How the Experts Algorithm Can Help Solve LPs Online

Anupam Gupta∗ Marco Molinaro†

July 22, 2014

Abstract

We consider the problem of solving packing/covering LPs online, when the columns of the constraint matrix are presented in random order. This problem has received much attention: the main open question is to figure out how large the right-hand sides of the LPs have to be (compared to the entries on the left-hand side of the constraint) to get $(1 + \varepsilon)$-approximations online? It is known that the RHS has to be $\Omega(\varepsilon^{-2} \log m)$ times the left-hand sides, where $m$ is the number of constraints.

In this paper we give a primal-dual algorithm to achieve this bound for all packing LPs, and also for a class of mixed packing/covering LPs. Our algorithms construct dual solutions using a regret-minimizing online learning algorithm in a black-box fashion, and use them to construct primal solutions. The adversarial guarantee that holds for the constructed duals help us to take care of most of the correlations that arise in the algorithm; the remaining correlations are handled via martingale concentration and maximal inequalities. These ideas lead to conceptually simple and modular algorithms, which we hope will be useful in other contexts.

1 Introduction

In this paper we consider the problem of solving packing-covering LPs online. For concreteness, consider as a simple example the packing LP \( \max \{ \pi^\top x \mid Ax \leq b, x \in [0, 1]^n \} \)—we will consider more complicated LPs in this paper—where the columns of $A$ are being revealed one by one, and we have to choose values for each $x_t$ and irrevocably before the next columns are revealed.

While such problems have been also studied in the worst-case competitive analysis [BN09, ABFP13], to avoid the pessimistic achievable guarantees a lot of recent research has focused on the random permutation model: the matrix $A$ is chosen adversarially but its columns are presented to the algorithm in random order. Several packing problems have been considered in this model, starting from a classic maximization version of the secretary problem [Dyn63], to single-knapsack problems [Kle03, BIKK07], the AdWords problem [MSVV07, GM08, DH09] and recently more general packing LPs [FHK+10, AWY09, MR12, KRTV13]. These models have a vast range of applications, like online advertisement, online routing, and airline revenue management. In this paper we consider this random permutation model.

A major question of research has been: how large must the right-hand sides of the LPs be (compared to the left-hand side coefficients) to get $(1 + \varepsilon)$-approximations online, say, in expectation?

Dual methods. Dual-based method are very important in optimization under uncertainty and have been used extensively in the operations management literature [Sim89, WI92, TR98]. Most of the works on packing LPs in the random permutation model [DH09, FHK+10, AWY09, MR12] use the

---

∗Computer Science Department, Carnegie Mellon University, Pittsburgh, PA. Research partly supported by NSF awards CCF-1016799 and CCF-1319811, and by a grant from the CMU-Microsoft Center for Computational Thinking.

†ISyE, Georgia Institute of Technology.
following form of this approach. If the optimal dual solution was known, then one could use it to set the primal variables according to their reduced costs and obtain an (almost) optimal solution. The idea is then to see the first few (say $\varepsilon n$) columns of the LP, use this sample to estimate a good dual solution for the LP, and then use this dual solution to decide how to select the primal values for the future columns.

This idea was first analyzed by Devanur and Hayes for the AdWords problem [DH09], and recently extended to more general packing LPs [FHK+10 AWY09]; the latter show that having right-hand sides roughly $\Omega(\varepsilon^{-2}m \log(n/\varepsilon))$ times larger than the left-hand side entries suffices, where $m$ is the number of packing constraints. Recently, Molinaro and Ravi [MR12] also used a modified version of this idea to show that $\Omega(\varepsilon^{-2}m^2 \log(m/\varepsilon))$ suffices, removing the dependence on $n$ at the expense of extra $m$ factors. Agrawal et al. [AWY09] showed that $\Omega(\varepsilon^{-2}\log m)$ is the best one can do, and this was believed to be the right answer.

The main difficulty in analyzing these algorithms lies in the fact that the primal and dual solutions are, by construction, heavily connected. In order to deal with the correlations that arise, all of these above analyses resort to a massive union bound, which is the root of the extra $\log n$ and $m$ factors.\footnote{Recently and independently, Kesselheim et al. [KRV13] circumvents these issues by giving a primal-only algorithm with optimal parameters; they leave the question of getting a primal-dual question open. We discuss their work at the end of Section 1.2}

1.1 Our Techniques

Decoupling primal and dual via regret minimization. One of our contributions in this paper is an alternative algorithmic construction where the primal and the dual are loosely coupled. For that, we construct duals using a regret-minimizing online learning algorithm (which offers adversarial guarantees); in the primal, we use a greedy strategy which ensures that, with respect to these duals, we are doing at least as well as the optimal offline solution.

The fact that the guarantees for the dual hold in the adversarial setting, together with the greedy primal step, reduces the analysis to that of comparing the dual constructed solution to the offline optimal solution (see Section 2.4). Since the latter is a fixed solution (modulo the permutation of the columns), it is only loosely correlated with our constructed dual; in fact, correlations only come from the fact we can think of the random permutation model as sampling without replacement.

As far as we know, this is the first explicit black-box connection between regret-minimization and optimization in the random permutation model; in the offline packing-covering LP setting such a connection was implicit in, e.g., [PST95 Yon04 GK07] and made explicit in, e.g., [Kha04 Haz06 AHK12 CHW12]. In the simpler i.i.d. case, where columns of the LP are sampled with replacement, [DJSW11] uses multiplicative-weights based dual updates, reminiscent of the multiplicative-weight expert algorithm for regret-minimization. Our analysis abstracts out the reliance on the details of that technique, giving a clean conceptual explanation of what is happening. Importantly, this isolation between the analysis of the regret-minimization algorithm and the comparison of our dual solution and the offline optimum is instrumental in dealing with the dependencies arising in the random permutation model.

Handling dependencies in random permutations. Handling the dependencies that arise from sampling without replacement is not immediate: the known bounds on the distances between sampling with and without replacement distributions (e.g., [DFS00]) do not seem to be strong enough, and a simple approach requires taking a union bound across all time steps, leading again to extra $\log n$ factors. Instead, we use a maximal inequality for sampling without replacement to compare sampling with/without replacement at every time step while avoiding a union bound, together with martingale
1.2 Our Results

Generalized Load Balancing. First we consider a simple packing-type problem, modeling the following generalization of the load balancing problem: We have a set of $m$ machines. For each arriving job $t$, there are $k$ different options on how to serve it, and the $j^{th}$ choice requires some amount of processing on each of the machines, given by the vector $(a_{t,j}^1, a_{t,j}^2, \ldots, a_{t,j}^m)$. When the job arrives, we must (fractionally) choose one of these options. The goal is to minimize the makespan, i.e. the maximum total processing assigned to any machine. Notice that when the matrices $A^t$’s are diagonal $m \times m$ this captures the classic problem of scheduling in unrelated machines.

If the jobs are i.i.d. draws from some distribution, the result of Devanur et al. [DJSW11, Section 3] mentioned above achieve with probability $1 - \delta$, a $(1 + \varepsilon)$-approximation to the optimum makespan assuming this optimum is $\Omega(\frac{\log(m/\delta)}{\varepsilon^2})$ times the maximum processing $a_{t,j}^i$. They left as an open question whether the same result holds in the more general random permutation model. Our first result, algorithm expertLB, answers this in the affirmative:

Informal Theorem 1.1 (Load Balancing). Consider the load-balancing problem where the jobs arrive in random order. Given $\varepsilon > 0$, if the optimal makespan $\lambda^*$ is at least $\Omega(\frac{\log(m/\delta)}{\varepsilon^2})$ times the maximum processing $a_{t,j}^i$. They left as an open question whether the same result holds in the more general random permutation model. Our first result, algorithm expertLB, answers this in the affirmative:

Informal Theorem 1.1 (Load Balancing). Consider the load-balancing problem where the jobs arrive in random order. Given $\varepsilon > 0$, if the optimal makespan $\lambda^*$ is at least $\Omega(\frac{\log(m/\delta)}{\varepsilon^2})$ times the maximum processing $a_{t,j}^i$. They left as an open question whether the same result holds in the more general random permutation model. Our first result, algorithm expertLB, answers this in the affirmative:

As is common in this context, our algorithm produces integer solutions that compare favorably to the optimal fractional solution.

Moreover, we can handle processing requirements that can be both positive or negative, as long as for each machine, the jobs are mostly positive or mostly negative (see Definition 2.1). This is useful, because we employ this abstraction for the next result, to solve packing-covering linear programs.

Packing-Covering LPs. Again we have $n$ jobs with $k$ processing options. Choice $j$ for a job $t$ has a profit $\pi^t_j$. The goal is to make fractional choices $\{x^t_j\}$ (subject to $\sum_j x^t_j \leq 1$) that maximize the profit $\sum_{t,j} \pi^t_j x^t_j$ subject to $m$ constraints, some of them of packing form $\sum_{t,j} a^t_{ij} x^t_j \leq b_i$ some of covering form $\sum_{t,j} c^t_{ij} x^t_j \geq d_i (a^t_{ij}$ and $c^t_{ij}$ non-negative). This is a (multiple-choice) packing-covering LP where the items arrive online, in random order; the multiple-choice emphasizes the presence of constraints like $\sum_j x^t_j \leq 1$ which have small RHS. (In the case $k = 1$, the multiple-choice model degenerates to the standard online packing-covering LP model where a single variable $x^t$ is revealed at each time, and must be assigned a fractional value in $[0,1]$.) A restricted version of this multiple-choice model with only packing constraints has been analyzed recently [EHK+10, AVY09, KRTV13]; even this restricted version already generalizes the well-studied AdWords problem [MSV07, GM08, DH09].

We show how to simply reduce the problem of solving such a packing-covering LP problem to a load balancing problem by viewing the covering constraints and the objective function as yet another packing constraints with (mostly) negative loads. This gives the next result.

Informal Theorem 1.2 (Packing-Covering LPs). Consider a feasible multiple-choice packing-covering LP where the items arrive in random order. Given $\varepsilon, \delta$, assume that each right hand side in the LP is at least $\Omega(\frac{\log(m/\delta)}{\varepsilon^2})$ times the maximum coefficient of the left-hand side in its row, and that the optimal profit is at least $\Omega(\frac{\log(m/\delta)}{\varepsilon^2})$ times each individual item’s profit. Then given an (under)estimate $O^*$ for the optimal profit, the algorithm LPviaLB computes an $\varepsilon$-feasible solution online that achieves a profit of $(1 - \varepsilon)O^*$ with probability $1 - \delta$. 


The astute reader will notice that the above two theorems required us to provide estimates of the optimal profit as input. This is sometimes a reasonable assumption (e.g. when enough historic knowledge of the problem directly provides such estimates). Nonetheless, we show how to construct such estimates as the columns come. However, to obtain good estimates we need to consider stable LPs, which informally means allowing solutions to violate the covering constraints by a small factor does not increase the optimal value by much (see Definition 4.1). We note that a packing-only LP is vacuously stable.

The rough idea to obtain the optimum estimate is to sample half the columns of the LP and solve it (with the capacities reduced by a factor of 2) to obtain a good estimate for the optimal value. This would, however, lose the profit from 50% of the LP right off the bat! We then use a doubling idea of [AWY09] to learn progressively better opt estimates (this was also used in [Kle05, DJSW11, MRT2]).

Informal Theorem 1.3 (Packing-Covering LPs II). Consider a feasible, stable, multiple-choice packing-covering LP in the random order setting. Given \( \varepsilon, \delta \), assume that each right hand side in the LP is at least \( \Omega\left(\frac{\log(m/\delta)}{\varepsilon^2}\right) \) times the maximum coefficient of the left-hand side in its row, and that the optimal profit is at least \( \Omega\left(\frac{\log(m/\delta)}{\varepsilon^2}\right) \) times each individual item’s profit. Then the \textbf{MultiphaseLP} algorithm computes a solution that is \( \varepsilon \)-feasible with probability \( 1 - \delta \) and has expected value at least \( (1 - \varepsilon) \) times the optimum.

As far as we know this is the first such guarantee for packing-covering LPs in the random permutation (or i.i.d.) model.

**Packing LPs.** The assumption of individual item profits being small compared to the optimal net profit is often a reasonable one. However, there are reasonable situations where we want a stronger result. Our final result removes all assumptions about the magnitude of the profit whenever we are dealing with packing-only LPs (without multiple-choice constraints of the form \( \sum_j x^t_j \leq 1 \)). Obtaining good estimates of \( \text{opt} \) without any assumptions on the item values in our setting requires significantly new ideas.

Informal Theorem 1.4 (Ultimate Packing LP). Consider a packing LP in the random order setting. Given \( \varepsilon > 0 \), suppose the capacities are \( \Omega\left(\frac{\log(m/\varepsilon)}{\varepsilon^2}\right) \) times any left hand side entry. Then the \textbf{MultiSkimLP} algorithm computes a solution online with expected value at least \((1 - \varepsilon)\) times the optimum.

As mentioned earlier, this problem has been approached using dual-based methods by several authors [DH09, AWY09, MRT2], showing that bounds \( \Omega(\varepsilon^{-2} m \log n) \) and \( \Omega(\varepsilon^{-2} m^2 \log m) \) are sufficient and \( \Omega(\varepsilon^{-2} \log m) \) is necessary. Hence our result is asymptotically optimal.

Very recently, and independently of our work, Kesselheim et al. [KRTV13] give a primal-only algorithm for the more general multiple-choice packing LPs that achieves a \((1+\varepsilon)\)-approximation when the bound on the capacities is \( \Omega(\varepsilon^{-2} \log d) \), where \( d \) is the column-sparsity of the constraint matrix (i.e. maximum number of non-zero entries in each column). Since \( d \leq m \), this bound is optimal. Their algorithm is elegant, where at time \( t \) the optimal solution of a primal LP comprising the columns seen up to time \( t \) is directly used to set the variable \( x_t \); somewhat surprisingly, this strategy, which does not directly impose consistency of the \( x_t \)’s set in different time steps, still works. While the guarantee for packing LPs we obtain does not fully recover the result of Kesselheim et al., we believe that it is still worth presenting, specially given the potential of this method to extend to more general cases, e.g. to packing-covering LPs, and also as a different primal-dual approach to contrast with their primal-only ideas.
2 Load-Balancing using Experts

In this section, we formally define the generalized load-balancing problem and present our online algorithm for it.

2.1 Definitions: Offline and Online Instances

An instance of the offline version of the Load-balance problem is a set of matrices \( \{A^t\} \), each in \( \mathbb{R}^{n \times k} \). The goal is to find vectors \( p^1, p^2, \ldots, p^n \in \Delta^k \) to minimize \( \|\sum_{t=1}^n A^tp^t\|_{\max} \), the load of the most-loaded machine, where \( \|v\|_{\max} = \max_j v_j \) and \( \Delta^k \) is the simplex \( \{p \in [0,1]^k : \sum_i p_i = 1\} \). We use throughout \( \lambda^* \) to denote the offline optimum value.

Notice that, by adding a column of zeroes, we can get an instance where one option is to do nothing (i.e., where \( \sum_j p^t_j \leq 1 \), denoted by \( p^t \in \Delta^k \)).

In the online version of this problem, we consider the random permutation model. Now the number of time steps \( n \) is the only information known upfront. Let \( A^1, A^2, \ldots, A^n \) be matrices sampled from the set \( \{A^1, \ldots, A^n\} \) uniformly without replacement. At time step \( t \), the algorithm has seen matrices \( A^1, \ldots, A^t \) and must decide on a (random) vector \( p^t \in \Delta \); the randomness is both from the random order and the internal coin flips of the algorithm (if any). The goal is to minimize \( \|\sum_{t=1}^n A^tp^t\|_{\max} \).

The vectors \( \{p^t\}_t \) output by the algorithm are called the online solution for the online instance \( \{A^t\}_t \) corresponding to the offline instance \( I \).

Ideally we want guarantees that for any given \( \epsilon, \delta > 0 \), we get \( \|\sum_{t=1}^n A^tp^t\|_{\max} \leq (1+\epsilon)\lambda^* \) with probability \( (1-\delta) \). Due to the online nature of the problem, such a guarantee is only conceivable if \( \lambda^* \) is “large” compared to the maximum size of the loads. In the rest of this section we show how to get the best possible such guarantee.

