability $1-\rho$,

$$\|\mathbf{w}_{t,\pi} - \boldsymbol{\omega}\| \le \|\mathbf{1}_d\|\sqrt{\frac{3s\log(d/\rho)}{t}}.$$

And, integrating over ρ, we obtain,

$$\mathbb{E}[\|\mathbf{w}_{t,\pi} - \boldsymbol{\omega}\|] \le O(\|\mathbf{1}_d\|\sqrt{\frac{s\log(d)}{t}}).$$

High Probability bounds. For high probability bounds, firstly from Equation (4.2) and (4.3),

$$\begin{aligned}\sum_t \mathbb{E}[g_t(\boldsymbol{\theta}_t)|\mathcal{F}_{t-1}] &\leq \sum_t \|\mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}_t^*]\| \\ &= \sum_t \|\mathbf{w}_{t,\pi} - \boldsymbol{\omega}\|,\end{aligned}$$

for uniform at random orderings π.

Then, as in above, using Corollary A.1 we obtain that for every t, with probability $1 - \frac{\rho}{T}$

$$||\mathbf{w}_{t,\pi} - \boldsymbol{\omega}|| \leq \|\mathbf{1}_d\|\sqrt{\frac{3s\log(dT/\rho)}{t}}.$$

Taking union bound over $t = 1, \ldots, T$, and summing over t we obtain that with probability $1 - \rho$,

$$\begin{aligned}\sum_t \mathbb{E}[g_t(\boldsymbol{\theta}_t)|\mathcal{F}_{t-1}] &\leq \sum_t \|\mathbf{w}_{t,\pi} - \boldsymbol{\omega}\| \\ &= O(\|\mathbf{1}_d\|\sqrt{T\log(dT/\rho)}).\end{aligned}$$

Now, using Lemma A.4 for dependent random variables $X_t = g_t(\boldsymbol{\theta}_t)$, with $|X_t| = |\boldsymbol{\theta}_t \cdot \boldsymbol{v}_t^\dagger - h_S(\boldsymbol{\theta}_t)| \leq \|\mathbf{1}_d\|$, we have,

$$\sum_t g_t(\boldsymbol{\theta}_t) - \sum_t \mathbb{E}[g_t(\boldsymbol{\theta}_t)|\mathcal{F}_{t-1}] \leq O(\|\mathbf{1}_d\|\sqrt{T\log(1/\rho)})$$

with probability at least $1 - \rho$.

Combining the above observations, we obtain that with probability $1 - \rho$,

$$\sum_t g_t(\boldsymbol{\theta}_t) \leq O(\|\mathbf{1}_d\|\sqrt{T\log(dT/\rho)}).$$

D Proof of Lemma 5.1

The offline optimal solution needs to pick $\boldsymbol{v}_t^* \in \text{Conv}(X_t)$ to serve request type X_t, where $\text{Conv}(X_t)$ denotes the convex hull of set X_t. Therefore, OPT^δ is defined as

$$\begin{aligned}\text{OPT}^\delta \quad :=& \quad \max_{\{\boldsymbol{v}_t \in \text{Conv}(X_t)\}} \quad \begin{matrix} f(\frac{1}{T}\sum_t \boldsymbol{v}_t) \\ d(\frac{1}{T}\sum_t \boldsymbol{v}_t, S) \leq \delta \end{matrix} \\ =& \min_{\lambda\geq 0} \max_{\{\boldsymbol{x}=\frac{1}{T}\sum_t \boldsymbol{v}_t, \boldsymbol{v}_t\in\text{Conv}(X_t)\}} \{ \\ & f(\boldsymbol{x}) - \lambda d(\boldsymbol{x}, S) + \delta\lambda \ \} \\ =& \min_{\lambda\geq 0} \max_{\{\boldsymbol{x}=\frac{1}{T}\sum_t \boldsymbol{v}_t, \boldsymbol{v}_t\in\text{Conv}(X_t)\}} \min_{||\boldsymbol{\phi}||_*\leq L, ||\boldsymbol{\theta}||_*\leq 1} \{ \\ & f^*(\boldsymbol{\phi}) - \boldsymbol{\phi}\cdot\boldsymbol{x} - \lambda\boldsymbol{\theta}\cdot\boldsymbol{x} + \lambda h_S(\boldsymbol{\theta}) + \delta\lambda \ \} \\ =& \min_{\lambda\geq 0, ||\boldsymbol{\phi}||_*\leq L, ||\boldsymbol{\theta}||_*\leq 1} \max_{\{\boldsymbol{x}=\frac{1}{T}\sum_t \boldsymbol{v}_t, \boldsymbol{v}_t\in\text{Conv}(X_t)\}} \{ \\ & f^*(\boldsymbol{\phi}) - \boldsymbol{\phi}\cdot\boldsymbol{x} - \lambda\boldsymbol{\theta}\cdot\boldsymbol{x} + \lambda h_S(\boldsymbol{\theta}) + \delta\lambda \ \} \\ =& \min_{\lambda\geq 0, ||\boldsymbol{\phi}||_*\leq L, ||\boldsymbol{\theta}||_*\leq 1} \{ \ f^*(\boldsymbol{\phi}) + \lambda h_S(\boldsymbol{\theta}) \\ & + \frac{1}{T}\sum_{t=1}^T h_{\text{Conv}(X_t)}(-\boldsymbol{\phi} - \lambda\boldsymbol{\theta}) + \delta\lambda \ \}\end{aligned} \tag{D.4}$$

where, recall that for any convex set X, $h_X(\boldsymbol{\theta})$ was defined as $h_X(\boldsymbol{\theta}) := \max_{\boldsymbol{v}\in X} \boldsymbol{\theta}\cdot\boldsymbol{v}$. Because a linear function is maximized at a vertex of a convex set, $h_{\text{Conv}(X_t)}(-\boldsymbol{\phi} - \lambda\boldsymbol{\theta})$ is same as $h_{X_t}(-\boldsymbol{\phi} - \lambda\boldsymbol{\theta})$. This allows us to rewrite the expression for OPT^δ as

$$\begin{aligned}\text{OPT}^\delta \quad =& \min_{\lambda\geq 0, ||\boldsymbol{\phi}||_*\leq L, ||\boldsymbol{\theta}||_*\leq 1} \{ \ f^*(\boldsymbol{\phi}) + \lambda h_S(\boldsymbol{\theta}) \\ & + \frac{1}{T}\sum_{t=1}^T h_{X_t}(-\boldsymbol{\phi} - \lambda\boldsymbol{\theta}) + \delta\lambda \ \}\end{aligned} \tag{D.5}$$