2.2 Well-Bounded Instances

While instances arising from scheduling-type applications usually consist solely of positive loads, for future sections it is useful to handle instances where some entries \( A_{ij}^t \) are also negative. To get good guarantees we need to control the class of permissible instances: loosely speaking, while we allow the entries of the load matrices \( A^t \) to be in the symmetric interval \([{-M,M}]\), on each machine the loads are either mostly positive on all steps (with any negative loads being very small), or mostly a negative load on all steps.

**Definition 2.1.** For \( M, \gamma \geq 0 \), an instance \( A^1, \ldots, A^n \) of the load-balance problem is \((M, \gamma)\)-well-bounded if \( A_{ij}^t \in [{-M,M}] \) for all \( i,j,t \), and moreover for each \( i \in [m] \) we have: either \( A_{ij}^t \in [{-\min\{\frac{2\lambda^*}{n},M\},M}] \) for all \( t \in [n] \), or \( A_{ij}^t \in [{-M,\min\{\frac{2\lambda^*}{n},M\}}] \) for all \( t \in [n] \).

In particular, this is satisfied with \( \gamma = 0 \) if the \( A^t \)'s are non-negative. The reader can think of \( \gamma \) as a small constant, say one. The main motivation behind this definition is that it allows us to control the variation of random processes defined over \( \{A^t\}_t \), as the next lemma shows.

**Lemma 2.2.** Suppose \( \{A^t\}_{t=1}^n \) is an \((M, \gamma)\)-well-bounded instance for some \( M, \gamma \geq 0 \), and consider \( p^1, \ldots, p^n \in \Delta^k \). Let the sequence \( A^1\tilde{p}^1, \ldots, A^n\tilde{p}^n \) be sampled without replacement from the set \( \{A^tp^t\}_t \). Then for every event \( \mathcal{E} \) and for every \( i \in [m] \) and \( t \in [n] \),

\[
\mathbb{E} \left[ \left| A_i^t\tilde{p}^t - \mathbb{E}[A_i^t\tilde{p}^t | \mathcal{E}] \right| \right] \leq \frac{2\gamma\lambda^*}{n} + 2\mathbb{E} \left[ |A_i^t\tilde{p}^t| | \mathcal{E} \right].
\]

**Proof.** Fix \( i \in [m] \). Define \( \tilde{o}^t_i \) as follows: if \( A_i^t\tilde{p}^t \geq -\frac{2\lambda^*}{n} \) for all \( t \), set \( \tilde{o}^t_i = A_i^t\tilde{p}^t + \frac{2\lambda^*}{n} \) for all \( t \); else set \( \tilde{o}^t_i = \frac{2\lambda^*}{n} - A_i^t\tilde{p}^t \). By definition of well-boundedness, for all \( t \) we have \( \tilde{o}^t_i \geq 0 \), and since we did a
uniform shift
\[
E \left[ |A_i^t \hat{p}^t - E[A_i^t \hat{p}^t] | \epsilon \right] = E \left[ |\hat{o}_i^t - E[\hat{o}_i^t] | \epsilon \right] \leq E \left[ |\hat{o}_i^t| | \epsilon \right] + E \left[ |\hat{o}_i^t| | \epsilon \right] = 2E \left[ |\hat{o}_i^t| | \epsilon \right].
\]
But it is easy to see that \( E[\hat{o}_i^t | \epsilon] \leq \frac{\lambda^*}{n} + E[A_i^t \hat{p}^t | \epsilon] \), which concludes the proof. \qed

2.3 The expertLB Algorithm and Its Guarantee

To define the algorithm expertLB, we need to recall the online experts problem [AHK12]: We are given an adversarial sequence of vectors \( o^1, o^2, \ldots, o^n \in [-1, 1]^m \). At time \( t \), up to time \( t - 1 \) (i.e. \( o^1, \ldots, o^{t-1} \)), we need to compute a vector \( \hat{w}^t \in \Delta^m \); then we incur a reward of \( \langle w^t, o^t \rangle \). The goal is to maximize the total reward \( \sum_t \langle w^t, o^t \rangle \).

Given an online instance \( \{A^t\} \) to the generalized load-balancing problem and values \( n, M, \) and \( \varepsilon \), the following algorithm expertLB (for “expert load-balancing”) runs a primal greedy strategy and a dual online experts algorithm, restarting at timestep \( n/2 \). The algorithm maintains primal vectors \( p^1, \ldots, p^{n/2} \in \Delta^k \) and “dual” vectors \( \hat{w}^1, \ldots, \hat{w}^{n/2} \in \Delta^m \) as follows:

(P) primal step: in step \( t \), the algorithm sees the random item \( A^t \), computes \( \hat{w}^t A^t \) and chooses \( \hat{p}^t \in \Delta^k \) so as to minimize \( \langle \hat{w}^t A^t, \hat{p}^t \rangle \); 

(D) dual step: run an online experts algorithm (with \( A^t p^1, \ldots, A^t p^t \) as the adversarial vectors) to compute \( \hat{w}^{t+1} \) (so the “reward” accrued by this experts algorithm at time \( t+1 \) is \( \langle \hat{w}^{t+1}, A^{t+1} p^{t+1} \rangle \)).

At time \( n/2 \), the algorithm restarts the online experts subroutine afresh, and computes the vectors \( p^{n/2+1}, \ldots, p^n \) and \( \hat{w}^{n/2+1}, \ldots, \hat{w}^n \) in the same way as before. For concreteness, the online experts algorithm we use is a multiplicative-weights update algorithm from [AHK12] Section 2.3, scaling down by \( M \) to ensure gains are in \([-1, 1]\), and using \( \varepsilon \) for the learning rate \( \eta \).

Note the simplicity of the algorithm: it is perhaps the “natural” algorithm, once we decide to reduce load-balancing to the experts algorithm. The main theorem of this section states the guarantee of this algorithm; given a vector \( v \in \mathbb{R}^m \), define \( |v| := (|v_1|, \ldots, |v_m|) \).

**Theorem 2.3** (Load Balancing Guarantee). Suppose \( \{A^t\} \) is \((M, \gamma)\)-well-bounded load-balancing instance for \( \gamma \geq 1 \). Let \( \lambda^* \) be its optimal load and suppose \( \varepsilon \leq \varepsilon_1 := \frac{1}{8M} \) and \( \delta \in (0, \varepsilon] \) are such that \( \lambda^* \geq \frac{3M \log(m/\delta)}{2\varepsilon^2} \). Given values of \( n, M \) and \( \varepsilon \), the algorithm expertLB finds an online solution \( \{p^t\}_t \) such that with probability at least \( 1 - \delta \), \( \|\sum_t A^t p^t - \varepsilon \sum_t |A^t p^t|\|_{\max} \leq \lambda^*(1 + c_1(1 + \gamma)\varepsilon) \) for a universal constant \( c_1 \).

When the load matrices are non-negative, well-boundedness means \( A^t_{ij} \in [0, M] \) and we can set \( \gamma = 1 \) to get:

**Corollary 2.4** (Positive Loads Guarantee). Suppose \( \{A^t\} \) is a load-balancing instance with \( A^t_{ij} \in [0, M] \), and optimal load \( \lambda^* \). Suppose \( \varepsilon \leq \varepsilon_1 := \frac{1}{8M} \) and \( \delta \in (0, \varepsilon] \) satisfy \( \lambda^* \geq \frac{3M \log(m/\delta)}{2\varepsilon^2} \). Given \( n, M \) and \( \varepsilon \), the algorithm expertLB finds an online solution \( \{p^t\}_t \) such that \( \|\sum_t A^t p^t\|_\infty \leq (1 + O(\varepsilon))\lambda^* \) with probability at least \( 1 - \delta \).

In the rest of Section 2, we prove Theorem 2.3. We first show in Section 2.4 that the maximum load is close in expectation to \( \lambda^* \). We then extend these ideas to prove the high-probability guarantee in Section 2.5.

2.4 The Guarantee in Expectation

Let \( \bar{p}^1, \ldots, \bar{p}^n \) be the optimal solution for the offline instance. We see this as a mapping from matrix \( A^t \) to solution \( \bar{p}^t \) (say \( \phi(A^t) = \bar{p}^t \)); this way, define \( \bar{p}^t \) as the random solution with respect to the
random matrix $A^t$ (i.e. $p^t = \phi(A^t)$). (Formally, if there are repeated matrices then we can assume that the optimal solution does the same thing for all of them.)

To simplify the notation, let $o^t := A^tp^t$ denote the load vector incurred at step $t$ by our algorithm; for an integer $\ell$, we use $A^{\leq \ell}$ to denote the sequence $A^1, \ldots, A^\ell$, and similarly for other objects.

Theorem 2.3 seeks to bound $\| \sum_{t=n/2+1}^{n} o^t \|_{\text{max}}$.

Fact 2.6 (Suffices to analyze first half). The random variables $\| \sum_{t=n/2}^{n/2+t} o^t - \varepsilon \sum_{t=n/2}^{n/2+t} |o^t| \|_{\text{max}}$ and $\| \sum_{t=n/2}^{n/2+t} o^t - \varepsilon \sum_{t=n/2}^{n/2+t} |o^t| \|_{\text{max}}$ have the same distribution.

We want to show that the computed dual solutions $w^t$ capture our (non-linear) total load, i.e., that $\sum_t w^t \cdot o^t \approx \| \sum_t o^t \|_{\text{max}}$. To this end we formally show:

Fact 2.6 (Dual captures load). For every scenario,

$$\sum_{t=1}^{n/2} \langle w^t, o^t \rangle \geq \left\| \sum_{t=n/2}^{n/2} o^t - \varepsilon \sum_{t=n/2}^{n/2} |o^t| \right\|_{\text{max}} - \frac{M \log m}{\varepsilon} \geq \left\| \sum_{t=n/2}^{n/2} o^t - \varepsilon \sum_{t=n/2}^{n/2} |o^t| \right\|_{\text{max}} - \varepsilon \lambda^*.$$

Proof. We use the online learning algorithm from [AHK12, Section 2.3] to compute $w^t$, where the vectors $o^t$ are the adversarial vectors, the "learning rate" $\eta$ is set to $\varepsilon$, and the standard "experts"-like guarantee follows from Corollary 2.6 of that survey.

Let $\hat{o}^t := A^t \hat{p}^t$ denote the load incurred by the optimal solution $\hat{p}$ at step $t$. By our primal greedy choice of the $p^t$'s, we directly have the following.

Fact 2.7 (Optimality of algorithm wrt duals).

$$\sum_{t=1}^{n/2} \langle w^t, o^t \rangle = \sum_{t=1}^{n/2} w^t A^t p^t \leq \sum_{t=1}^{n/2} w^t A^t \hat{p}^t = \sum_{t=1}^{n/2} \langle w^t, \hat{o}^t \rangle.$$

From these facts, in order to give a guarantee in expectation, it suffices to show that

$$\mathbb{E}[\sum_{t=n/2}^{n/2} \langle w^t, \hat{o}^t \rangle] \leq \sum_{t=n/2}^{n/2} \langle \mathbb{E}[w^t], \mathbb{E}[\hat{o}^t] \rangle \leq \| \sum_{t=n/2}^{n/2} \hat{o}^t \|_{\text{max}} = \frac{\lambda^*}{2}. \quad (2.1)$$

Notice that these are easy implications if we were sampling with replacement, since in that case $w^t$ and $\hat{o}^t$ are independent. For the random permutation model we need to work harder.

Let $G_{i,t}$ be the good event that $\mathbb{E}[\hat{o}_i^t | A^{<t}] \leq (1 + 80(1 + \gamma)\varepsilon)\frac{\lambda^*}{n}$, i.e., that the expected occupation at timestep $t$ is at most $\approx \frac{\lambda^*}{n}$ even after conditioning on the history up to time $t - 1$. What we are able to show is that with high probability, the good events $G_{i,t}$ hold for all $t \leq n/2$ simultaneously; effectively, this says that for our purposes, sampling with and without replacement has essentially the same effect. Note that applying Bernstein’s inequality to each $\hat{o}_i^t$ and taking a union bound over the $t$'s would only give $\mathbb{E}[\hat{o}_i^t | A^{<t}] \leq (1 + \gamma \log n\varepsilon)\frac{\lambda^*}{n}$, with an extra log $n$ factor. To avoid this we employ a maximal version of Bernstein’s inequality.

Lemma 2.8. With probability $1 - \delta/m$, $G_{i,t}$ holds for all $i$ and all $t \leq n/2$; i.e., $\Pr(\mathbb{V}_i \cap \bigcup_{t \leq n/2} G_{i,t}) \leq \delta/m$.

Proof. Fix $i \in [m]$. Let $\mu = \frac{1}{n} \sum_{t=n}^{n} A_i^t \hat{p}^t \leq \frac{\lambda^*}{n}$ be the expected value of $\hat{o}_i^t$, which is independent of $t$. For a sequence of matrices $H^{<t} = (H^1, \ldots, H^{t-1})$, let $g(H^{<t})$ denote the occupation $\sum_{j<t} \hat{o}_i^j$.
conditioned on $A^{<t} = H^{<t}$. Notice that in every scenario, $\sum_{j=1}^{n} \hat{o}^j_i - g(A^{<t})$ gives the total $i$th load from items that have not shown up at times $j < t$, and hence

$$E[\hat{o}^j_i \mid A^{<t}] = \frac{\sum_{j=1}^{n} \hat{o}^j_i - g(A^{<t})}{n - (t - 1)} = \frac{n\mu - g(A^{<t})}{n - (t - 1)}. \quad (2.2)$$

Let $\alpha = 2n|\mu| + 2\gamma\lambda^*$ and let $E$ be the event that we have $\sum_{j=1}^{n} \hat{o}^j_i \geq t\mu - 20\varepsilon\alpha$ for all $t \leq n/2$; so the complement event $E^c$ equals $\max_{t \leq n/2}(\sum_{j=1}^{t} \hat{o}^j_i - t\mu) < -20\varepsilon\alpha$. To upper bound $Pr[E^c]$, let $V = \frac{1}{n}(\sum_{j=1}^{n} (A^j_i - \hat{o}^j_i - \mu)^2) = \text{Var}(\hat{o}^j_i)$. Since $\hat{o}^j_i \in [-M, M]$, we have that $V \leq M \cdot E[|\hat{o}^j_i - E[\hat{o}^j_i]|]$; from Lemma 2.2 we then have $V \leq M(2|\mu| + 2\gamma\lambda^*) \leq M\alpha/n$. By the maximal Bernstein’s inequality for sampling without replacement (Lemma 1.1 in Appendix B),

$$Pr[E^c] \leq 30 \exp\left(-\frac{20\varepsilon^2}{24} \frac{\alpha^2}{nV + 20\varepsilon\alpha(M/24)}\right) \leq 30 \exp\left(-\frac{\alpha^2}{1/24 + 20\varepsilon(M/24)}\right) = \frac{\delta}{m^2},$$

where we used that $\varepsilon \leq 1/80$, $\gamma \geq 1$, $\lambda^* \geq \frac{3M\log(m/\delta)}{\varepsilon^2}$, and $\delta \leq \varepsilon \leq 1/30$.

Now we claim that the event $E$ implies $\bigwedge_{t \leq n/2} G_{i,t}$. Every $A^{<n/2}$ satisfying $E$ has $g(A^{<t}) \geq (t - 1)\mu - 20\varepsilon(2n|\mu| + 2\gamma\lambda^*)$ for all $t \leq n/2$, and hence from (2.2)

$$E[\hat{o}^j_i \mid A^{<t}] \leq \frac{n\mu - (t - 1)\mu + 20\varepsilon(2n|\mu| + 2\gamma\lambda^*)}{n - (t - 1)} \leq \mu + 40\varepsilon\left(\frac{2|\mu| + 2\gamma\lambda^*}{n}\right) = \mu + 80\varepsilon\left(|\mu| + \frac{\gamma\lambda^*}{n}\right).$$

Now notice that since $80\varepsilon \leq 1$, we have $\mu + 80\varepsilon|\mu| \leq (1 + 80\varepsilon)\max\{|\mu, 0| \leq (1 + 80\varepsilon)\frac{\lambda^*}{n}$, and the claim that $E \implies \bigwedge_{t \leq n/2} G_{i,t}$ then holds. Finally, since for each $i \in [m]$ we have $\Pr(\bigwedge_{t \leq n/2} G_{i,t}) \geq \Pr(E) \geq 1 - \frac{\delta}{m^2}$, the lemma follows from a union bound over all $i \in [m]$.

Using this lemma it is not difficult to show that the bound in Theorem 2.3 holds in expectation. We sketch how this can be done in order to illustrate that little of the structure of the problem is actually used (and thus it might carry over to other settings); this follows directly from the lemma below and from Facts 2.5, 2.7.

**Lemma 2.9.** We have $E[\sum_{t \leq n/2} \langle w^t, \hat{o}^t \rangle] \leq \frac{\lambda^*}{n} + O(\gamma\varepsilon\lambda^*)$.

**Proof sketch.** Let $G := \bigwedge_{t \leq n/2} G_{i,t}$ be the “good” event. Let us split the expression:

$$E\left[\sum_{t \leq n/2} \langle w^t, \hat{o}^t \rangle\right] = E\left[\sum_{t \leq n/2} \langle w^t, \hat{o}^t \rangle \mid G\right] \Pr(G) + E\left[\sum_{t \leq n/2} \langle w^t, \hat{o}^t \rangle \mid G^c\right] \Pr(G^c).$$

For the second term, $\Pr(G^c) \leq \frac{\delta}{m} \leq \frac{\lambda^*}{m}$. Moreover, the sum $\sum_{t \leq n/2} \langle w^t, \hat{o}^t \rangle$ is independent of $G$, and at most $\sum_{t \leq n} \sum_{i} \hat{o}^t_i \leq m\lambda^*$. Multiplying the two gives us $O(\varepsilon\lambda^*)$.

To upper bound the first term, we use linearity of expectations to write it as $\sum_{t \leq n/2} E[\langle w^t, \hat{o}^t \rangle \mid G] \Pr(G)$, and bound each term. Since $w^t$ is determined by $A^t$, and $E[\hat{o}^t \mid A^{<t}, G] \leq \frac{\lambda^*}{n} (1 + O(\gamma\varepsilon))$, we get

$$E[\langle w^t, \hat{o}^t \rangle \mid A^{<t}, G] = \langle w^t, E[\hat{o}^t \mid A^{<t}, G] \rangle \leq \frac{\lambda^*}{n} (1 + O(\gamma\varepsilon)).$$

Taking expectations over $A^{<t}$ gives $E[\langle w^t, \hat{o}^t \rangle \mid G] \leq \frac{\lambda^*}{n} (1 + O(\gamma\varepsilon))$. This concludes the proof of the lemma. \[\square\]
2.5 The Guarantee with High Probability

We now show that Lemma 2.9 actually holds with high probability.

**Lemma 2.10.** With probability at least $1 - 3\delta$ we have

$$\sum_{t \leq n/2} \langle w^t, \hat{o}^t \rangle \leq (1 + 88\epsilon(1 + \gamma)) \frac{\lambda^*}{2}.$$  

If we were sampling with replacement, then it is not difficult to show that $\{\langle w^t, \hat{o}^t \rangle - \langle w^t, E[\hat{o}^t] \rangle\}_t$ is a martingale difference sequence, and consequently martingale concentration inequalities would imply $\sum_{t \leq n/2} \langle w^t, \hat{o}^t \rangle \leq \sum_{t \leq n/2} \|E[\hat{o}^t]\|_{\text{max}} \leq \frac{\lambda^*}{2}$ with good probability. For our case of sampling with replacement, we proceed similarly, but using the martingale difference sequence $\{\langle w^t, \hat{o}^t \rangle - \langle w^t, E[\hat{o}^t | A^<t] \rangle\}_t$, and using Lemma 2.3 to get that with good probability for all $t$ we have $E[\hat{o}^t | A^<t] \approx E[\hat{o}^t]$.

Define $Y_t := \langle w^t, \hat{o}^t - E[\hat{o}^t | A^<t] \rangle$. Since $w^t$ is determined by $A^<t$ we have that $E[Y_t | A^<t] = 0$ and hence $\{Y_t\}_t$ is a martingale difference sequence (adapted to $\{A^t\}_t$). The following controls the conditional variance of this martingale.

**Lemma 2.11.** With probability at least $1 - \frac{\delta}{m}$ we have

$$\sum_{t \leq n/2} \text{Var}(Y_t | A^<t) \leq 4M \cdot \sum_{t \leq n/2} \langle w^t, |E[\hat{o}^t | A^<t]| \rangle + 2M \gamma \lambda^*.$$  

**Proof.** Since $\hat{o}^t_i \in [-M,M]$, we have $Y_t \in [-2M,2M]$ and hence $Y_t^2 \leq 2M|Y_t|^2$. Since $E[Y_t | A^<t] = 0$, we have

$$\text{Var}(Y_t | A^<t) = E(Y_t^2 | A^<t) \leq 2M \cdot E|Y_t| | A^<t| = 2M \cdot E \left[ \sum_i w^t_i (\hat{o}^t_i - E[\hat{o}^t_i | A^<t]) \right] | A^<t| \leq 2M \cdot \sum_i w^t_i \cdot |E[\hat{o}^t_i | A^<t]| \leq 4M \cdot \sum_i w^t_i \cdot |E[\hat{o}^t_i | A^<t]| + \frac{4\gamma M \lambda^*}{n},$$

where the last inequality also uses $w^t \in \Delta^m$. Adding over all $t \leq n/2$ then gives the lemma. \qed

Now we use our martingale to prove the following.

**Lemma 2.12.** With probability at least $1 - \frac{2\delta}{m}$ we have

$$\sum_{t \leq n/2} \langle w^t, \hat{o}^t \rangle \leq \sum_{t \leq n/2} \langle w^t, E[\hat{o}^t | A^<t] \rangle + 4\epsilon \sum_{t \leq n/2} \langle w^t, |E[\hat{o}^t | A^<t]| \rangle + 2\epsilon \gamma \lambda^*.$$  

**Proof.** Letting $\alpha = 4 \sum_{t \leq n/2} \langle w^t, |E[\hat{o}^t | A^<t]| \rangle + 2\gamma \lambda^*$, we need to prove $\text{Pr}(\sum_{t \leq n/2} Y_t > \varepsilon \alpha) \leq \frac{2\delta}{m}$. Let $\mathcal{E}$ be the event that $\sum_{t \leq n/2} \text{Var}(Y_t | A^<t) \leq \alpha M$; Lemma 2.11 says that $\text{Pr}(\mathcal{E}) \leq \frac{\delta}{m}$. Then from Freedman’s inequality (Theorem A.3 in the Appendix) we get that

$$\text{Pr} \left( \sum_{t \leq n/2} Y_t > \varepsilon \alpha \wedge \mathcal{E} \right) \leq \exp \left( -\frac{\varepsilon^2 \alpha^2}{4M \alpha} \right) \leq \exp \left( -\frac{\varepsilon^2 2\gamma \lambda^*}{4M} \right) \leq \frac{\delta}{m},$$

where the last inequality uses $\gamma \geq 1$ and $\lambda^* \geq \frac{2M \log(m/\delta)}{\epsilon^2}$. Since from Lemma 2.11 $\text{Pr}(\mathcal{E}) \leq \frac{\delta}{m}$ a union bound shows that $\text{Pr}(\sum_{t \leq n/2} Y_t > \varepsilon \alpha) \leq \frac{2\delta}{m}$, concluding the proof. \qed
Proof of Lemma 2.10. Let $G = \bigwedge_{i} \bigwedge_{t \leq n/2} G_{i,t}$. Notice that since $4\varepsilon \leq 1$, $E[\hat{\omega}_{i}^{t} | A^{<t}] + 4\varepsilon E[\hat{\omega}_{i}^{t} | A^{<t}] \leq (1 + 4\varepsilon) \max\{E[\hat{\omega}_{i}^{t} | A^{<t}], 0\}$. Thus conditioned on $G$ we have $E[\hat{\omega}_{i}^{t} | A^{<t}] + 4\varepsilon E[\hat{\omega}_{i}^{t} | A^{<t}] \leq (1 + 4\varepsilon)(1 + 80\varepsilon(1 + \gamma)) \frac{\lambda^{*}}{m}$. Now taking a union bound over the failure probability of Lemma 2.12 and of the good event $G$ not holding true (from Lemma 2.8), with probability at least $1 - \frac{3\delta}{m}$ we have

$$\sum_{t \leq n/2} \langle w^{t}, \hat{\omega}_{i}^{t} \rangle \leq \sum_{t \leq n/2} (1 + 4\varepsilon)(1 + 80\varepsilon(1 + \gamma)) \frac{\lambda^{*}}{n} + 2\varepsilon \gamma \lambda^{*} = (1 + 88\varepsilon(1 + \gamma)) \frac{\lambda^{*}}{2}$$

and the lemma holds. \qed

Proof of Theorem 2.3. From triangle inequality we have

$$\left\| \sum_{t} o^{t} - \varepsilon \sum_{t} |o^{t}| \right\|_{\max} \leq \left( \left\| \sum_{t \leq n/2} o^{t} - \varepsilon \sum_{t \leq n/2} |o^{t}| \right\|_{\max} + \left\| \sum_{t > n/2} o^{t} - \varepsilon \sum_{t > n/2} |o^{t}| \right\|_{\max} \right)$$

Using Facts 2.6 and 2.7 and Lemma 2.10 with probability at least $1 - \frac{3\delta}{m}$ the first term in the right-hand side is at most $(1 + 88\varepsilon(1 + \gamma)) \frac{\lambda^{*}}{2}$. Since from Fact 2.5 the same bound holds for the second term of the right-hand side, a union bound concludes the proof of the theorem with constant $c_{1} = 88$. \qed

3 Solving Packing-Covering LPs via Load-Balancing

This section shows how to solve packing-covering LPs via a reduction to the load-balancing algorithm we developed in Theorem 2.3. Theorem 2.3 provided an estimate of the optimal value is available. The packing-covering LPs (PCLPs) we will solve will be of the following form:

$$\begin{align*}
\max & \sum_{t=1}^{n} \pi^{t} x^{t} \\
\text{st} & \sum_{t=1}^{n} A^{t} x^{t} \leq b \\
& \sum_{t=1}^{n} C^{t} x^{t} \geq d \\
x^{t} \in \Delta^{k} & \forall t \in [n]
\end{align*}$$

(3.3)

(3.4)

(3.5)

where all the data $\pi^{t}, A^{t}, C^{t}, b, d$ is non-negative and $\Delta^{k}$ denotes the “full simplex” $\{x \in [0,1]^{k} : \sum_{j} x_{j} \leq 1\}$. We use $m_{p}$ to denote the number packing constraints (3.3) and $m_{c}$ to denote the number of covering constraints (3.4), so our matrices have dimensions $A^{t} \in \mathbb{R}^{m_{p} \times k}$ and $C^{t} \in \mathbb{R}^{m_{c} \times k}$. We allow for either $m_{c}$ or $m_{p}$ to be zero—i.e., it could be a pure packing or covering LP. Note that the variables $\{x_{1}^{t}, x_{2}^{t}, \ldots, x_{k}^{t}\}$ in each block $t$ are bound together by $\sum_{j} x_{j}^{t} \leq 1$ — this is the “multiple-choice” setting, which is useful in many contexts. Given a packing-covering LP $\mathcal{L}$, we use $\text{opt}(\mathcal{L})$ to denote its optimal value; we omit $\mathcal{L}$ when it is clear from the context.

In the online version of PCLP, we are presented with a fixed PCLP $\mathcal{L}$, but in random order. We know upfront the number of time steps $n$ and the right-hand sides $b, d$. At each time step $t = 1, \ldots, n$, a “block” is sampled from $\mathcal{L}$ without replacement and revealed to the algorithm, which then outputs a vector $x^{t} \in \Delta^{k}$. Formally, define the random variables $\pi^{t}, A^{t}, C^{t}$ so that triples $(\pi^{1}, A^{1}, C^{1}), \ldots, (\pi^{n}, A^{n}, C^{n})$ are sampled from $\{ (\pi^{t}, A^{t}, C^{t}) \}_{t}$ uniformly without replacement. Define the randomly permuted LP $\mathcal{L}$

$$\mathcal{L} = \max \left\{ \sum_{t} \pi^{t} x^{t} : \sum_{t} A^{t} x^{t} \leq b, \sum_{t} C^{t} x^{t} \geq d, x^{t} \in \Delta^{k} \forall t \right\}$$

(3.6)

At time step $t$, the algorithm computes a (random) vector $x^{t} \in \Delta^{k}$ based on the information seen up to time $t$, i.e., $(\pi^{1}, A^{1}, C^{1}), \ldots, (\pi^{t}, A^{t}, C^{t})$, plus $n$ and $b, d$. We call such $\{x^{t}\}_{t}$ an online solution. The
online solution \( \{x^t\} \) is \( \varepsilon \)-feasible for \( \mathcal{L} \) if it satisfies the packing constraints (i.e., \( \sum_{t \leq n} A^t x^t \leq b \)) and almost satisfies the covering constraints (i.e., \( \sum_{t \leq n} C^t x^t \geq (1 - \varepsilon) d \)). The goal in the online PCLP is to obtain an online solution \( \{x^t\} \) which (with high probability) is \( \varepsilon \)-feasible, and gets a reward \( \sum_{t \leq n} \pi^t x^t \geq (1 - O(\varepsilon))\text{opt}(\mathcal{L}) \) (notice that \( \text{opt}(\mathcal{L}) = \text{opt}(\mathcal{L}) \), so again we are comparing against the optimal offline solution).

To state our main result for this section, we need one more concept. For a packing-covering LP \( \mathcal{L} \) of the form \( \text{PCLP} \), define its \textit{generalized width} to be

\[
\min \left\{ \min_{i,j,t} \frac{b_i}{a_{i,j}^t}, \min_{i,j,t} \frac{d_i}{c_{i,j}^t}, \min_{i,j} \frac{\text{opt}}{\pi_j^t} \right\}.
\]

(3.7)

Recall that \( \varepsilon_1 \) was the constant in Theorem 2.3.

\textbf{Theorem 3.1.} \textit{Consider a feasible (multiple-choice) packing-covering LP \( \mathcal{L} \), and let \( \varepsilon \leq \frac{\varepsilon_1}{4} \) and \( \delta \in (0, \varepsilon] \) be such that its generalized width is at least \( \frac{\log(m/\delta)}{\varepsilon^2} \). Suppose the algorithm is given an approximation \( \hat{\text{opt}} \) to the optimal value, as well as the right-hand sides \( b, d \) and values \( \varepsilon, \delta, n \). If the estimate \( \hat{\text{opt}} \in [\frac{\text{opt}}{2}, \text{opt}] \), then the algorithm LPviaLB below finds an online solution \( \{x^t\} \) that with probability at least \( 1 - \delta \) is \( (c_2 \varepsilon) \)-feasible for \( \mathcal{L} \) and has value \( \sum_t \pi^t x^t \geq (1 - c_3 \varepsilon) \hat{\text{opt}} \), for constants \( c_2, c_3 \geq 1 \).}

Before we describe the algorithm and give its analysis, let us record some useful corollaries. In the packing-only case where \( m_c = 0 \) and we have no covering constraints of the form (3.4), we can get a similar result that no longer requires exact knowledge of \( b \), the RHS of the packing constraints (3.3); we only need an approximation to \( b \). Moreover, we no longer require the estimate \( \hat{\text{opt}} \) to lie in \([\text{opt}/2, \text{opt}]\). Here is the theorem, which we will use in the subsequent sections. (We omit the proof, which is similar to that of Theorem 3.1.)

\textbf{Theorem 3.2} \textit{(Solving Packing LPs). Consider a packing LP \( \mathcal{L} \), and let \( \varepsilon \leq \varepsilon_3 := 2\sqrt{2} \varepsilon_1 \) and \( \delta \in (0, \varepsilon] \) be such that its generalized width is at least \( \frac{\log(m/b)}{\varepsilon^2} \). Suppose we are given as input an underestimate \( \hat{\text{opt}} \leq \text{opt} \), an approximation \( \hat{b} = (1 \pm \varepsilon)b \) to the packing RHS, and also the values of \( \varepsilon, \delta, n \). Define the estimation error \( \hat{\varepsilon} := \text{opt} - \hat{\text{opt}} \). Then the algorithm LPviaLB finds an online solution \( \{x^t\} \subseteq A^\varepsilon \) that with probability at least \( 1 - \delta \) is feasible for \( \mathcal{L} \) and has value \( \sum_t \pi^t x^t \geq (1 - c_3 \hat{\varepsilon}) \text{opt} - 2\hat{\varepsilon} \) for a universal constant \( c_3 \geq 1 \).}

Observe: both Theorems 3.1 and 3.2 require estimates of the optimal value. In Section 3, we show how to remove this assumption for “stable” packing-covering LPs, and hence fully solving these PCLPs online as long as the generalized width is small. Then in Section 6 we show how to solve packing-only LPs online without any assumptions on the profits. But for now, let return to the reduction from PCLPs to load-balancing and the proof of Theorem 3.1.