From above, it is clear that OPT^δ is a non-decreasing concave function of δ, with gradient as $\lambda^*(\delta) \geq 0$, where $\lambda^*(\delta)$ is the optimal dual variable corresponding to the distance constraint. And,

$$\lim_{\delta\to 0} \frac{\text{OPT}^\delta - \text{OPT}}{\delta} = \lambda^*$$

where λ^* is the optimal dual variable for OPT (i.e., the case of $\delta = 0$). This proves the lemma.

E Proof of Theorem 2.1

We provide proof of a more detailed theorem statement.

THEOREM E.1. *Given Z that satisfies Assumption 1, Algorithm 5.1 achieves the following regret bounds for online stochastic CP, in RP model:*

$$\begin{aligned}&\mathbb{E}[\textit{avg-regret}_1(T)] \leq \\ &\tfrac{(Z+L)}{T}\cdot O\left(\mathcal{R}(T) + \mathcal{Q}(T)\right) + O(\tfrac{\mathcal{R}'(T)}{T}),\end{aligned}$$

$$\begin{aligned}&\mathbb{E}[\textit{avg-regret}_2(T)] \leq \\ &\tfrac{1}{T}\cdot O\left(\mathcal{R}(T) + \mathcal{Q}(T)\right) + \tfrac{1}{(Z+L)}O(\tfrac{\mathcal{R}'(T)}{T}),\end{aligned}$$

where $\mathcal{Q}(T) = O(||\mathbf{1}_d||\sqrt{sT\log(d)})$, $\mathcal{R}'(T)$ is the regret bound for OCO on $\psi_t(\cdot)$, $\mathcal{R}(T)$ is the regret bound for OCO on $g_t(\cdot)$. And, $s \leq 1$ is the coordinate-wise largest value a vector in S can take.

Then, substituting OCO regret bounds from Corollary 3.1 gives the statement of Theorem 2.1.

Proof. Denote by $(\boldsymbol{v}_t^*)$ the choice made by the offline optimal solution to satisfy request A_t. Then,

$$f(\mathbb{E}[\boldsymbol{v}_t^*]) \geq \text{OPT}, \text{ and } \mathbb{E}[\boldsymbol{v}_t^*] \in S,$$

where expectation is over A_t drawn uniformly at random from $X_1, \ldots, X_T$.

Lemma E.1 provides

$$\begin{aligned}f(\mathbb{E}[\boldsymbol{v}_t^*]) + \frac{1}{T}\sum_t \mathbb{E}[\psi_t(\boldsymbol{\phi}_t) + 2(Z+L)g_t(\boldsymbol{\theta}_t)] \\ \leq (Z+L)\frac{\mathcal{Q}(T)}{T}\end{aligned}$$

where $\mathcal{Q}(T) = O(||\mathbf{1}_d||\sqrt{s\log(d)T})$. Using Fenchel duality and OCO guarantees, it follows that

$$\begin{aligned}\min_{||\boldsymbol{\theta}||_*\leq 1} \frac{1}{T}\sum_t g_t(\boldsymbol{\theta}) &= d(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger, S) \\ &\leq \frac{1}{T}\sum_t g_t(\boldsymbol{\theta}_t) + \frac{1}{T}\mathcal{R}(T),\end{aligned}$$

$$\min_{||\boldsymbol{\phi}||_*\leq L} \psi_t(\boldsymbol{\phi}) = -f(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger) \leq \frac{1}{T}\sum_t \psi_t(\boldsymbol{\theta}_t) + \frac{1}{T}\mathcal{R}'(T).$$

Then, using above observations, along with $f(\mathbb{E}[\boldsymbol{v}_t^*]) \geq \mathrm{OPT}$, we obtain

$$\begin{aligned}&\mathrm{OPT} - \mathbb{E}[f(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger)] + 2(Z+L)\mathbb{E}[d(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger, S)] \\ &\leq \frac{2(Z+L)}{T}(\mathcal{Q}(T)+\mathcal{R}(T)) - \frac{1}{T}\mathcal{R}'(T).\end{aligned}$$

This gives

$$\begin{aligned}\mathbb{E}[f(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger)] &\geq \mathrm{OPT} + 2(Z+L)\mathbb{E}[d(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger, S)] \\ &\quad -\frac{2(Z+L)}{T}(\mathcal{Q}(T)+\mathcal{R}(T)) \\ &\quad -\frac{1}{T}\mathcal{R}'(T) \end{aligned} \tag{E.6}$$

Now, we use Assumption 1, to upper bound the reward obtained by the algorithm in terms of OPT and distance from set S. In particular, we obtain that for $\delta := \mathbb{E}[d(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger, S)]$,

$$\begin{aligned}\mathbb{E}[f(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger)] &\leq f(\mathbb{E}[\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger]) \\ &\leq \mathrm{OPT}^\delta \\ &\leq \mathrm{OPT} + Z\delta \\ &= \mathrm{OPT} + Z\cdot\mathbb{E}[d(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger, S)].\end{aligned} \tag{E.7}$$

Combining the above two inequalities, we obtain

$$\mathbb{E}[d(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger, S)] \leq \frac{2}{T}(\mathcal{R}(T)+\mathcal{Q}(T)) + \frac{1}{(Z+L)}\mathcal{R}'(T).$$

And, from (E.6) (using $\mathbb{E}[d(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger, S)] \geq 0$),

$$\begin{aligned}&\mathbb{E}[f(\frac{1}{T}\sum_t \boldsymbol{v}_t^\dagger)] \\ &\geq \mathrm{OPT} - \frac{2(Z+L)}{T}\cdot(\mathcal{R}(T)+\mathcal{Q}(T)) - \frac{\mathcal{R}'(T)}{T}.\end{aligned} \tag{E.8}$$

This gives the theorem statement.