\textbf{3.1 The Algorithm LPviaLB}

The idea of the reduction is very simple: take the covering constraints (and the objective function) and flip their signs to make them packing constraints; also shift these negated values by adding a constant to both sides and then rescale the constraints so that all of them have same right-hand side. Notice that this creates negative entries in the left-hand side (negative loads); this is precisely why we had considered a load-balancing problem of this generality in the previous section.

To begin, we assume that \( \varepsilon^2 \geq \frac{\log(m/\delta)}{\varepsilon^2} \). Indeed, if \( \varepsilon^2 < \frac{\log(m/\delta)}{\varepsilon^2} \) then the assumption that the generalized width is at least \( \frac{\log(m/\delta)}{\varepsilon^2} \) implies \( c_{i,j}^t < \frac{d_i}{n} \) for all \( t, i \); this means the covering constraints,
if any, cannot be satisfied and the LP is infeasible. On the other hand, if \( m_c = 0 \) and there are no covering constraints, then another implication of the low value of \( \varepsilon \) is that \( a_{i,j}^c < \frac{b_i}{d_i} \); in this case the packing constraints are so loose that the optimal solution is to choose the most profitable choice for each time \( t \) independently.

Given a packing-covering LP \( \mathcal{L} \), define the matrices \( H^1, \ldots, H^n \) with \( k + 1 \) columns and \( m_p + m_c + 1 \) rows (indexed from 0 to \( m := m_p + m_c \)) as follows: the zeroth row of \( H^t \) equals the vector

\[
H^t_{0,*} := \left( \frac{2}{n} - \frac{\pi^t_{\text{opt}}}{\pi^t_{\text{opt}}}, \ldots, \frac{2}{n} - \frac{\pi^t_{\text{opt}}}{\pi^t_{\text{opt}}}, \frac{2}{n} \right);
\]

for \( i \in \{1, \ldots, m_p\} \), the \( i^{th} \) row of \( H^t \) is

\[
H^t_{i,*} := \left( \frac{a_i^t}{b_i}, \ldots, \frac{a_i^t}{b_i}, 0 \right),
\]

and for \( i \in \{m_p + 1, \ldots, m_p + m_c\} \), the \( (m_p + i)^{th} \) row of \( H^t \) is

\[
H^t_{m_p+i,*} := \left( \frac{2}{n} - \frac{\lambda^{t,\text{opt}}}{\lambda^{t,\text{opt}}}, \ldots, \frac{2}{n} - \frac{\lambda^{t,\text{opt}}}{\lambda^{t,\text{opt}}}, \frac{2}{n} \right).
\]

(For simplicity, we assume that \( \hat{\text{opt}} \) and all \( b_i, d_i \)s are strictly positive.)

The algorithm LPviaLB can be thought of as having three phases. In the first phase, it computes the matrices \( H^1, \ldots, H^n \). In the second, it runs load-balancing algorithm expertLB over the instance \( \{H^t\}_t \) with parameters \( M = \frac{2\varepsilon^2}{\log(m/\delta)} \), \( \varepsilon' = 4\varepsilon \) and \( \delta \) (given), obtaining a solution \( \{\hat{x}^t\}_t \). In the last phase, the algorithm simply outputs the scaled and truncated solution \( \{\hat{x}^t\}_t \), for \( \hat{x}^t_j := \frac{(1-2\varepsilon')}{(1+c_1\varepsilon')} x^t_j \) with \( j = 1, \ldots, k \), as an (approximate) solution to \( \mathcal{L} \), where \( c_1 \) is the constant from Theorem 2.3. Notice that all the steps in this algorithm can be implemented to run in an online manner.

### 3.2 The Proof of Theorem 3.1

The next claim bounds the value of the optimal solution to the instance \( \{H^t\}_t \) by relating feasible solutions of this instance to those of the LP \( \mathcal{L} \). Let \( \lambda^* \) denote the optimal value for the load-balancing instance \( \{H^t\}_t \). We will repeatedly use the assumption that \( \text{opt} \) is strictly positive and \( \frac{\text{opt}}{2} \leq \hat{\text{opt}} \leq \text{opt} \).

**Claim 3.3.** \( \frac{1}{2} \leq \lambda^* \leq 1. \)

**Proof.** (Upper bound.) Let \( \{x^t\}_t \) be an optimal solution for \( \mathcal{L} \). The feasibility of \( \{x^t\}_t \) guarantees that for \( i \in \{1, \ldots, m_p\} \), the load \( \sum_t H^t_{i,*} x^t = \frac{1}{b_i} \sum_t A^t_{i,*} x^t \leq 1 \). For \( i \in \{1, \ldots, m_c\} \), the load \( \sum_t H^t_{m_c+i,*} x^t = 2 - \frac{1}{d_i} \sum_t C^t_{i,*} x^t \leq 1 \). Finally, the optimality of \( \{x^t\}_t \) guarantees that \( \sum_t H^t_{0,*} x^t = 2 - \frac{1}{\text{opt}} \left( \sum_t \pi^t x^t \right) \leq \frac{2}{\text{opt}} \leq 1 \), where the last inequality follows from \( \text{opt} \geq \hat{\text{opt}} \).

(Lower bound.) Consider an optimal solution \( x^1, \ldots, x^n \in \Delta^k \) for the load-balancing instance \( \{H^t\}_t \) and hence \( \lambda^* := \|\sum_t H^t x^t\|_{\text{max}} \). For any \( i \in \{1, \ldots, m_p\} \), we have \( \lambda^* \geq \sum_t H^t_{i,*} x^t = \frac{1}{b_i} \sum_t A^t_{i,*} x^t \); it then follows that \( \{x^t\}_t \) is feasible for the packing constraints. Moreover, for any \( i \in \{1, \ldots, m_c\} \), we have \( \lambda^* \geq \sum_t H^t_{m_c+i,*} x^t = 2 - \frac{1}{d_i} \sum_t C^t_{i,*} x^t \); and hence \( \sum_t C^t \frac{x^t}{\lambda^*} \geq \frac{2}{\lambda^*} \). Using \( \lambda^* \leq 1 \) from the previous argument, the above expression is at least \( d \), so \( \{x^t\}_t \) is also feasible for the covering constraints. Moreover, we have \( \lambda^* \geq \sum_t H^t_{0,*} x^t = 2 - \frac{1}{\text{opt}} \sum_t \pi^t x^t \), and hence \( \sum_t \pi^t \frac{x^t}{\lambda^*} \geq \frac{2}{\lambda^*} \text{opt} \). This must be at most the optimal value \( \text{opt} \) for the LP \( \mathcal{L} \), so \( \text{opt} \geq \frac{2}{\lambda^*} \hat{\text{opt}} \), or \( \lambda^* \geq \frac{2 \text{opt}}{\text{opt} + \hat{\text{opt}}} \geq \frac{1}{2} \), where the last inequality follows from \( \frac{\text{opt}}{2} \leq \hat{\text{opt}} \leq \text{opt} \). \( \square \)

**Claim 3.4.** The instance \( \{H^t\}_t \) is \((M,4)\)-well-bounded.
Proof. Using the lower bound $\lambda^* \geq \frac{1}{2}$ from the previous claim, it suffices to show that each row of $H^t$ contains entries that all lie in the interval $[-M, \frac{2}{n}]$, or all in the interval $[-\frac{2}{n}, M]$.

For the zeroth row, using our width lower bound, we have $H^t_{0,j} \geq \frac{-\varepsilon^t}{\text{opt}} \geq \frac{-\text{Mopt}}{2\text{opt}} \geq -M$, where the last inequality uses $\text{opt} \geq \frac{1}{n}$; since $H^t_{0,j} \leq \frac{2}{n}$, it follows that $H^t_{0,j} \in [-M, \frac{2}{n}]$. For $i \in \{1, \ldots, m_t\}$, we have $0 \leq H^t_{i,j} \leq \frac{\varepsilon^t}{b_i} \leq M$. And for $i \in \{1, \ldots, m_c\}$, we have $H^t_{m_p+i,j} \leq \frac{2}{n}$ and $H^t_{m_p+i,j} \geq -\varepsilon^t \frac{c_i}{n} \geq -M$. 

Proof of Theorem 3.1. Claim 3.4 guarantees that $\{H^t_i\}_i$ is $(M, 4)$-well-bounded; Claim 3.3 plus the definition of $\varepsilon'$ guarantee that $\lambda^* \geq \frac{3M\log(m/c)}{(\varepsilon')^2}$. Applying Theorem 2.3 the returned solution $\{\tilde{x}^t_i\}_i$ satisfies with probability at least $1 - \delta$

$$\forall i = 0, \ldots, m_c, \sum_t H^t_{i,*} \tilde{x}^t - \varepsilon' \sum_t |H^t_{i,*} \tilde{x}^t| \leq \lambda^*(1 + c_1\varepsilon') \leq \frac{(1 + \varepsilon'(1 + c_1\varepsilon'))}{(1 - \varepsilon)}$$ (3.8)

where the last inequality follows from Claim 3.3. Let $E$ be the event that this bound holds. For $i \in \{1, \ldots, m_p\}$, since $H^t_{i,*} = \frac{A^t_i}{b_i}$ is non-negative, this means that conditioned on $E$, $\sum_t \sum_{j \leq k} A^t_{i,j} \tilde{x}^t_j \leq b_i \frac{(1 + \varepsilon(1 + c_1\varepsilon'))}{(1 - \varepsilon)}$ (since $\varepsilon' \geq 2\varepsilon$) and so $\{\tilde{x}^t_i\}_i$ is feasible for the packing constraints.

For $i = 0$ we have $|H^t_{i,*} \tilde{x}^t| \leq \frac{2}{n} + \frac{\text{opt}}{\text{opt}} \sum_{j \leq k} \pi^t_{j} \tilde{x}^t_j$ and the left-hand side of inequality (3.8) is at least

$$\left(2 - \frac{\text{opt}}{\text{opt}} \sum_{j \leq k} \pi^t_{j} \tilde{x}^t_j\right) - \varepsilon' \left(2 + \frac{\text{opt}}{\text{opt}} \sum_{j \leq k} \pi^t_{j} \tilde{x}^t_j\right) = 2(1 - \varepsilon') - (1 + \varepsilon'(1 + c_1\varepsilon') \frac{1}{(1 - \varepsilon)} \sum_{j \leq k} \pi^t_{j} \tilde{x}^t_j. $$

Conditioned on $E$, inequality (3.8) and algebraic manipulations give $\sum_t \pi^t \tilde{x}^t = (1 - O(\varepsilon')) \sum_t \sum_{j \leq k} \pi^t_{j} \tilde{x}^t_j \geq (1 - O(\varepsilon')) \text{opt}$. An identical proof shows that conditioned on $E$, each covering constraint is $O(\varepsilon')$-approximately satisfied, i.e., $\sum_t C^t \tilde{x}^t \geq (1 - O(\varepsilon'))d$. Replacing the value of $\varepsilon' = 4\varepsilon$ concludes the proof.

4 Estimating the Optimal Value for Packing/Covering LPs

We now turn to the problem of estimating the value of $\text{opt}$ for a packing-covering problem, so that we can use it in conjunction with Theorem 3.1. Doing this for packing LPs with large generalized width is not difficult, but the covering constraints cause certain complications. To handle these, we require a kind of stability property, loosely asking that the covering constraints are not “very tight” — formally, if we reduce the covering requirements by a little, the optimal value should not increase by a lot.

**Definition 4.1 (Stability).** A packing-covering LP $\mathcal{L}$ is called $(\varepsilon, \sigma)$-stable if any optimal $\varepsilon$-feasible solution for it has value at most $(1 + \sigma\varepsilon)\text{opt}(\mathcal{L})$.

For a randomly permuted LP $\mathcal{L}$, the basic idea to obtain an estimate of $\text{opt}(\mathcal{L})$ is to see the first $n/2$ random blocks and form a sampled LP with these blocks and right-hand side scaled by a factor of 1/2; computing an optimal $O(\varepsilon)$-solution for this sampled LP should give a good estimate of $\text{opt}(\mathcal{L})/2$. To make this formal, we introduce the following operations on LPs.

**Definition 4.2 (Restricted LP).** Given a packing-covering LP $\mathcal{L}$, and a subset $I$ of $[n]$, the restricted LP $\mathcal{L}^I$ is obtained by retaining only the columns of $\mathcal{L}$ that belong to $I$, and setting the right hand side $\text{rhs}(\mathcal{L}^I)$ to be $\frac{|I|}{n} \text{rhs}(\mathcal{L})$.

Notice that $\mathcal{L}^{\{1, \ldots, \frac{n}{2}\}}$ corresponds to the sampled LP described in the previous paragraph.
Definition 4.3. Given a packing-covering LP \( \mathcal{L} \), and \( \varepsilon \in [0, 1) \), we define \( \mathcal{L}(1 - \varepsilon) \) as the LP obtained by multiplying the right-hand side of the covering constraints by \( (1 - \varepsilon) \); the packing constraints remain unchanged.

Notice that this operation captures exactly set of the \( \varepsilon \)-feasible solutions of an LP. For an LP of the form \( \text{PCLP} \), its width is defined as

\[
\min \left\{ \min_{t,i,j} \frac{b_{ij}}{a_{ij}}, \min_{t,i,j} \frac{d_{ij}}{c_{ij}} \right\} ;
\]

(4.9)

unlike the generalized width (defined in (3.7)) it does not depend on the the ratio of individual item values to \( \text{opt} \).

The following lemma gives bounds on how close the optimal value of \( \mathcal{L}' \) is to the value of the original LP \( \mathcal{L} \). The first part of the lemma is a straightforward application of Bernstein’s inequality, while the part uses \((\varepsilon, \sigma)\)-stability to control the variance of the (dual of the) sampled LP. Its proof is presented in Appendix D.

Lemma 4.4. For \( \varepsilon, \delta \in (0, 1) \), consider a packing-covering LP \( \mathcal{L} \) with width at least \( \frac{32 \log(m/\delta)}{\varepsilon^2} \). Assume that all item values are at most \( \rho \cdot \text{opt}(\mathcal{L}) \left( \frac{\varepsilon^2}{32 \log(m/\delta)} \right) \) for \( \rho \geq 1 \). Let \( k \geq \varepsilon^2 n \) and let \( I \) be a \( k \)-subset of \([n]\). Then for \( \varepsilon' = \varepsilon \sqrt{k} \) we have the following guarantees for the corresponding randomly permuted LP \( \mathcal{L}_I \):

(a) If \( \bar{x} \) is feasible for \( \mathcal{L} \), then \( (1 - \frac{\varepsilon'}{2})\bar{x}|_I \) is feasible for \( \mathcal{L}'(1 - \varepsilon') \) with probability \( 1 - \frac{\delta}{2} \).

(b) With probability at least \( 1 - \delta \), \( \text{opt}(\mathcal{L}'(1 - \varepsilon')) \geq (1 - \varepsilon' - \rho \varepsilon'/2) \cdot \frac{k}{n} \text{opt}(\mathcal{L}) \).

(c) \( \mathbb{E}[\text{opt}(\mathcal{L}'(1 - \varepsilon'))] \geq (1 - \varepsilon' - \rho \varepsilon'/2)(1 - \delta) \cdot \frac{k}{n} \text{opt}(\mathcal{L}) \).

Moreover, if \( \mathcal{L} \) is \((\varepsilon, \sigma)\)-stable, then the following upper bounds hold:

(d) With probability at least \( 1 - \delta \), \( \text{opt}(\mathcal{L}'(1 - \varepsilon')) \leq (1 + 2\sigma \varepsilon' + \rho \varepsilon') \cdot \frac{k}{n} \text{opt}(\mathcal{L}) \)

(e) \( \mathbb{E}[\text{opt}(\mathcal{L}'(1 - \varepsilon'))] \leq (1 + \sigma \varepsilon') \cdot \frac{k}{n} \text{opt}(\mathcal{L}) \).

5 Solving Packing/Covering LPs with Unknown OPT

We can now combine the online LP solver from Section 3 that required estimates on \( \text{opt} \), with the estimates of \( \text{opt} \) from Section 4 to get \( \varepsilon \)-feasible solutions to online packing/covering LPs. Observe, however, that a good estimate of \( \text{opt} \) requires us to look at a constant fraction of the LP, and hence a naive implementation uses the first \( n/2 \) blocks for the estimate, and uses this on the second \( n/2 \) blocks to get a solution with value \( \approx \frac{1}{2} \text{opt} \); this is given in Theorem 5.1. We use this idea to learn \( \text{opt} \) dynamically at multiple scales and get value \((1 - O(\varepsilon))\text{opt} \), which appears as Theorem 5.2.

We use the shorthand \( \mathcal{L}^{\leq k} \) to denote \( \mathcal{L}^{(1, \ldots, k)} \), and similarly \( \mathcal{L}^{> k} := \mathcal{L}^{(k + 1, \ldots, n)} \).

5.1 One-time Learning OPT

Theorem 5.1. Let \( \varepsilon \leq \frac{1}{4} \), and \( \delta \in (0, \varepsilon] \). Consider a packing-covering LP \( \mathcal{L} \) with generalized width at least \( \frac{32 \log(m/\delta)}{\varepsilon^2} \) and that is \((\varepsilon, \sigma)\)-stable for \( \sigma \in [1, \frac{1}{12 \sqrt{2}} \varepsilon] \). Suppose that the number \( n \) of columns of \( \mathcal{L} \), the right-hand sides of \( \mathcal{L} \), \( \varepsilon \), \( \delta \) and \( \sigma \) are known a priori. Then with probability at least \( 1 - 5\delta \) we can find an online \( \varepsilon(\sqrt{2} + c_2) \)-feasible solution to the random LP \( \mathcal{L}^{> \frac{n}{2}} \) with value at least

\[
(1 - c_3 \varepsilon) \text{opt}(\mathcal{L}^{\leq \frac{n}{2}}) - \varepsilon (5 \sqrt{2} \sigma) \text{opt}(\mathcal{L}).
\]
Proof. Let $\varepsilon' = \varepsilon \sqrt{2}$. From Lemma 4.4 we have that
\[
\Pr \left[ (1 - 2\varepsilon') \leq \frac{\text{opt}(\mathcal{L}^{\leq n/2}(1 - \varepsilon'))}{2 \cdot \text{opt}(\mathcal{L})} \leq (1 + 3\varepsilon' \sigma) \right] \geq 1 - 2\delta,
\]
and similarly for $\mathcal{L}^{> n/2}$ (these estimates use $\sigma \geq 1$). Then define $\hat{\text{opt}} := \text{opt}(\mathcal{L}^{\leq n/2}(1 - \varepsilon'))/(1 - 3\varepsilon' \sigma)$; with probability at least $1 - 4\delta$ the value of $\hat{\text{opt}}$ lies between $\text{opt}(\mathcal{L}^{> n/2}(1 - \varepsilon')) - \Delta$ and $\text{opt}(\mathcal{L}^{> n/2}(1 - \varepsilon'))$, where $\Delta \leq 5\varepsilon' \sigma \cdot \text{opt}(\mathcal{L})$ (standard calculations provide this upper bound on $\Delta$), and also $\Delta \geq \text{opt}(\mathcal{L}^{> n/2}(1 - \varepsilon'))/2$ (this uses the upper bound on $\sigma$).

Then we can employ algorithm LPviaLB to $\mathcal{L}^{> n/2}(1 - \varepsilon')$ using $\hat{\text{opt}}$. Theorem 3.1 guarantees that with probability at least $1 - 5\delta$ we get an $(c_2 \varepsilon + \varepsilon')$-feasible solution for $\mathcal{L}^{> n/2}$ with value at least $(1 - c_3 \varepsilon) \hat{\text{opt}} \geq (1 - c_3 \varepsilon) \text{opt}(\mathcal{L}^{> n/2}) - 5\varepsilon' \sigma \cdot \text{opt}(\mathcal{L})$, and the theorem follows.

Observe that Theorem 5.1 gives us value $\approx \text{opt}(\mathcal{L}^{\leq n/2}) \approx \frac{1}{2} \text{opt}(\mathcal{L})$. To do better, the next section shows how to dynamically learn $\text{opt}$ at $\log \varepsilon^{-1}$ different scales.

5.2 Dynamically Learning OPT

Theorem 5.2. Let $\varepsilon \leq (\frac{d}{4})^2$, and $\delta \in (0, \varepsilon]$. Consider a packing-covering LP $\mathcal{L}$ with generalized width at least $\frac{32 \log(m/d)}{\varepsilon^2}$ and that is $(\varepsilon, \sigma)$-stable for $\sigma \in \left[ 1, \frac{1}{12 \sqrt{2} \varepsilon^2} \right]$. Suppose that the number $n$ of columns of $\mathcal{L}$, the right-hand sides of $\mathcal{L}$, $\varepsilon$, $\delta$ and $\sigma$ are known a priori. Then we can find an online solution to the random LP $\mathcal{L}$ that is $\varepsilon(\sqrt{2} + c_2)$-feasible with probability at least $1 - 5\delta \log \varepsilon^{-1}$, and has expected value at least $1 - O(\varepsilon \sigma + \delta \log \varepsilon^{-1})\text{opt}(\mathcal{L})$.

For $i = 0, 1, \ldots, \log \varepsilon^{-1}$, define the set $S_i := \{1, \ldots, 2^i \varepsilon n\}$. Also let $\varepsilon_i := \varepsilon \cdot \sqrt{\frac{2}{|S_i|}}$, so that $\varepsilon_i = \sqrt{\varepsilon/2^i}$. Since $\varepsilon_i \leq 1/4$, note that the restricted LPs $\mathcal{L}^{S_i}(1 - \varepsilon_i)$ have generalized width at least $\frac{32 \log(m/d)}{\varepsilon_i^2}$.

Moreover, from Lemma 3.2 with probability at least $1 - 2\delta$ the LP $\mathcal{L}^{S_i}(1 - \varepsilon_i)$ is $(\varepsilon_i, 10\sigma)$-stable. In this case, Theorem 5.1 can be employed to obtain an almost-feasible approximate solution for $\mathcal{L}^{S_i}$. (Formally, we condition on the set of columns in $\mathcal{L}^{S_i}$ (but not their order), so for each conditioning we get a random permutation LP to which we apply the theorem.) By the doubling sizes of the sets $S_i$, note that the sets $S_i \setminus S_{i-1}$ for $i \geq 1$ are disjoint, and partition the integers in $(\varepsilon n, n]$.

The algorithm for solving $\mathcal{L}$ is simple: for each $i \geq 1$, compute a solution $\mathbf{x}(i)$ for $\mathcal{L}^{S_i}(1 - \varepsilon_i)$ using Theorem 5.1 and output their “union” $\mathbf{x}$ as the solution for $\mathcal{L}$; more precisely, set $\mathbf{x} = \mathbf{x}(i)^t$ for the unique $i$ such that $t \in S_i \setminus S_{i-1}$.

To analyze this solution, we first show feasibility. Let $\mathcal{E}$ denote the event that all LPs $\mathcal{L}^{S_i}(1 - \varepsilon_i)$ are $(\varepsilon_i, 10\sigma)$-stable and that the guarantee of Theorem 5.1 holds for all $\mathcal{L}^{S_i}(1 - \varepsilon_i)$’s. By a union bound, $\Pr[\mathcal{E}] \geq 1 - 7\delta \log(1/\varepsilon)$. Conditioned on $\mathcal{E}$, each $\mathbf{x}(i)$ is $O(\varepsilon_i)$-feasible for $\mathcal{L}^{S_i}$ and hence
\[
\sum_t A^t \mathbf{x} \leq \sum_{i=1}^{\log(1/\varepsilon)} \sum_{t \in S_i \setminus S_{i-1}} A^t \mathbf{x}(i)^t \leq \sum_{i=1}^{\log(1/\varepsilon)} \frac{|S_i|}{n} b = b \sum_{i=1}^{\log(1/\varepsilon)} 2^i \varepsilon \leq b,
\]
and hence $\mathbf{x}$ satisfies the packing constraints. Similarly, using the definition of $\varepsilon_i$ we have
\[
\sum_t C^t \mathbf{x} \geq \sum_{i=1}^{\log(1/\varepsilon)} (1 - O(\varepsilon_i)) \frac{|S_i|}{n} d \geq d(1 - \varepsilon - \varepsilon^2) - d \sum_{i=1}^{\log(1/\varepsilon)} 2^i \varepsilon O(\varepsilon_i) = d(1 - \varepsilon - \varepsilon^2) - d \left( O(\varepsilon^{3/2}) \sum_{i=1}^{\log(1/\varepsilon)} \sqrt{2^i} \right) \geq d(1 - O(\varepsilon)).
\]
Thus, $\bar{x}$ is $O(\varepsilon)$-feasible for $\mathcal{L}$ whenever $\mathcal{E}$ holds, which happens with probability $1 - O(\delta \log \varepsilon^{-1})$.

Now we analyze the value $\text{val} := \sum \pi^t \bar{x}^t$ of this solution. Let $\text{val}_i := \text{opt}(\mathcal{L}^{S_i \backslash S_{i-1}}(1 - \varepsilon_i)) - O(\varepsilon_i)\text{opt}(\mathcal{L}^{S_i}(1 - \varepsilon_i))$, which is the guarantee of Theorem 5.1 with respect to $\mathcal{L}^{S_i}(1 - \varepsilon_i)$. Then using the non-negativity of the objective function we have

$$E[\text{val}] \geq E[\text{val} | \mathcal{E}] \Pr(\mathcal{E}) \geq \sum_{i \geq 1} E[\text{val}_i | \mathcal{E}] \Pr(\mathcal{E}) = \sum_{i \geq 1} \left( E[\text{val}_i] - E[\text{val}_i | \mathcal{E}] \Pr(\mathcal{E}) \right). \quad (5.10)$$

The next two claims will bound the right-hand side of this expression.

**Claim 5.3.** $\sum_{i \geq 1} E[\text{val}_i | \mathcal{E}] \leq (1 + O(\sigma \varepsilon)) \text{opt}(\mathcal{L})$.

*Proof.* For $i \geq 1$, let $x(i)$ be an optimal solution for $\mathcal{L}^{S_i \backslash S_{i-1}}(1 - \varepsilon_i)$, and let $x$ be its “union” given by $x^t = x^t(i)$ for the unique $i$ such that $t \in S_i \backslash S_{i-1}$.

The argument above about the feasibility of $\bar{x}$ shows that actually $x$ is $O(\varepsilon)$-feasible for $\mathcal{L}$ in every scenario; thus, $\sum_{i \geq 1} \text{val}_i$ is at most the value of the best $O(\varepsilon)$-feasible solution for $\mathcal{L}$. The $(\varepsilon, \sigma)$-stability of $\mathcal{L}$ plus Lemma 4.4.1 implies that $\sum_{i \geq 1} \text{val}_i \leq (1 + O(\sigma \varepsilon)) \text{opt}(\mathcal{L})$. The claim follows by taking expectation conditioned on $\mathcal{E}$ and noticing that $\text{opt}(\mathcal{L})$ is deterministic. $\square$

**Claim 5.4.** $\sum_{i \geq 1} E[\text{val}_i] \geq (1 - O(\varepsilon) - \delta \log \varepsilon^{-1}) \text{opt}(\mathcal{L})$.

*Proof.* Using Lemma 4.4.1(e) and the upper bound on $\sigma$, $E[\text{opt}(\mathcal{L}^{S_i}(1 - \varepsilon_i))] \leq 2^i(1 + \sigma \varepsilon) \text{opt}(\mathcal{L}) \leq O(\varepsilon)2^i \text{opt}(\mathcal{L})$, and hence

$$\sum_{i \geq 1} O(\varepsilon_i) E[\text{opt}(\mathcal{L}^{S_i}(1 - \varepsilon_i))] \leq \sum_{i=1}^{\log(1/\varepsilon)} O(\varepsilon_i) \left( \varepsilon 2^i \text{opt}(\mathcal{L}) \right) \leq O(\varepsilon) \text{opt}(\mathcal{L}). \quad (5.11)$$

Now, to lower bound $\sum_{i \geq 1} \text{opt}(\mathcal{L}^{S_i \backslash S_{i-1}}(1 - \varepsilon_i))$, let $x$ be a optimal solution for $\mathcal{L}$ and let $x(i)$ be the solution scaled down by $(1 - \frac{\varepsilon_i}{2})$ and restricted to the $2^{i-1} \varepsilon_n$ columns in $S_i \backslash S_{i-1}$. From Lemma 4.4.1 we have that $x(i)$ is feasible for $\mathcal{L}^{S_i \backslash S_{i-1}}(1 - \varepsilon_i)$ with probability at least $1 - \delta$.

Let $\mathcal{G}$ be the good event that for all $i$ we have $x(i)$ feasible for $\mathcal{L}^{S_i \backslash S_{i-1}}(1 - \varepsilon_i)$; we get $\Pr[\mathcal{G}] \geq 1 - (\log \varepsilon^{-1}) \delta$ by a union bound. Then, by definition of $x(i)$’s, for every scenario in $\mathcal{G}$ we have

$$\sum_{i \geq 1} \text{opt}(\mathcal{L}^{S_i \backslash S_{i-1}}(1 - \varepsilon_i)) \geq \sum_i \pi^t x^t - \sum_{i \geq 1} \frac{\varepsilon_i}{2} \sum_{t \in S_i \backslash S_{i-1}} \pi^t x^t.$$

Here we have defined $\varepsilon_0 = 1$ for convenience. Using the non-negativity of the objective function, we get

$$E \left[ \sum_{i \geq 1} \text{opt}(\mathcal{L}^{S_i \backslash S_{i-1}}(1 - \varepsilon_i)) \right] \geq E \left[ \sum_{i \geq 1} \text{opt}(\mathcal{L}^{S_i \backslash S_{i-1}}(1 - \varepsilon_i)) \mid \mathcal{G} \right] \Pr[\mathcal{G}]$$

$$\geq \left( \sum_i \pi^t x^t \right) \Pr[\mathcal{G}] - E \left[ \sum_{i \geq 0} \frac{\varepsilon_i}{2} \sum_{t \in S_i \backslash S_{i-1}} \pi^t x^t \mid \mathcal{G} \right] \cdot \Pr[\mathcal{G}]. \quad (5.12)$$

Moreover, again by the non-negativity of the $\pi^t$’s, we get

$$E \left[ \sum_{i \geq 0} \frac{\varepsilon_i}{2} \sum_{t \in S_i \backslash S_{i-1}} \pi^t x^t \right] \geq E \left[ \sum_{i \geq 0} \frac{\varepsilon_i}{2} \sum_{t \in S_i \backslash S_{i-1}} \pi^t x^t \mid \mathcal{G} \right] \cdot \Pr[\mathcal{G}]$$
Putting these two inequalities together, and using the optimality of \( \bf{x} \), eq. (5.12) is at least

\[
\opt(L) \cdot \Pr[\mathcal{G}] - \mathbb{E} \left[ \sum_{t \geq 0} \frac{\varepsilon_i}{2} \sum_{t \in S \setminus S_{t-1}} \pi^t x^t \right].
\]

(5.13)

The rest is immediate: \( \mathbb{E}[\sum_{t \in S \setminus S_{t-1}} \pi^t x^t] = \varepsilon 2^{t-1} \opt(L) \), and hence

\[
\mathbb{E} \left[ \sum_{t \geq 0} \frac{\varepsilon_i}{2} \sum_{t \in S \setminus S_{t-1}} \pi^t x^t \right] = \sum_{t \geq 0} \varepsilon_i \varepsilon 2^{t-1} \opt(L) \leq O(\varepsilon) \opt(L).
\]

Putting it all together, we get that the right-hand side of (5.13) is at least \( (\Pr[\mathcal{G}] - O(\varepsilon)) \opt(L) \geq (1 - \delta \log \varepsilon^{-1} - O(\varepsilon)) \opt(L) \), and the claim follows.