Lemma E.1.

$$\begin{aligned}&f(\mathbb{E}[\boldsymbol{v}_t^*]) + \frac{1}{T}\sum_t \mathbb{E}[\psi_t(\boldsymbol{\phi}_t) + 2(Z+L)g_t(\boldsymbol{\theta}_t)] \\ &\quad\leq \frac{1}{T}(Z+L)O(||\mathbf{1}_d||\sqrt{sT\log(d)}).\end{aligned}$$

Proof.

$$\begin{aligned}&\psi_t(\boldsymbol{\phi}_t) + 2(Z+L)g_t(\boldsymbol{\theta}_t) \\ &= \boldsymbol{\phi}\cdot\boldsymbol{v}_t^\dagger - (-f)^*(\boldsymbol{\phi}) + 2(Z+L)(\boldsymbol{\theta}_t\cdot\boldsymbol{v}_t^\dagger - h_S(\boldsymbol{\theta}_t)) \\ &\leq \boldsymbol{\phi}\cdot\boldsymbol{v}_t^* - (-f)^*(\boldsymbol{\phi}) + 2(Z+L)(\boldsymbol{\theta}_t\cdot\boldsymbol{v}_t^* - h_S(\boldsymbol{\theta}_t)).\end{aligned}$$

$$\begin{aligned}&\mathbb{E}[\psi_t(\boldsymbol{\phi}_t) + 2(Z+L)g_t(\boldsymbol{\theta}_t)|\mathcal{F}_{t-1}] \\ &\leq \boldsymbol{\phi}_t\cdot\mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}] - (-f)^*(\boldsymbol{\phi}_t) \\ &\qquad +2(Z+L)(\boldsymbol{\theta}_t\cdot\mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}] - h_S(\boldsymbol{\theta}_t)) \\ &\leq -f(\mathbb{E}[\boldsymbol{v}_t^*]) + \boldsymbol{\phi}_t\cdot(\mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}_t^*]) \\ &\qquad +2(Z+L)\boldsymbol{\theta}_t\cdot(\mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}_t^*])\end{aligned}$$

where the last inequality uses $\boldsymbol{\phi}_t\cdot\mathbb{E}[\boldsymbol{v}_t^*] - (-f)^*(\boldsymbol{\phi}_t) \leq -f(\mathbb{E}[\boldsymbol{v}_t^*])$ (using Fenchel duality) and $\boldsymbol{\theta}_t\cdot\mathbb{E}[\boldsymbol{v}_t^*] - h_S(\boldsymbol{\theta}_t) \leq d(\mathbb{E}[\boldsymbol{v}_t^*], S) = 0$. Then, as in proof of Lemma 4.1, $\mathbb{E}[\sum_t ||\mathbb{E}[\boldsymbol{v}_t^*|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{v}_t^*]||]$ can be upper bounded by $O(\sqrt{||\mathbf{1}_d||sT\log(d)})$. Using this along with observation that $||\boldsymbol{\phi}_t||_* \leq L, ||\boldsymbol{\theta}_t||_* \leq 1$, we get the desired lemma statement.

F Estimating the parameter Z

Let Z^* denote the minimum value of Z that satisfies the property in Equation (5.5). As discussed in the proof of Lemma 5.1, $Z^* = \lambda^*$, the value of optimal dual variable corresponding to feasibility constraint. To obtain low regret bounds, ideally we would like to use $Z = Z^*$ in Algorithm 5.1, which would provide the minimum possible regret bound of $O((Z^*+L)\sqrt{\frac{C}{T}})$ in objective according to Theorem 2.1. The regret in constraints does not depend on Z. However, in the absence of knowledge of Z^*, we need to obtain a good enough approximation. Following lemma provides a relaxed condition to be satisfied by Z in order to obtain the same order of regret bounds, as those obtained with $Z = Z^*$.

Lemma F.1. *Assume that $Z \geq 0$ satisfies the following property, for all $\delta \geq 3\gamma$ where $\gamma = ||\mathbf{1}_d||\sqrt{\frac{\log(dT)}{T}}$,*

$$\frac{OPT^\delta - OPT^{2\gamma}}{\delta} \leq Z = O(Z^*+L).$$

Then, Algorithm 5.1 using such a Z will achieve an expected regret bound of $O((Z^*+L)\gamma)$ in objective, and $O(\gamma)$ in constraints.

To compare with Theorem 2.1, note that $\gamma = O(\sqrt{\frac{C\log(T)}{T}})$, therefore, using such a Z degrades the regret bounds by only an $O(\sqrt{\log(T)})$ factor.

Proof. Recall that in the proof of Theorem 2.1, the condition $\mathrm{OPT}^\delta \leq \mathrm{OPT} + Z\delta$ was used in the following way. We had the inequality,

$$\begin{aligned} \mathrm{OPT}^{\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)]} &\geq \mathbb{E}[f(\boldsymbol{v}^\dagger_{\text{avg}})] \\ &\geq \mathrm{OPT} + 2(Z+L)\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] \\ &\quad -\ell(T), \end{aligned} \tag{F.9}$$

where $\ell(T) = O((Z+L)\sqrt{\frac{C}{T}})$. Then, we applied $\mathrm{OPT}^{\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)]} \leq \mathrm{OPT} + Z\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)]$, to obtain $\mathrm{OPT} + Z\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] \geq \mathrm{OPT} + 2(Z+L)\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] - \ell(T)$, yielding $\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] \leq \frac{1}{(Z+L)}O(\ell(T)) = O(\sqrt{\frac{C}{T}})$.

Now, we will show that it suffices to have $Z \geq \frac{\mathrm{OPT}^\delta - \mathrm{OPT}^{2\gamma}}{\delta}$, for $\delta > 3\gamma$ to obtain the given regret bounds.

We first bound $\mathbb{E}[\text{avg-regret}_2(T)] = \mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)]$. Starting with Equation F.9, observe that if $\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] \leq 3\gamma$, then the distance is bounded by $O(\gamma)$ as required anyway, therefore, assume that $\delta := \mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] \geq 3\gamma$. Then, from the given property of Z we have $\mathrm{OPT}^{\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)]} = \mathrm{OPT}^\delta \leq \mathrm{OPT}^{2\gamma} + Z\delta = \mathrm{OPT}^{2\gamma} + Z\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)]$. Substituting back in Equation (F.9), we get

$$\begin{aligned} &\mathrm{OPT}^{2\gamma} + Z\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] \\ \geq\ &\mathrm{OPT} + 2(Z+L)\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] - \ell(T) \end{aligned}$$

which gives

$$\begin{aligned} (Z+L)\mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] &\leq \ell(T) + \mathrm{OPT}^{2\gamma} - \mathrm{OPT} \\ &\leq \ell(T) + 2Z^*\gamma \\ &= O((Z+L)\gamma) + 2Z^*\gamma \end{aligned}$$

Then, using $Z = O(Z^* + L)$, we get

$$\begin{aligned} \mathbb{E}[\text{avg-regret}_2(T)] &= \mathbb{E}[d(\boldsymbol{v}^\dagger_{\text{avg}},S)] \\ &= O(\gamma) = O(\sqrt{\frac{C\log(T)}{T}}). \end{aligned}$$

The bound on $\mathbb{E}[\text{avg-regret}_1(T)]$ depends only on the upper bound on Z used, and $Z = O(Z^*+L)$ makes this regret bound to be $O((Z^*+L)\sqrt{\frac{C}{T}})$.

Next, we provide method for estimating a Z that satisfies the property stated in Lemma F.1. Define

$$\hat{\mathrm{OPT}}^\delta(n) = \max_{\{\boldsymbol{v}_t \in \mathrm{Conv}(A_t)\}} \quad f(\tfrac{1}{n}\textstyle\sum_{t=1}^n \boldsymbol{v}_t) \quad d(\tfrac{1}{n}\textstyle\sum_{t=1}^n \boldsymbol{v}_t, S) \leq \delta \tag{F.10}$$

with $\hat{\mathrm{OPT}}(n)$ denoting $\hat{\mathrm{OPT}}^\delta(n)$ for $\delta = 0$. We will divide the timeline into phases of size $1, 1, 2^1, 2^2, ..., 2^r, \ldots$. Note that phase $r \geq 2$ consists of $T_r = 2^{r-2}$ time steps, and there are T_r time steps before phase r. The first phase of a single step, we make an arbitrary choice. Then, in every phase $r \geq 2$, we will rerun the algorithm, using Z constructed using observations from the previous T_r time steps as

$$Z := \frac{(\hat{\mathrm{OPT}}^{4\gamma}(T_r) - \hat{\mathrm{OPT}}^{\gamma}(T_r))}{\gamma} + 2L \tag{F.11}$$

with $\gamma = ||\mathbf{1}_d||\sqrt{\frac{\log(dT_r)}{T_r}}$.

Algorithm F.1. [Algorithm for online CP with Z estimation]

Choose any option in the first step.
for all phases $r = 2, ..., \log(T)+1$ **do**
  COMPUTE Z using observations in steps 1 to $T_r = 2^{r-2}$ as

$$Z = \frac{(\hat{\mathrm{OPT}}^{4\gamma}(T_r) - \hat{\mathrm{OPT}}^{\gamma}(T_r))}{\gamma} + 2L$$

  with $\gamma = ||\mathbf{1}_d||\sqrt{\frac{\log(dT_r)}{T_r}}$.
  Run Algorithm 5.1 for T_r steps $t = \{T_r + 1, \ldots, 2T_r\}$ of phase r using Z as computed above.
end for

We prove the following lemma regarding the estimate Z used in above. Here we use the observation that in RP model, the first n time steps provide a random sample of observations from the T observations.

Lemma F.2. *For all $\rho > 0$ and for all natural numbers n, let $\gamma = ||\mathbf{1}_d||\sqrt{\frac{\log(d/\rho)}{n}}$, and*

$$Z := \frac{(\hat{OPT}^{4\gamma}(n) - \hat{OPT}^{\gamma}(n))}{\gamma} + 2L.$$

Then, for all $\delta > 3\gamma$, with probability $1 - O(\rho)$,

$$\frac{(OPT^\delta - OPT^{2\gamma})}{\delta} \leq Z \leq O(L + Z^*).$$

The proof of above lemma is provided later. We now state the regret bounds for Algorithm F.1.

Theorem F.1. *Algorithm F.1 has an expected regret of $\tilde{O}(\sqrt{\frac{C}{T}})$ in the objective and $(Z^*+L)\tilde{O}(\sqrt{\frac{C}{T}})$ in the constraints.*