Using the claims with inequality (5.10) implies that

\[
\mathbb{E}[val] \geq (1 - O(\varepsilon) - \delta \log \varepsilon^{-1}) \opt(L) - (1 + O(\varepsilon)) \opt(L) \cdot \Pr(\mathcal{E})
= \opt(L)(1 - O(\varepsilon) - \delta \log \varepsilon^{-1} - O(\delta \varepsilon \sigma \log \varepsilon^{-1})) = \opt(L)(1 - O(\varepsilon + \delta \log \varepsilon^{-1})),
\]

using that \( \delta \leq \varepsilon \) and \( \sigma \geq 1 \), which concludes the proof of Theorem 5.2.

**Observation 5.5.** The above proof of Theorem 5.2 does not explicitly use that fact that the LP \( L \) has item values at most \( \opt(L)/B \) for some \( B = \Theta(\log(B/\delta)) \), this is only required to invoke Theorem 5.7. This will be useful, since we can use the proof of Theorem 5.2 for the next section where we do not assume anything about the magnitudes of item values.

### 6 Solving Arbitrary Packing LPs

In this section we consider packing-only LPs (PPLPs) of the form \( L = \max \{ \sum_i \pi^t x_t \mid Ax \leq b, x \in [0, 1]^n \} \). Note that we do not have any covering constraints, and also there are no multiple-choice constraints—each of the load matrices \( A^t \) are just column vectors, and at each step one more column of the matrix \( A \) is revealed to us. In this setting we show how find a solution of value \( (1 - \varepsilon) \opt \) without any assumptions about the magnitudes of the values \( \pi_t \).

The main idea in this section is to show that items that have high enough value can just be added to our solution (i.e., we can set \( x_i = 1 \) for these items), such that (a) the value of these items plus the LP on the remaining items is still a \( (1 - \varepsilon) \)-approximation to the original optimal value, and (b) the right-hand side of the remaining LP is still large, so that we can use the algorithms developed in the previous sections.

Some notation used throughout this section. Let us denote the set of PPLPs with \( n \) columns by \( \text{PPLP}_n \). Recall the **width** (defined in (4.9)) of a packing LP is \( \max \{ \frac{b_i}{a_{ij}} \} \). We define \( B := \frac{\log(B/\delta)}{\varepsilon} \), and will want the width of all LPs to be \( \Omega(B) \). By scaling, we will always assume that entries in each column \( A^t \) lie in \([0, 1]\) and that the right-hand sides \( b_i \)'s are at least the width of the LP.

#### 6.1 Skimming the High Value Items

The first piece of the puzzle is to show the existence of a good skimming threshold—this is a value \( \tau \) such that blindly picking all the items with value at least \( \tau \) and solving the residual LP is a \( (1 - \varepsilon) \opt \)-approximation to the optimal value; moreover, we have \( \tau \ll \opt \) so that we can use sampling ideas like the previous section on the residual LP. Towards the goal of showing a good skimming threshold, the following definitions are useful.
Definition 6.1 (Skimmed LP). For $\mathcal{L} \in PLP_n$ and a threshold $\tau \geq 0$, let $\text{skim}_\tau(\mathcal{L})$ be the LP defined by retaining only the columns $t \in [n]$ for which $\pi_t \leq \tau$, and setting $\text{rhs}(\text{skim}_\tau(\mathcal{L})) := \text{rhs}(\mathcal{L}) - \sum_{t : \pi_t > \tau} A^t$. (I.e., this is the residual LP we get from $\mathcal{L}$ if we imagine hard-coding $x_t = 0$ for all $t$ with $\pi_t > \tau$.)

Definition 6.2 (Good Skimming Threshold). A value $\tau \in \mathbb{R}$ is a $\Delta$-good skimming threshold for $\mathcal{L} \in PLP_n$ if:

1. $\text{opt}(\text{skim}_\tau(\mathcal{L})) + \sum_{t : \pi_t > \tau} \pi_t \geq (1 - \epsilon) \text{opt}(\mathcal{L})$,
2. $\text{rhs}(\text{skim}_\tau(\mathcal{L})) \geq \frac{1}{2} \text{rhs}(\mathcal{L})$, and
3. $\tau \leq \frac{160 \cdot \Delta}{B}$.

Lemma 6.3 (Finding Good Skimming Threshold). Given $\mathcal{L} \in PLP_n$ with width $cB$ for $c \geq 1024$, $\epsilon \leq 1/32$ and $\delta > 0$, let $J$ be a random $(n/4)$-subset of $[n]$ and let $I$ be a random $k$-subset of $[n] \setminus J$, for $k \geq n/4$. Define $\tau := \frac{160}{B} \text{opt}(\mathcal{L}^J)$. Then $\tau$ is an $\text{opt}(\mathcal{L})$-good skimming threshold for $\mathcal{L}^I$ with probability $1 - 2\delta$.

6.1.1 Proof of Lemma 6.3

We start with the following crucial but simple claim, which gives a “perfect” skimming threshold $\tau$—one where we do not lose any value if we pick the items of higher value and solve the residual LP. It is a generalization of the “value/weight” rule for solving single knapsack problems, and says that every optimal fractional solution must pick all the sufficiently high value items.

Lemma 6.4 (Perfect Threshold). For a packing LP $\mathcal{L}'$, and for every $\tau \geq \frac{\text{opt}(\mathcal{L}')}{\min_i b_i}$, we have

$$\text{opt}(\mathcal{L}') = \text{opt}(\text{skim}_\tau(\mathcal{L}')) + \sum_{t : \pi_t > \tau} \pi_t.$$ 

Proof. The dual of $\mathcal{L}'$ is given by

$$\mathcal{D}' = \min \left\{ \sum_{i \in [m]} b_i p_i + \sum_{t \in [n]} y_t \mid \langle p, A^t \rangle + y_t \geq \pi_t \quad \forall t \in [n], \quad p, y \geq 0 \right\}$$

Let $(x^*, (p^*, y^*))$ be an optimal primal-dual pair for $\mathcal{L}'$ and $\mathcal{D}'$. We claim that $x^*$ select all items with $\pi_t > \|p^*\|_1$. To see this, observe that $\pi_t > \langle p^*, A^t \rangle$ means $y_t > 0$ and hence $x^*_t = 1$ by complementary slackness. But since the entries of $A^t$ are upper bounded by 1, we have $\langle p^*, A^t \rangle \leq \|p^*\|_1$. Thus, $\pi_t > \|p^*\|_1 \geq \langle p^*, A^t \rangle$ implies $x^*_t = 1$, giving the claim. This means that

$$\text{opt}(\text{skim}_{\|p^*\|_1}(\mathcal{L}')) + \sum_{t : \pi_t > \|p^*\|_1} \pi_t = \text{opt}(\mathcal{L}').$$

Now it suffices to show that $\|p^*\|_1 \leq \frac{\text{opt}(\mathcal{L}')}{\min_i b_i}$. This is easy from strong duality: $\text{opt}(\mathcal{L}') = \text{opt}(\mathcal{D}') \geq \sum_{i \in [m]} b_i p^*_i \geq (\min_i b_i) \cdot \|p^*\|_1$, which completes the proof.

Such a perfect threshold does not give us a handle on the RHS of the skimmed LP—to satisfy property (2) we want the skimmed RHS to be at least half the original RHS. Achieving this will be our next goal.

In addition to the restricted and skimmed LPs from Definitions 4.2 and 6.1, we need one more operator on LPs. Given an LP $\mathcal{L} = \max \{ p \xi : Ax \leq b, \xi \in [0,1]^n \}$ and $K \in \mathbb{Z}$, we use $\mathcal{L}_{<K}$ to denote the LP obtained by zeroing out the value $\pi$ of the $K$ highest valued items (without changing the right hand side).

\footnote{We assume items have distinct values for simplicity; one can break ties consistently and get the same result.}
The next lemma proves that if we take a random sample of a quarter of the columns of $L$, the optimal value on this sample gives us a good estimate for $\text{opt}$ (perhaps minus the value of the top few items, about $O(\varepsilon B)$ of them).

**Lemma 6.5.** For $L \in \mathcal{PLP}_n$ with width at least $8B$ and $\varepsilon \leq 1/32$, $S$ a random $(n/4)$-subset of $[n]$ and $K := 64\varepsilon B$,

$$
\Pr\left[\text{opt}(L^S) \geq \frac{\text{opt}(L_{<K})}{16}\right] \geq 1 - 2\delta.
$$

**Proof.** Let $T$ be the indices of the top $K = 64\varepsilon B$ values of $\pi$ and let $M$ be the value of the $K^{th}$ largest value of $\pi$. We use Lemma 6.4(b) on $L_{<K}$ (using the fact that all item values in $L_{<K}$ are at most $M$) to get with probability at least $1 - \delta$,

$$
\text{opt}(L^S) \geq \text{opt}(L_{<K}^S) \geq (1 - 2\varepsilon) \cdot \frac{1}{4} \text{opt}(L_{<K}) - \varepsilon (8BM)
$$

$$
\leq (1/4 - 2\varepsilon) \text{opt}(L_{<K}) - (KM/8).
$$

(6.14)

Moreover, we claim that $\text{opt}(L^S) \geq (KM)/8$ with probability at least $1 - \varepsilon$. Indeed, consider the solution $\hat{x}$ that sets $\hat{x}_t = 1$ if $t \in S \cap T$ and $\hat{x}_t = 0$ if $t \not\in S \cap T$. We know that $\text{rhs}(L^S) = \frac{1}{4} \text{rhs}(L) \geq 2B1$; moreover, this solution $\hat{x}$ can (at most) pick all $K = 64\varepsilon B \leq 2B$ elements of $T$ (since $\varepsilon \leq 1/32$), so $\hat{x}$ will be feasible. The expected number of non-zeros in $\hat{x}$ is $E[|S \cap T|] = K/4$, and the concentration bound Theorem A.1 implies $\Pr(|S \cap T| < K/8) \leq \exp(-\varepsilon B) \leq \frac{\delta}{m}$. Each item with $\hat{x}_t = 1$ has value at least $M$, so with probability at least $1 - \frac{\delta}{m}$,

$$
\text{opt}(L^S) \geq \sum_{t \in S} \pi_t \hat{x}_t \geq (K/8) \cdot M \geq (KM/8).
$$

(6.15)

Taking the average of inequalities (6.14) and (6.15), and using $\varepsilon \leq 1/16$ concludes the proof. \hfill \square

We now have all the ingredients to prove Lemma 6.3.

**Proof of Lemma 6.3.** Recall that $J$ and $I$ are random (disjoint) subsets of size $n/4$ and $\geq n/4$ respectively, and that we defined the random threshold $\tau = \frac{160}{B} \text{opt}(L^J)$. We want to show $\tau$ is an $\text{opt}(L)$-good skimming threshold for the LP $L^J$. To avoid notational ambiguity, denote $L' := L^J$. Let $T \subseteq I$ be the set of $K := 64\varepsilon B$ indices in $I$ with the highest values of $\pi_t$. Let $H = \{t \in I : \pi_t > \tau\}$ be all items in $I$ with values at least $\tau$.

Property (3) of a good skimming threshold asks for $\tau$ to be at most $\frac{160 \text{opt}(L)}{B}$. Since $\text{opt}(L^J) \leq \text{opt}(L)$, this is true with probability 1.

Let $\mathcal{E}$ be the event that $\text{opt}(L^J) \geq \frac{1}{16} \text{opt}(L_{<K}^J)$. From Lemma 6.3 we know $\text{opt}(L^J) \geq \frac{1}{16} \text{opt}(L_{<K})$ with probability at least $1 - 2\delta$. Since $\text{opt}(L_{<K}) \geq \text{opt}(L'_{<K})$, we get $\Pr[\mathcal{E}] \geq 1 - 2\delta$. We condition on event $\mathcal{E}$ for the rest of the proof.

Now for property (2): the number of items in $H \cap T$ is at most $B/10$, or else picking $B/10 + 1$ items in $H \cap T$ would give a feasible solution for $L'_{<K}$ with value strictly greater than $\tau \cdot (B/10) = 160 \text{opt}(L^J) \geq \text{opt}(L'_{<K})$, a contradiction. Hence $\text{rhs}(\text{skim}_\tau(L'))$ is

$$
\text{rhs}(L') - \sum_{t \in H} A_t \geq \text{rhs}(L') - (|H \cap T| + |T|) \geq \text{rhs}(L') - \left(\frac{B}{10} + K\right)1,
$$

using that $A_t \leq 1$ for all $t$. We know $K = 64\varepsilon B \leq 2B$ (since $\varepsilon \leq 1/32$) so $\frac{1}{10} + \frac{K}{B} \leq 3$. Also, $\text{rhs}(L') = \frac{|H|}{n} \text{rhs}(L) \geq \frac{1}{4} B1$ and $c \geq 24$, so $\text{rhs}(\text{skim}_\tau(L')) \geq \text{rhs}(L') - 3B1 \geq \frac{1}{4} \text{rhs}(L')$. This means we satisfy property (2) of a good skimming threshold for $L' = L^J$ (still conditioned on $\mathcal{E}$).
Finally, to show property (1), we want to prove that 
\[ \text{opt}(\text{skim}_\tau(L')) + \sum_{t \in H} \pi_t \geq (1 - \varepsilon) \text{opt}(L'). \]

Let us define \( L'' := L'_{< K} \); recall this is the LP where we zero out the values of the top \( K \) valued items in \( L' \) (those with indices in \( T \)). We first claim that 
\[ \text{opt}(\text{skim}_\tau(L')) \geq (1 - \varepsilon) \text{opt}(\text{skim}_\tau(L'')) + \sum_{t \in \overline{H} \cap T} \pi_t. \]
Indeed, given an optimal solution \( x^* \) for \( \text{skim}_\tau(L'') \), consider the solution \( \hat{x} \) obtained by setting \( \hat{x}_t = (1 - \varepsilon)x^*_t \) for \( t \in I \cap \overline{H} \cap \overline{T} \) and \( \hat{x}_t = 1 \) for \( t \in I \cap \overline{H} \cap T \). To get that \( \hat{x} \) is feasible for \( \text{skim}_\tau(L') \), note that 
\[ \text{rhs}(\text{skim}_\tau(L'')) - \text{rhs}(\text{skim}_\tau(L')) \leq K \mathbf{1}, \]
and moreover we need an additional \( K \mathbf{1} \) space to accommodate the items in \( \overline{H} \cap T \). On the other hand, scaling \( x^* \) down by \( (1 - \varepsilon) \) reduces its space usage by \( \varepsilon \text{rhs}(\text{skim}_\tau(L'')) \geq \varepsilon \text{rhs}(\text{skim}_\tau(L')) \geq \frac{cB}{8} \mathbf{1} \geq 2K \mathbf{1} \), so the solution \( \hat{x} \) is feasible for \( \text{skim}_\tau(L') \). Moreover, it has value \( (1 - \varepsilon) \text{opt}(\text{skim}_\tau(L'')) + \sum_{t \in \overline{H} \cap T} \pi_t \). Therefore,
\[
\text{opt}(\text{skim}_\tau(L')) \geq (1 - \varepsilon) \text{opt}(\text{skim}_\tau(L'')) + \sum_{t \in \overline{H} \cap T} \pi_t = (1 - \varepsilon) \text{opt}(L'') + \sum_{t \in T} \pi_t \geq (1 - \varepsilon) \text{opt}(L').
\]

The second inequality follows from Lemma 6.4 observe \( \tau = \frac{100 \text{opt}(L)}{B} \geq \frac{10 \text{opt}(L'')}{} \) (by event \( E \)), which is at least \( \frac{\text{opt}(L'')}{} \), the threshold mandated by Lemma 6.4. The final inequality just observes that \( L'' \) is same as \( L' \) with values of items in \( T \) zeroed out. This proves property (1).

So far we were conditioning on the event \( E \). Since \( \Pr[E] \geq 1 - 2\delta \), we get that \( \tau \) is an \( \text{opt}(L) \)-good skimming threshold for \( L^I \) with probability \( 1 - 2\delta \).

### 6.2 Estimating Parameters of a Skimmed LP

Given a \( \Delta \)-good skimming threshold \( \tau \) for \( L \), we can estimate the right sides and values of the skimmed LP \( \text{skim}_\tau(L) \) using sampling: indeed, the RHS and value of \( L^I \) for \( I \) being a random constant fraction of the columns is a very good estimate.