Proof. For phase $r \geq 2$, using $n = 2^{r-2} = T_r$, the number of time steps in phase r, and $\rho = \frac{1}{T_r^2}$, from Lemma F.2 we obtain that with probability $1 - O(\frac{1}{T_r^2})$, Z available to phase r satisfies the property required by Lemma F.1 (with T substituted by T_r), which gives the following regret bounds for phase r: let $\boldsymbol{v}^\dagger_{\text{avg}}(r)$ be the average of played vectors in the T_r time steps of phase r. Let $\mathcal{F}_{r-1}$ denote the history till phase $r-1$. Then, with probability $1-O(\frac{1}{T_r^2})$ the history $\mathcal{F}_{r-1}$ is such that in phase r the regret in distance is bounded by $\mathbb{E}[d(\boldsymbol{v}_{\text{avg}}^\dagger(r),S)|\mathcal{F}_{r-1}] \leq \tilde{O}(\sqrt{\frac{C}{T_r}})$. With remaining probability $O(\frac{1}{T_r^2})$, the distance can be at most $T_r||\mathbf{1}_d||$. Let $\boldsymbol{v}_{\text{avg}}^\dagger$ denote the average of played vectors from the entire period of T time steps. Then, we get that total regret,

$$\begin{aligned}
\mathbb{E}[d(\boldsymbol{v}_{\text{avg}}^\dagger,S)] &\leq \frac{\|\mathbf{1}_d\|}{T} + \sum_{r=2}^{\log(T)+1} \frac{T_r}{T}\mathbb{E}[d(\boldsymbol{v}_{\text{avg}}^\dagger(r),S)] \\
&\leq \frac{\|\mathbf{1}_d\|}{T} + \sum_{r=2}^{\log(T)+1} \frac{T_r}{T}\tilde{O}(\sqrt{\frac{C}{T_r}} + \frac{T_r||\mathbf{1}_d||}{T_r^2}) \\
&= \tilde{O}(\sqrt{\frac{C}{T}}).
\end{aligned}$$

Similarly, we obtain bounds on regret in the objective,

$$\begin{aligned}
&\text{OPT} - \mathbb{E}[f(\boldsymbol{v}_{\text{avg}}^\dagger)] \\
&\leq \frac{1}{T} + \sum_{r=2}^{\log(T)+1} \frac{T_r}{T}(\text{OPT} - \mathbb{E}[f(\boldsymbol{v}_{\text{avg}}^\dagger(r))]) \\
&\leq \frac{1}{T} + \sum_{r=2}^{\log(T)+1} \frac{T_r}{T}(Z^*+L)\tilde{O}(\sqrt{\frac{C}{T_r}} + \frac{T_r||\mathbf{1}_d||}{T_r^2}) \\
&= (Z^*+L)\tilde{O}(\sqrt{\frac{C}{T}}).
\end{aligned}$$

Proof. [Proof of Lemma F.2] From Lemma 5.1, OPT^δ is concave in δ, therefore, for all $\delta > 3\gamma$

$$\begin{aligned}
\frac{(\text{OPT}^\delta - \text{OPT}^{2\gamma})}{\delta} &\leq \frac{(\text{OPT}^\delta - \text{OPT}^{2\gamma})}{\delta - 2\gamma} \\
&\leq \frac{(\text{OPT}^{3\gamma} - \text{OPT}^{2\gamma})}{\gamma}.
\end{aligned}$$

So, it suffices to prove that

$$\frac{(\text{OPT}^{3\gamma} - \text{OPT}^{2\gamma})}{\gamma} \leq Z \leq O(L+Z^*).$$

In Lemma F.4 and Lemma F.5, we prove that for every $\delta \geq \gamma$, with probability $1-O(\rho)$

$$\begin{aligned}
\hat{\text{OPT}}^\delta + L\gamma &\geq \text{OPT}^{\delta-\gamma}, \\
\text{OPT}^\delta + L\gamma &\geq \hat{\text{OPT}}^{\delta-\gamma}
\end{aligned} \tag{F.12}$$

Using above for $\delta = 4\gamma$, and $\delta = 2\gamma$, respectively, we get

$$\begin{aligned}
Z &:= \frac{(\hat{\text{OPT}}^{4\gamma}(n) - \hat{\text{OPT}}^{\gamma}(n))}{\gamma} + 2L \\
&\geq \frac{(\text{OPT}^{3\gamma} - \text{OPT}^{2\gamma})}{\gamma}.
\end{aligned}$$

In Lemma F.6, we prove that for any $\delta \geq \gamma$,

$$\hat{\text{OPT}}^\delta \leq \text{OPT} + O(\delta(Z^*+L)) \tag{F.13}$$

Using this along with $\hat{\text{OPT}}^\gamma \geq \text{OPT} - L\gamma$ from the first inequality in Equation (F.12), we get

$$\begin{aligned}
Z &= \frac{\hat{\text{OPT}}^{4\gamma} - \hat{\text{OPT}}^{\gamma}}{\gamma} \\
&\leq \frac{(\text{OPT} + 4\gamma O(Z^*+L)) - (\text{OPT} - L\gamma)}{\gamma} \\
&= O(Z^*+L).
\end{aligned}$$

This completes the proof.

Lemma F.3. *Given fixed $\{\boldsymbol{v}_t\}_{t=1}^T$, and a vector $\boldsymbol{\mu}$, for all $\rho > 0$ and $n \in [T]$, let $\gamma = ||\mathbf{1}_d||\sqrt{\frac{\log(d/\rho)}{n}}$. Then for a uniformly random permutation over $1,\dots,T$, with probability $1-O(\rho)$, the following holds for the first n time steps.*

$$\|\frac{1}{n}\sum_{t=1}^n \boldsymbol{v}_t - \frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t\| \leq \gamma,$$

$$\left|\frac{1}{n}\sum_{t=1}^n h_{A_t}(\boldsymbol{\mu}) - \frac{1}{T}\sum_{t=1}^T h_{A_t}(\boldsymbol{\mu})\right| \leq \gamma\|\boldsymbol{\mu}\|_*.$$

Proof. The first inequality is obtained by simple application of Chernoff-Hoeffding bounds (Lemma A.2) for every coordinate $v_{t,j}$, which gives

$$\left|\frac{1}{n}\sum_{t=1}^n v_{t,j} - \frac{1}{T}\sum_{t=1}^T v_{t,j}\right| \leq \sqrt{\frac{\log(d/\rho)}{n}},$$

with probability $1-O(\rho/d)$. Then taking union bound over the d coordinates, we get the required inequality.