**Lemma 6.6.** Let \( \varepsilon \leq 1/20 \). Given \( L \in \text{PLP}_n \) with width \( cB \) for \( c \geq 64 \), let \( \tau \) be an \( \text{opt}(L) \)-good skimming threshold for \( L \). Let \( K \) be a random \( pn \)-subset of \([n]\) for \( p \in \left[ \frac{1}{4}, \frac{3}{4} \right] \). Then w.p. at least \( 1 - O(\delta) \),

\[
\text{rhs}(\text{skim}_\tau(L^K)) \in p \text{rhs}(\text{skim}_\tau(L)) \pm \varepsilon cB \mathbf{1} \\
\text{opt}(\text{skim}_\tau(L^K)) \in (1 \pm 10\varepsilon) p \text{opt}(\text{skim}_\tau(L)) \pm 10\varepsilon \cdot \text{opt}(L).
\]

**Proof.** Let \( H := \{ t \mid \pi_t > \tau \} \subseteq [n] \) be the set of “high” values. If \( A^t \) is the \( t \)-th column of \( A \), we get that

\[
\begin{align*}
b_L := \text{rhs}(\text{skim}_\tau(L)) &= \text{rhs}(L) - \sum_{t \in H} A^t ; \\
b_L^K := \text{rhs}(\text{skim}_\tau(L^K)) &= \text{rhs}(L^K) - \sum_{t \in K, \pi_t > \tau} A^t = \frac{|K|}{n} \text{rhs}(L) - \sum_{t \in H \cap K} A^t.
\end{align*}
\]

Since \( |K| \) is a random \( pn \)-subset of the columns, the expectation of the second line is precisely \( p \) times the first. Moreover, the vector \( \sum_{t \in H \cap K} A^t \) is coordinate-wise within \( p \sum_{t \in H} A^t \pm \varepsilon cB \mathbf{1} \) with probability at least

\[
1 - 2m \exp \left\{ -\frac{(\varepsilon cB)^2}{4pcB + \varepsilon cB} \right\} \geq 1 - 2m \exp \left\{ -\frac{\varepsilon^2 cB}{5} \right\} \geq 1 - 2m \frac{\delta^2}{m^2} \geq 1 - \frac{2\delta^2}{m}
\]

20
Corollary D.1 to (skim)

Notice that the width of skim rhs

A similar argument shows that − probability 1

ǫ arithmetic using the upper bound on − 1 with probability 1

τ

Now given an LP L := [n] \ H be the low value items. Let skimτ(L)( 1 p · bL K)

be the LP skimτ(L) with the right-hand side replaced by 1 p · bL K and notice that

( skimτ(L)( 1 p · bL K))K = \max \{ \sum_{t \in L \cap K} \pi_t x_t | \sum_{t \in L \cap K} A^t x_t \leq bL K \} = skimτ(LK).

Moreover, from the first part of the lemma, with probability 1 − O(δ) we have 1 p · bL K ∈ bL ± 2εcB p ∈ bL(1 + 8ε), where the last inequality uses p ≥ 1 4, bL ≥ 1 4 rhs(L) (since τ is a good skimming threshold) and rhs(L) ≥ cB1 (from our width lower bound and normalization assumption). Then we can employ Corollary D.1 to (skimτ(L)( 1 p · bL K))K (by setting η = 8ε and ρ = 1 160opt(L) from our width lower bound and normalization assumption). Then we can employ Corollary D.1 to (skimτ(L)( 1 p · bL K))K (by setting η = 8ε and ρ = 1 160opt(L) to get with probability at least 1 − O(δ)

opt(skimτ(LK)) ≥ (1 − 10ε)p opt(skimτ(L)) − 3ε · opt(L)

opt(skimτ(LK)) ≤ (1 + 8ε)p opt(skimτ(L)) + 10ε · opt(L).

(Notice that the width of skimτ(L) is at least min_{i∈I} (bh_L)_i ≥ 1 10 \min_i rhs(L)_i ≥ 32B and p opt(skimτ(L))B ≥ 1 2 + \pi_t for all t ∈ L, so indeed we can apply Corollary D.1.) This proves the lemma.

Now given an LP L we can sample half the columns to find (i) a good skimming threshold for the other half (using Lemma 6.6) and also (ii) estimates of the RHS and optimal value of the other half’s skimmed version (using Lemma 6.6).

Theorem 6.7. Let ε ≤ 1/32. Given L ∈ PLP_n with width cB where c ≥ 1024, let I be a random (n/2)-subset of [n]. We can use the columns in [n] \ I to get estimates τ, ˆO, ˆb which satisfy

(a) τ is an opt(L)-good skimming threshold for L^I,

(b) ˆO ∈ [opt(skimτ(L^I)) − 30ε · opt(L), opt(skimτ(L^I))], and

(c) ˆb ∈ (1 ± 11ε) · rhs(skimτ(L^I))

with probability 1 − O(δ).

Proof. Let J be a random n/4-subset of [n] \ I, and K = [n] \ (I \cup J). Define estimators τ := 160opt(L^J) , ˆO′ := 2opt(skimτ(LK)) and ˆb := 2 rhs(skimτ(LK)). By Lemma 6.3 τ is an opt(L)-good skimming threshold for all three of of L^I, L^K, and L^I\K with probability 1 − O(δ). Notice that conditioned on I \cup K (or equivalently on J, since [n] = I \cup K \cup J), K is a random n/4-subset of I \cup K. Then applying Lemma 6.6 to L^I\K in every instantiation of J where τ is an opt(L)-good skimming threshold for L^I\K, we get that that rhs(skimτ(LK)) ∈ 1 7 rhs(skimτ(LI\K)) ± 2εB1 with probability at least 1 − O(δ). By running the same argument with K replaced by I, we get rhs(skimτ(L^I)) ≥ 2 rhs(skimτ(LI\K)) ± 2εB1 with probability at least 1 − O(δ), and hence

rhs(skimτ(L^I)) ∈ 2 rhs(skimτ(LK)) ± 2εB1

with probability 1−O(δ) by a union bound. Since whenever τ is good for L^K we have rhs(skimτ(LK)) ≥ 1 2 rhs(LK) ≥ 24B1 8 (the last inequality using the width of L and our normalization assumption), simple arithmetic using the upper bound on ε gives the claimed bound of ˆb ∈ (1 ± 11ε) · rhs(skimτ(L^I)) with probability 1 − O(δ).

A similar argument shows that

opt(skimτ(L^I)) ∈ (1 ± 10ε) ˆO′ ± 20ε · opt(L)
with probability $1 - O(\delta)$. Setting $\hat{O} := \frac{\hat{O} - 20\varepsilon \text{opt}(L)}{1 + 10\varepsilon}$ ensures that $\hat{O}$ is now at most $\text{opt}(\text{skim}_\tau(L^1))$, and still at least $\hat{O} \geq \text{opt}(\text{skim}_\tau(L^1)) - 30\varepsilon \cdot \text{opt}(L)$. A union bound over all these results completes the argument. \hfill \Box

### 6.3 One-Time Learning OPT (Part II)

Given a packing LP $L$ in the random permutation model, the first $n/2$ columns in this random order give us a random $n/2$-subset of the columns of $A$; let us denote that by $[n] \setminus \mathbf{I}$. We can now estimate the various parameters using Theorem 6.7 above, and then apply Theorem 3.2 using these estimates to the LP $L^1$ and get almost all its value—which is about half the value of $L$ in expectation. Finally, to improve this to $(1 - O(\varepsilon))$, we use ideas from Section 5.2 to dynamically learn the optimum value over several phases.

**Theorem 6.8.** Let $\varepsilon \in (0, \frac{1}{32}]$, and $\delta \in (0, \varepsilon]$. Consider an LP $L \in \text{PLP}_n$ with width $cB$ for $c \geq 1024$. Suppose that the number of columns $n$, $\varepsilon$ and $\delta$ are known a priori. Then with probability at least $1 - O(\delta)$, we can find an online solution to the random LP $L^{>n/2}$ with value at least

$$\text{opt}(L^{>n/2}) - O(\varepsilon \text{opt}(L)).$$

**Proof.** Let $S$ be a random $\frac{n}{2}$-subset; we make the natural identification between the LPs $L^{>n/2}$ and $L^S$. Let $\tau, \hat{O}, \hat{b}$ be estimates based on $L^{[n]\setminus S}$ that satisfy the properties in Theorem 6.7 for $L^S$ with probability $1 - O(\delta)$.

The algorithm is simple: First, it picks all items in $S$ with values greater than the threshold $\tau$; let $H := \{t \in S : \pi_t > \tau\}$ be these “high” valued elements. Then we essentially run algorithm $\text{LPviaLB}$ from Theorem 3.2 on the residual LP $\text{skim}_\tau(L^S)$. More precisely, define $\varepsilon' := \varepsilon \cdot \max\{11, \sqrt{B\tau/\hat{O}}\}$. If $\varepsilon' \leq \varepsilon_3$ (the required upper bound on Theorem 3.2), we pass to $\text{LPviaLB}$ the LP $\text{skim}_\tau(L^S)$ (in random order), the number of columns $n$, the right-hand side approximation $\hat{b}$, the optimal value approximation $\hat{O}$, the error parameter $\varepsilon'$ and the probability error $\delta$ (given); otherwise, we do not pick any items in $\text{skim}_\tau(L^S)$ (i.e., just keep the items in $S$).

For the rest of the proof, condition on a scenario of $S$ (or equivalently of $[n] \setminus S$) where the guarantees of Theorem 6.7 hold for $\tau$, $\hat{O}$, and $\hat{b}$. Notice that in this case, by construction, whenever we invoke $\text{LPviaLB}$ the requirements of Theorem 3.2 are all satisfied. Then (still conditioned on the set $S$) we have that the following bound holds with probability at least $1 - O(\delta)$ ($\text{val}$ denotes the value of the solution obtained by the above algorithm)

$$\text{val} \geq \sum_{t \in H} \pi_t + \left(1 - \frac{\varepsilon'}{\varepsilon_3}\right) \hat{O} - 60\varepsilon \cdot \text{opt}(L) : \quad (6.16)$$

if $\varepsilon' \leq \varepsilon_3$, then $\text{LPviaLB}$ is executed and this follows from the guarantee from Theorem 3.2 and if $\varepsilon' > \varepsilon_3$ then this is implies by the valid inequality $\text{val} \geq \sum_{t \in H} \pi_t$. Now notice that $\varepsilon' \hat{O} \leq \varepsilon \cdot \max\{11\hat{O}, \sqrt{160 \text{opt}(L)} \hat{O}\} \leq \varepsilon \sqrt{160 \cdot \text{opt}(L)}$ (this uses the fact we are in a scenario of $S$ where the guarantee $\hat{O} \leq \text{opt}(\text{skim}_\tau(L))$ holds). Then using the lower bound $\hat{O} \geq \text{opt}(\text{skim}_\tau(L)) - 30\varepsilon \cdot \text{opt}(L)$, inequality \hfill (6.16) implies that

$$\text{val} \geq \sum_{t \in H} \pi_t + \text{opt}(\text{skim}_\tau(L^S)) - \frac{\varepsilon \sqrt{160}}{\varepsilon_3} \text{opt}(L) - 90\varepsilon \cdot \text{opt}(L) \quad (6.17)$$

with probability $1 - O(\delta)$. From property (1) of good skimming threshold, we then get that $\text{val} \geq (1 - \varepsilon)\text{opt}(L^S) - O(\varepsilon)\text{opt}(L)$ with probability $1 - O(\delta)$. \hfill \Box
Since this holds for every scenario of \( S \) where the guarantees of Theorem 6.7 hold, and since we have \( 1 - O(\delta) \) mass in these scenarios, a union bound concludes the proof.

Observe this guarantee is very similar to that achieved by Theorem 5.1. We can now apply the idea of dynamically learning \( \text{opt} \) over multiple phases exactly as in that case; the details are the same as in Section 5.2 and omitted here. This gives us the following theorem.

**Theorem 6.9.** Let \( \varepsilon \in (0, \frac{1}{32}] \), and \( \delta \in (0, \varepsilon] \). Consider a random packing LP \( L \) with width at least \( 1024 \frac{\log(m/\delta)}{\varepsilon^2} \). Suppose that the number \( n \) of columns and the RHS of \( L \), \( \varepsilon, \delta \) are known a priori. Then with probability at least \( 1 - O(\delta \log \varepsilon^{-1}) \) we can find an online feasible solution to the packing LP \( L \) with value at least \( (1 - O(\delta \log \varepsilon^{-1})) \text{opt}(L) \).

**References**


A Concentration Bounds

Theorem A.1 (Theorem 2.14.19 in [vdVW96]). Let \( Y = \{Y_1, \ldots, Y_n\} \) be a set of real numbers in the interval \([0,1]\). Let \( S \) be a random subset of \( Y \) of size \( s \) and let \( Y_S = \sum_{i \in S} Y_i \). Setting \( \mu = \frac{1}{n} \sum_i Y_i \) and \( \sigma^2 = \frac{1}{n} \sum_i (Y_i - \mu)^2 \), we have that for every \( \tau > 0 \)

\[
\Pr(|Y_S - s\mu| \geq \tau) \leq 2 \exp\left(-\frac{\tau^2}{2s\sigma^2} + \tau\right)
\]

Notice that, since the \( Y_i \)'s belong to the interval \([0,1]\), we can upper bound the variance by the mean as follows:

\[
\sigma^2 \leq \frac{1}{n} \sum_i |Y_i - \mu| \leq \frac{1}{n} \left( \sum_i |Y_i| + \sum_i |\mu| \right) = 2\mu.
\]

Using this observation and a scaling argument gives the following corollary.
Corollary A.2. Let $Y = \{Y_1, \ldots, Y_n\}$ be a set of real numbers in the interval $[0, M]$. Let $S$ be a random subset of $Y$ of size $s$ and let $Y_S = \sum_{i \in S} Y_i$. Setting $\mu = \frac{1}{n}\sum_{i} Y_i$, we have that for every $\tau > 0$,
\[
\Pr(|Y_S - s\mu| \geq \tau) \leq 2\exp\left(-\frac{\tau^2}{M(4s\mu + \tau)}\right)
\]

Theorem A.3 (Freedman’s Inequality \cite{Fre75}). Let $X_0, X_1, \ldots, X_n$ be a real-valued martingale, with $X_0 = 0$, with respect to a filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n$. Let $Y_t = X_t - X_{t-1}$ be the martingale differences, let $M$ be such that $|Y_t|$ $\leq M$ for all $t$ and let $L = \sum_{i \leq n} \text{Var}(Y_t \mid \mathcal{F}_{t-1})$ be the predictable variation. Then for every $\alpha, \sigma^2 \geq 0$,
\[
\Pr\left( X_n \geq \alpha \text{ and } L \leq \sigma^2 \right) \leq \exp\left(-\frac{\alpha^2}{2\sigma^2 + 2M\alpha}\right)
\]

B Maximal Bernstein for Sampling Without Replacement

Lemma B.1. Consider a set of real values $x_1, \ldots, x_n$ in $[0, 1]$, and let $X_1, \ldots, X_k$ be sampled without replacement from this collection. Assume $k \leq n/2$. Let $S_i = X_1 + \ldots + X_i$. Also let $\mu = \frac{1}{n}\sum_{i} x_i$ and $\sigma^2 = \frac{1}{n}\sum_i (x_i - \mu)^2$. Then for every $\alpha > 0$
\[
\Pr\left( \max_{i \leq k} |S_i - i\mu| \geq \alpha \right) \leq 30\exp\left(\frac{(\alpha/24)^2}{2k\sigma^2 + (\alpha/24)}\right).
\]

The lemma follows directly from Theorem A.1 and the following Levy-type maximal inequality for exchangeable random variables of \cite{Pru98}.