The second inequality follows using Chernoff-Hoeffding bounds (Lemma A.2) for bounded random variables $Y_t = h_{A_t}(\boldsymbol{\mu})$, where $|Y_t| = |h_{A_t}(\boldsymbol{\mu})| \leq ||\boldsymbol{\mu}||_* \cdot ||\mathbf{1}_d||$ (from the definition of the dual norm). This gives with probability $1-O(\rho)$,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{t=1}^n h_{A_t}(\boldsymbol{\mu}) - \frac{1}{T}\sum_{t=1}^T h_{A_t}(\boldsymbol{\mu})\right| \\
&= |\frac{1}{n}\sum_{t=1}^n (Y_t - \mathbb{E}[Y_t])| \\
&\leq (||\boldsymbol{\mu}||_* \cdot ||\mathbf{1}_d||)\sqrt{\frac{\log(1/\rho)}{n}} \\
&\leq ||\boldsymbol{\mu}||_*\gamma.
\end{aligned}$$

Lemma F.4. *For all $\rho > 0$ and $n \in [T]$, let $\gamma = ||\mathbf{1}_d||\sqrt{\frac{\log(d/\rho)}{n}}$. For all $\delta \geq \gamma$, with probability $1-O(\rho)$,*

$$\hat{OPT}^\delta(n) \geq OPT^{\delta-\gamma} - L\gamma.$$

Proof. To prove $\hat{\mathrm{OPT}}^{\delta}(n) \geq \mathrm{OPT}^{\delta-\gamma} - L\gamma$, we prove that there exists a feasible primal solution of $\hat{\mathrm{OPT}}^{\delta}(n)$ that is at most γ distance from the optimal primal solution of $\mathrm{OPT}^{\delta-\gamma}$. Then, the lemma follows from the L-Lipschitz property of f.

Let $\{\boldsymbol{v}_t\}_{t=1}^T$ be the optimal primal solution for $\mathrm{OPT}^{\delta-\gamma}$, so that $d(\frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t, S) \leq \delta - \gamma$. Then,

$$\begin{aligned}
& d(\frac{1}{n}\sum_{t=1}^n \boldsymbol{v}_t, S) \\
\leq \ & ||\frac{1}{n}\sum_{t=1}^n \boldsymbol{v}_t - \frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t|| + d(\frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t, S) \\
\leq \ & \gamma + (\delta - \gamma) = \delta,
\end{aligned}$$

where we used the concentration bouds from Lemma F.3 to bound $||\frac{1}{n}\sum_{t=1}^n \boldsymbol{v}_t - \frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t||$ by γ. Therefore, $\{\boldsymbol{v}_t\}_{t=1}^n$ is a primal feasible solution of $\hat{\mathrm{OPT}}^{\delta}(n)$ with objective value $f(\frac{1}{n}\sum_{t=1}^n \boldsymbol{v}_t) \geq f(\frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t) - L||\frac{1}{n}\sum_{t=1}^n \boldsymbol{v}_t - \frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t|| \geq f(\frac{1}{T}\sum_{t=1}^T \boldsymbol{v}_t) - L\gamma = \mathrm{OPT}^{\delta-\gamma} - L\gamma$. Therefore, $\hat{\mathrm{OPT}}^{\delta} \geq f(\frac{1}{n}\sum_{t=1}^n \boldsymbol{v}_t) \geq \mathrm{OPT}^{\delta-\gamma} - L\gamma$.

Lemma F.5. *For all $\rho > 0$ and $n \in [T]$, let $\gamma = ||\mathbf{1}_d||\sqrt{\frac{\log(d/\rho)}{n}}$. For all $\delta \geq \gamma$, with probability $1 - O(\rho)$,*

$$OPT^{\delta} + L\gamma \geq \hat{OPT}^{\delta-\gamma}(n).$$

Proof. Define S^δ as the set $\{\boldsymbol{v} : d(\boldsymbol{v}, S) \leq \delta\}$. Then, using the derivation in Equation (D.5), we have that

$$\begin{aligned}
\mathrm{OPT}^{\delta} \ = \ & \min_{\lambda \geq 0, ||\boldsymbol{\phi}||_* \leq L, ||\boldsymbol{\theta}||_* \leq 1} \{\ f^*(\boldsymbol{\phi}) + \lambda h_{S^\delta}(\boldsymbol{\theta}) \\
& + \frac{1}{T}\sum_{t=1}^T h_{A_t}(-\boldsymbol{\phi} - \lambda\boldsymbol{\theta})\ \}.
\end{aligned}$$

Let λ^*, $\boldsymbol{\theta}^*$, $\boldsymbol{\phi}^*$ be the optimal dual solutions in above. Then,

$$\begin{aligned}
\hat{\mathrm{OPT}}^{\delta-\gamma}(n) \ = \ & \min_{\lambda \geq 0, ||\boldsymbol{\phi}||_* \leq L, ||\boldsymbol{\theta}||_* \leq 1} \{\ f^*(\boldsymbol{\phi}) + \lambda h_{S^{\delta-\gamma}}(\boldsymbol{\theta}) \\
& + \frac{1}{n}\sum_{t=1}^n h_{A_t}(-\boldsymbol{\phi} - \lambda\boldsymbol{\theta})\ \} \\
\leq \ & f^*(\boldsymbol{\phi}^*) + \lambda^* h_{S^{\delta-\gamma}}(\boldsymbol{\theta}^*) \\
& + \frac{1}{n}\sum_{t=1}^n h_{A_t}(-\boldsymbol{\phi}^* - \lambda^*\boldsymbol{\theta}^*)
\end{aligned}$$

Now, using concentration bounds from Lemma F.3 for the sum of h_{A_t}'s, we obtain,

$$\begin{aligned}
\hat{\mathrm{OPT}}^{\delta-\gamma}(n) \ \leq \ & f^*(\boldsymbol{\phi}^*) + \lambda^* h_{S^{\delta-\gamma}}(\boldsymbol{\theta}^*) \\
& + \frac{1}{T}\sum_{t=1}^T h_{A_t}(-\boldsymbol{\phi}^* - \lambda^*\boldsymbol{\theta}^*) \\
& + \gamma(\lambda^*||\boldsymbol{\theta}^*||_* + ||\boldsymbol{\phi}^*||_*).
\end{aligned}$$

Now, observe that for any $\boldsymbol{\theta}$, $h_{S^\delta}(\boldsymbol{\theta}) \geq h_{S^{\delta-\gamma}}(\boldsymbol{\theta}) + \gamma||\boldsymbol{\theta}||_*$. To see this, let $\boldsymbol{v}$ be the maximizer in the definition of $h_{S^{\delta-\gamma}}$, i.e., $\boldsymbol{v} = \arg\max_{\mathbf{u} \in S^{\delta-\gamma}} \mathbf{u} \cdot \boldsymbol{\theta}$. Then consider $\boldsymbol{v}' = \boldsymbol{v} + \gamma\frac{\boldsymbol{\theta}}{||\boldsymbol{\theta}||}$. We have that $||\boldsymbol{v}' - \boldsymbol{v}|| = \gamma$, so that $\boldsymbol{v} \in S^{\delta-\gamma}$ implies that $\boldsymbol{v} \in S^\delta$. Therefore $h_{S^\delta}(\boldsymbol{\theta}) \geq \boldsymbol{v}' \cdot \boldsymbol{\theta} = \boldsymbol{v} \cdot \boldsymbol{\theta} + \gamma||\boldsymbol{\theta}||_* = h_{S^{\delta-\gamma}}(\boldsymbol{\theta}) + \gamma||\boldsymbol{\theta}||_*$. Substituting, we get,

$$\begin{aligned}
\hat{\mathrm{OPT}}^{\delta-\gamma}(n) \ \leq \ & f^*(\boldsymbol{\phi}^*) + \lambda^* h_{S^\delta}(\boldsymbol{\theta}^*) - \gamma\lambda^*||\boldsymbol{\theta}^*||_* \\
& + \frac{1}{T}\sum_{t=1}^T h_{A_t}(-\boldsymbol{\phi}^* - \lambda^*\boldsymbol{\theta}^*) \\
& + \gamma(\lambda^*||\boldsymbol{\theta}^*||_* + ||\boldsymbol{\phi}^*||_*) \\
= \ & \mathrm{OPT}^{\delta} + \gamma||\boldsymbol{\phi}^*||_* \\
\leq \ & \mathrm{OPT}^{\delta} + \gamma L
\end{aligned}$$

Lemma F.6. *For all $\delta \geq \gamma$, with probability $1 - O(\rho)$,*

$$\hat{OPT}^{\delta}(n) \leq OPT + 2\delta(L + Z^*)$$

Proof. Using the derivations in Equation (D.5),

$$\begin{aligned}
\hat{\mathrm{OPT}}^{\delta}(n) \ = \ & \min_{\lambda \geq 0, ||\boldsymbol{\phi}||_* \leq L, ||\boldsymbol{\theta}||_* \leq 1} \{\ f^*(\boldsymbol{\phi}) + \lambda h_S(\boldsymbol{\theta}) \\
& + \frac{1}{n}\sum_{t=1}^n h_{A_t}(-\boldsymbol{\phi} - \lambda\boldsymbol{\theta}) + \delta\lambda\ \},
\end{aligned}$$

Let $\lambda^*, \boldsymbol{\phi}^*, \boldsymbol{\theta}^*$ denote the optimal dual solution for OPT, then,

$$\begin{aligned}
\hat{\mathrm{OPT}}^{\delta}(n) \ \leq \ & f^*(\boldsymbol{\phi}^*) + \lambda h_S(\boldsymbol{\theta}^*) \\
& + \frac{1}{n}\sum_{t=1}^n h_{A_t}(-\boldsymbol{\phi}^* - \lambda^*\boldsymbol{\theta}^*) + \delta\lambda^*
\end{aligned}$$

Now, using concentration bounds from Lemma F.3 for the sum of h_{A_t}'s, we obtain,

$$\begin{aligned}
\hat{\mathrm{OPT}}^{\delta}(n) \ \leq \ & f^*(\boldsymbol{\phi}^*) + \lambda h_S(\boldsymbol{\theta}^*) \\
& + \frac{1}{T}\sum_{t=1}^T h_{A_t}(-\boldsymbol{\phi}^* - \lambda^*\boldsymbol{\theta}^*) \\
& + \gamma(\lambda^*||\boldsymbol{\theta}^*||_* + ||\boldsymbol{\phi}^*||_*) + \delta\lambda^* \\
= \ & \mathrm{OPT} + \gamma(\lambda^*||\boldsymbol{\theta}^*||_* + ||\boldsymbol{\phi}^*||_*) + \delta\lambda^* \\
\leq \ & \mathrm{OPT} + (L + \lambda^*)\gamma + \delta\lambda^* \\
\leq \ & \mathrm{OPT} + 2(L + \lambda^*)\delta \\
= \ & \mathrm{OPT} + 2\delta(L + Z^*)
\end{aligned}$$

G Proof of Lemma 6.3

Given an instance of the online packing problem, recall that $(r_t^*, \boldsymbol{v}_t^*)$ denotes the optimal offline solution. Then $\mathrm{OPT}_{\mathrm{SUM}} = \sum_{t=1}^T r_t^*$, and $\sum_{t=1}^T \boldsymbol{v}_t^* \leq B\mathbf{1}$. Given $\rho > 0$, let $\eta = \sqrt{3\log(\frac{d+2}{\rho})}$. Let the given random subset of δ fraction of requests be Γ. *Define $\hat{\mathrm{OPT}}$ to be $1/\delta$ times the optimum value of the following scaled optimization problem*: pick $(r_t^\dagger, \boldsymbol{v}_t^\dagger)$ for each $t \in \Gamma$, to maximize the total reward $\sum_{t\in\Gamma} r_t^\dagger$ such that $\sum_{t\in\Gamma} \boldsymbol{v}_t^\dagger \leq (\delta B + \eta\sqrt{\delta B})\mathbf{1}$.

The bounds we need on $\hat{\mathrm{OPT}}$ follow from considering the optimal primal and dual solutions to the given packing problem restricted to the sample and using Corollary A.1 to bound their values on the sample. Applying Corollary A.1 to the set of r_t^* for all $t \in [T]$ we get that with probability at least $1 - \rho/(d+2)$,

$$\begin{aligned}\sum_{t\in\Gamma} r_t^* &\geq \delta\mathrm{OPT}_{\mathrm{SUM}} - \sqrt{3\delta\mathrm{OPT}_{\mathrm{SUM}} \log(\tfrac{d+2}{\rho})} \\ &= \delta\mathrm{OPT}_{\mathrm{SUM}} - \eta\sqrt{\delta\mathrm{OPT}_{\mathrm{SUM}}}.\end{aligned}$$

Similarly, applying Corollary A.1 to each co-ordinate of the set of $\boldsymbol{v}_t^*$s, and taking a union bound, we get that with probability at least $1 - \rho d/(d+2)$,

$$\begin{aligned}\sum_{t\in\Gamma} \boldsymbol{v}_t^* &\leq (\delta B + \sqrt{3\delta B \log(\tfrac{d+2}{\rho})})\mathbf{1} \\ &= (\delta B + \eta\sqrt{\delta B})\mathbf{1}.\end{aligned}$$

Therefore with probability $1 - \rho(d+1)/(d+2)$ both the inequalities above hold and $(r_t^*, \boldsymbol{v}_t^*)_{t\in\Gamma}$ is a feasible solution to the scaled optimization problem used to define $\hat{\mathrm{OPT}}$. Hence

$$\delta\hat{\mathrm{OPT}} \geq \sum_{t\in\Gamma} r_t^* \geq \delta\mathrm{OPT}_{\mathrm{SUM}} - \eta\sqrt{\delta\mathrm{OPT}_{\mathrm{SUM}}}$$

and the first bound on $\hat{\mathrm{OPT}}$ follows from dividing the above inequality throughout by δ. For the second bound, we need to consider the dual of the packing problem. The packing problem has the following natural LP relaxation. (The dual LP follows.)

$$\begin{aligned}&\max \sum_{t=1}^{T} \sum_{\mathbf{v}\in A_t} r(\mathbf{v}) x_{t,\mathbf{v}} \\ &\text{s.t.} \forall\ t, \sum_{\mathbf{v}\in A_t} x_{t,\mathbf{v}} \leq 1 \\ &\sum_{t=1}^{T} \sum_{\mathbf{v}\in A_t} \mathbf{v} x_{t,\mathbf{v}} \leq B\mathbf{1}.\end{aligned}$$

$$\begin{aligned}&\min \sum_{t=1}^{T} \beta_t + B\boldsymbol{\theta}\cdot\mathbf{1} \\ &\text{s.t.} \forall\ t, \forall\ \mathbf{v} \in A_t, \beta_t \geq r(\mathbf{v}) - \mathbf{v}\cdot\boldsymbol{\theta}, \\ &\quad \forall\ t, \beta_t \geq 0, \boldsymbol{\theta} \geq 0.\end{aligned}$$

First of all, we ignore the integrality gap and assume that the value of the optimal dual (and primal) solution is equal to the optimal value $\mathrm{OPT}_{\mathrm{SUM}}$ for the offline packing problem. Let $(\beta_t^*)_{t=1}^T, (\theta_j^*)_{j=1}^d$ be the optimal dual solution for the given instance, and $\mathrm{OPT}_{\mathrm{SUM}} = \sum_t \beta_t^* + \sum_j B\theta_j^*$. It can be shown that $\beta_t^* \in [0,1]$ for all t: all the constraints involving β_t are of the form $\beta_t \geq (\cdot)$ so at least one of these constraints is tight for the optimal solution. Also for each of these constraints, the RHS is at most 1, and one of the constraints is $\beta_t \geq 0$. Further note that these constraints are local, i.e., they only depend on the request indexed by t. This means that $(\beta_t^*)_{t\in\Gamma}, (\theta_j^*)_{j=1}^d$ is a feasible solution to the dual of the scaled optimization problem. The objective value of this solution to this dual is

$$\sum_{t\in\Gamma} \beta_t^* + \sum_j (\delta B + \eta\sqrt{\delta B})\theta_j^* \geq \delta\hat{\mathrm{OPT}}.$$

Using Corollary A.1 on the set of β_t^*s, we get that with probability at least $1 - \rho/(d+2)$,

$$\begin{aligned}\sum_{t\in\Gamma} \beta_t^* &\leq \delta\sum_{t=1}^{T} \beta_t^* + \sqrt{3\delta\mathrm{OPT}_{\mathrm{SUM}} \log(\tfrac{d+2}{\rho})} \\ &= \delta\sum_{t=1}^{T} \beta_t^* + \eta\sqrt{\delta\mathrm{OPT}_{\mathrm{SUM}}}.\end{aligned}$$

Putting the two inequalities above together,

$$\begin{aligned}\frac{\delta\hat{\mathrm{OPT}}}{1+\eta/\sqrt{\delta B}} &\leq \sum_{t\in\Gamma} \beta_t^* + \delta\sum_j B\theta_j^* \\ &\leq \delta\left(\sum_{t=1}^{T} \beta_t^* + \sum_j B\theta_j^*\right) + \eta\sqrt{\delta\mathrm{OPT}_{\mathrm{SUM}}} \\ &= \delta\mathrm{OPT}_{\mathrm{SUM}} + \eta\sqrt{\delta\mathrm{OPT}_{\mathrm{SUM}}}.\end{aligned}$$

The lemma follows by taking the union bound over the probabilities for the two inequalities as required. Finally, we ignored the integrality gap, but it is easy to show that this gap is at most $1 - \frac{1}{B}$, which can be absorbed in the $1 + \eta/\sqrt{\delta B}$ factor.