Theorem B.2 (Theorem 1 of \cite{Pru98}, with $\gamma = 2$). Let $Y_1, Y_2, \ldots, Y_n$ be an exchangeable sequence of random variables taking values in a Banach space. Then for every $\lambda \geq 0$ and every $k \leq n/2$, we have
\[
\Pr\left( \max_{j \leq k} \left\| \sum_{i \leq j} Y_i \right\| > \lambda \right) \leq 15\Pr\left( \left\| \sum_{i \leq k} Y_i \right\| > \frac{\lambda}{24} \right).
\]

Proof of Lemma B.1. The sequence $\{\sum_{i \leq j} X_i - j\mu\}_{i=1}^n$ is exchangeable, so applying the theorem above and then Theorem A.1 we have
\[
\Pr\left( \max_{j \leq k} \left| \sum_{i \leq j} X_i - j\mu \right| > \alpha \right) \leq 15\Pr\left( \left| \sum_{i \leq k} X_i - k\mu \right| > \frac{\alpha}{24} \right) \leq 30\exp\left(-\frac{(\alpha/24)^2}{2k\sigma^2 + (\alpha/24)}\right).
\]

C Useful Properties of $(\varepsilon, \sigma)$-Stability

For any $\varepsilon'$ and subset $I \subseteq [n]$, the dual of $\mathcal{L}^I(1 - \varepsilon')$ is
\[
\mathcal{D}^I(1 - \varepsilon') := \left\{ \min_{\beta, \gamma} \frac{k}{\gamma} (\alpha, b) - (1 - \varepsilon') \frac{k}{\gamma} (\beta, d) + \sum_{t \in I} \gamma_t \left( \langle A^t_{k, j}, \alpha \rangle - \langle C^t_{k, j}, \beta \rangle + \gamma_t \geq \pi^t_j \right) \quad \forall j, \forall t \in I \right\}
\]
where again we use $A^t_{k, j}$ to denote the $j$th column of $A^t$, and similarly for $C^t$.

The next essentially follows from the fact that $\eta \mapsto \text{opt}(\mathcal{L}(1 - \eta))$ is concave, and that its derivative is obtained by looking at an optimal dual solution.
Lemma C.1. Let $\mathcal{L}$ be an $(\varepsilon, \sigma)$-stable packing/covering LP. Then:

1. For every $\varepsilon' \geq \varepsilon$, $\mathcal{L}$ is $(\varepsilon', \sigma)$-stable.

2. If $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ is an optimal solution for the dual of $\mathcal{L}(1-\varepsilon)$, then $\langle \bar{\beta}, d \rangle \leq \sigma \cdot \text{opt}(\mathcal{L})$.

Proof. Take an optimal solution $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ for $\mathcal{D}(1-\varepsilon)$. Notice that for every $\eta$, $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ is feasible for $\mathcal{D}(1-\varepsilon + \eta)$ and hence $\text{opt}(\mathcal{D}(1-\varepsilon + \eta)) \leq \langle \bar{\alpha}, b \rangle - (1-\varepsilon + \eta) \langle \bar{\beta}, d \rangle + \sum \bar{\gamma} = \text{opt}(\mathcal{D}(1-\varepsilon)) - \eta \langle \bar{\beta}, d \rangle$.

Using Strong Duality, this gives $\text{opt}(\mathcal{L}(1-\varepsilon)) \geq \text{opt}(\mathcal{L}(1-\varepsilon + \eta)) + \eta \langle \bar{\beta}, d \rangle$.

Setting $\eta = \varepsilon$, $(\varepsilon, \sigma)$-stability of $\mathcal{L}$ gives that $\langle \bar{\beta}, d \rangle \leq \sigma \cdot \text{opt}(\mathcal{L})$, proving Part 2. Also, setting $\eta = \varepsilon - \varepsilon'$ we get $\text{opt}(\mathcal{L}(1-\varepsilon')) \leq \text{opt}(\mathcal{L}(1-\varepsilon)) + (\varepsilon' - \varepsilon)\sigma \cdot \text{opt}(\mathcal{L})$; again using $(\varepsilon, \sigma)$-stability of $\mathcal{L}$ we get $\text{opt}(\mathcal{L}(1-\varepsilon')) \leq (1+\varepsilon')\text{opt}(\mathcal{L})$ for $\varepsilon' \geq \varepsilon$, concluding the proof of Part 1.

We can use the concentration of the value of sampled LP’s $\text{opt}(\mathcal{L}_1)$ to show that these sampled LP’s are also stable. (The following lemma uses Lemma C.1 which in turn uses the above Lemma C.1.)

Lemma C.2. For $\varepsilon, \delta \in (0, 1)$, let $\mathcal{L}$ be a packing-covering LP with generalized width at least $\frac{32\log(m/\delta)}{\varepsilon^2}$. Assume that $\mathcal{L}$ is $(\varepsilon, \sigma)$-stable for $\sigma \geq 1$. Let $k \geq 7\varepsilon^2n$ and let $I$ be a random $k$-subset of $[n]$. Then with probability at least $1 - 2\delta$ the LP $\mathcal{L}_I(1 - \varepsilon')$ is $(\varepsilon', 10\sigma)$-stable, where $\varepsilon' = \varepsilon\sqrt{\frac{n}{k}}$.

Proof. We can apply Lemma 4.4(d) to get that with probability at least $1 - \delta$, $\text{opt}(\mathcal{L}_I(1 - 2\varepsilon')) \leq (1 + 6\varepsilon')^k \text{opt}(\mathcal{L})$. Applying Lemma 4.4(b), we get with probability at least $1 - \delta$, $\text{opt}(\mathcal{L}_I(1 - \varepsilon')) \geq (1 - 2\varepsilon')^k \text{opt}(\mathcal{L})$. Together, these inequalities give

$$\text{opt}(\mathcal{L}_I(1 - 2\varepsilon')) \leq \frac{(1 + 6\varepsilon')^k}{(1 - 2\varepsilon')} \text{opt}(\mathcal{L}_I(1 - \varepsilon')) \leq (1 + 10\varepsilon') \text{opt}(\mathcal{L}_I(1 - \varepsilon')),$$

where the last inequality uses the facts that $\varepsilon' \leq 2/5$ and $\sigma \geq 1$. The result then follows.

D OPT Estimation for Packing/Covering LP’s

D.1 Proof of Lemma 4.4

Since the LP has width at least $\frac{32\log(m/\delta)}{\varepsilon^2}$, by scaling the constraints we assume without loss of generality that its right-hand sides are all exactly equal to $\vartheta := \frac{32\log(m/\delta)}{\varepsilon^2}$ and the matrices $A^t, C^t$ on the left-hand side have entries in $[0, 1]$.

Part (a). Consider a feasible solution $\bar{x}$ for $\mathcal{L}$. Since for all $i$, $\sum_i A_i^t \bar{x}^t \leq \vartheta$, from Bernstein’s inequality we have

$$\Pr\left( \sum_{t \in I} A_i^t \bar{x}^t \geq \frac{k}{n} \vartheta + \tau \vartheta \right) \leq \exp \left( - \min \left( \frac{\tau^2 \vartheta^2}{8kn^2}, \frac{\tau \vartheta}{2} \right) \right) \leq \exp \left( - \frac{\log(2m/\delta)}{\varepsilon^2} \min \left( \frac{4\tau^2 n}{k}, 2\tau \right) \right);$$

setting $\tau = (\varepsilon/2)\sqrt{k/n}$ makes the right-hand side at most $\delta/m$. Then taking a union bound over all $i$’s, we have $\sum_{t \in I} A_i^t \bar{x}^t \leq (1 + \varepsilon^2/2)\frac{k}{n} \vartheta 1$ with probability at least $1 - \delta(m_p/2m)$.

Similarly, fix $i$ and let $\mu = \mathbb{E}[\sum_{t \in I} C_i^t \bar{x}^t]$; then using the fact $\mu \geq (k/n) \vartheta$,

$$\Pr\left( \sum_{t \in I} C_i^t \bar{x}^t \leq \mu - \tau \mu \frac{n}{k} \right) \leq \exp \left( - \min \left( \frac{\tau^2 \mu(n/k)^2}{8}, \frac{\tau \mu(n/k)}{2} \right) \right) \leq \exp \left( - \frac{\log(2m/\delta)}{\varepsilon^2} \min \left( \frac{4\tau^2 n}{k}, 2\tau \right) \right).$$

Setting $\tau = (\varepsilon/2)\sqrt{k/n}$ and taking a union bound over all $i$’s, we get $\sum_{t \in I} C_i^t \bar{x}^t \geq (1 - \varepsilon'/2)\frac{k}{n} \vartheta 1$ with probability at least $1 - \delta(m_c/2m)$.

Then, using the fact that $(1 - \varepsilon'/2)(1 + \varepsilon'/2) \leq 1$, with probability at least $1 - \frac{\delta}{2}$, $(1 - \varepsilon'/2)\bar{x}|I$ is $(1 - \varepsilon'/2)^2$-feasible for $\mathcal{L}_I$, and since $(1 - \varepsilon'/2)^2 \geq (1 - \varepsilon')$, Part 1 follows.
**Part (b).** Let $\bar{x}$ be an optimal solution for $\mathcal{L}$. Then by Bernstein’s inequality

$$
\Pr\left(\sum_{t \in I} \pi_t \bar{x}^t \leq \frac{k}{n} \opt(L) - \tau \opt(L)\right) \leq \exp\left(-\min\left\{\frac{\tau^2 \OPT(L)^2 \vartheta}{8 \rho (k/n) \OPT(L)^2}, \frac{\tau \OPT(L) \vartheta}{2 \rho (k/n) \OPT(L)}\right\}\right);
$$

setting $\tau = (\varepsilon \rho / 2) \sqrt{(k/n)}$ makes the right-hand side at most $\delta / 2$.

Then using Part 1, with probability at least $1 - \delta$ we have $\bar{x} | I$ feasible for $\mathcal{L}^i (1 - \varepsilon')$ and with value at least $(1 - \varepsilon' - \rho \varepsilon'/2) \frac{k}{n} \opt(L)$; this lower bounds the value of $\opt(L^i)$ and Part 2 follows.

**Part (c).** Let $\mathcal{E}$ be the event that $\opt(L^i) \geq (1 - \varepsilon' - \rho \varepsilon'/2) \frac{k}{n} \opt(L)$. Since the item values in $\mathcal{L}$ are non-negative, it follows that $\mathbb{E}[\opt(L^i)] \geq \mathbb{E}[\opt(L^i) | \mathcal{E}] \Pr(\mathcal{E})$. Part 3 then follows from Part 2.

Now for the arguments about the upper bounds. Here we use the $(\varepsilon, \sigma)$-stability of $\mathcal{L}$. We consider the (random) dual $\mathcal{D}^i$ of $\mathcal{L}^i$ (recall from (C.18) the expression for the dual).

Let $(\alpha, \beta, \gamma)$ be an optimal solution for $\mathcal{D}(1 - \varepsilon)$. Notice that in every scenario this solution is feasible for $\mathcal{D}^i (1 - \varepsilon')$, so taking expectations $\mathbb{E}[\mathcal{D}^i (1 - \varepsilon')] \leq \frac{k}{n} \opt(L^i) (1 - \varepsilon) + (\varepsilon' - \varepsilon) \alpha \opt(L) + (\sum_{t \in I} \gamma_t - \mathbb{E}[\sum_{t \in I} \gamma_t]) \leq \frac{k}{n} (1 + \varepsilon') \opt(L^i) + (\sum_{t \in I} \gamma_t - \mathbb{E}[\sum_{t \in I} \gamma_t])$. Thus it suffices to show that with probability at least $1 - \delta$, $\sum_{t \in I} \gamma_t \leq \mathbb{E}[\sum_{t \in I} \gamma_t] + \varepsilon' \frac{k}{n} (\rho + \sigma) \opt(L^i)$.

For that, we need to control how large $\gamma_t$ can be. By the optimality of $\gamma_t$ in each scenario notice that for all $t$ we have $\gamma_t \leq \max_j \{\pi^i_t, (\beta, C^i_{s, j})\}$, or else we could reduce $\gamma_t$ and obtain a feasible solution for $\mathcal{D}(1 - \varepsilon)$ with strictly smaller cost. Again from Lemma (C.1) we have $\sum_i \tilde{\beta}_i \leq \sigma \cdot \opt(L) / \vartheta$, and since $C^i \leq 1$, we get $\langle \beta, C^i_{s, j} \rangle \leq \sigma \cdot \opt(L) / \vartheta$. Using our assumption that $\pi^i_t \leq \rho \opt(L) / \vartheta$, we get $\gamma_t \leq (\rho + \sigma) \opt(L) / \vartheta$.

In addition, we claim that $\mathbb{E}[\sum_{t \in I} \gamma_t] \leq \frac{k}{n} (1 + \sigma) \opt(L)$. Since $\opt(L(1 - \varepsilon)) = \langle \alpha, \vartheta \mathbb{1} \rangle - (1 - \varepsilon) \langle \beta, \vartheta \mathbb{1} \rangle + \sum_{t} \gamma_t$, reorganizing we get $\sum_{t} \gamma_t \leq \opt(L(1 - \varepsilon)) + (1 - \varepsilon) \langle \beta, \vartheta \mathbb{1} \rangle$. Using $(\varepsilon, \sigma)$-stability of $\mathcal{L}$ and again the bound from Lemma (C.1) this gives $\sum_{t} \gamma_t \leq (1 + \sigma) \opt(L)$, and the claim follows.

Then using Bernstein’s inequality, we get (letting $\mu = \mathbb{E}[\sum_{t \in I} \gamma_t]$)

$$
\Pr\left(\sum_{t \in I} \gamma_t \geq \mu + \varepsilon \tau \opt(L)\right) \leq \exp\left(-\frac{\log(1/\delta)}{\rho + \sigma} \min\left\{\frac{4 \tau^2}{(k/n)^2}, \frac{2 \tau}{\varepsilon}\right\}\right);
$$

setting $\tau = (\rho + \sigma) \sqrt{(k/n)}$ makes the right-hand side at most $\delta$ (this uses the fact that $\rho \geq 1$ and $k \geq \varepsilon^2 n$). Since $\varepsilon' \frac{k}{n} (\rho + \sigma) \geq \varepsilon \tau$, this concludes the proof of the lemma.

### D.2 Packing LP’s with Perturbed Right-hand Sides

Now we give a corollary of Lemma (D.4) for packing-only LPs that will be convenient for Section 5.

**Corollary D.1.** For $\varepsilon, \delta \in (0, 1)$, consider a packing LP $\mathcal{L}$ with width at least $\frac{32 \log(m / \delta)}{\varepsilon^2}$. Assume that all item values are at most $\rho \cdot \opt(L) \frac{\varepsilon^2}{32 \log(m / \delta)}$ for $\rho \geq 1$. Let $b$ be a random vector such that with probability at least $1 - \delta$, $b \in b(1 \pm \eta)$ for some $\eta \in (0, 1)$, and let $\mathcal{L}(b)$ be the LP $\mathcal{L}$ with right-hand
Let \( k \geq \varepsilon^2 n \) and let \( I \) be a \( k \)-subset of \([n]\). Then for \( \varepsilon' = \varepsilon \sqrt{\frac{n}{k}} \), we have the following guarantees for the randomly permuted LP \( \mathcal{L}(b) \):

(a) With probability at least \( 1 - 2\delta \), \( \text{opt}(\mathcal{L}(b)^I) \geq (1 - \varepsilon' - \rho \varepsilon'/2)(1 - \eta) \cdot \frac{k}{n} \text{opt}(\mathcal{L}) \)

(b) With probability at least \( 1 - 2\delta \), \( \text{opt}(\mathcal{L}(b)^I) \leq (1 + \rho \varepsilon')(1 + \eta) \cdot \frac{k}{n} \text{opt}(\mathcal{L}) \)

Proof. For part (a), notice that whenever \( b \in b(1 \pm \eta) \) we have \( \text{opt}(\mathcal{L}(b)^I) \geq (1 - \eta)\text{opt}(\mathcal{L}^I) \) (since scaling a solution for \( \mathcal{L}^I \) by \((1 - \eta)\)-factor gives a feasible solution for \( \mathcal{L}(b)^I \)). Using Lemma 4.4 and a union bound, we have that with probability at least \( 1 - 2\delta \)

\[
\text{opt}(\mathcal{L}(b)^I) \geq (1 - \eta)\text{opt}(\mathcal{L}^I) \geq (1 - \varepsilon' - \rho \varepsilon'/2)(1 - \eta) \cdot \frac{k}{n} \text{opt}(\mathcal{L}).
\]

(Notice that a packing LP is vacuously \((\varepsilon, 0)\)-stable.)

Part (b) follows similarly by noticing that whenever \( b \in b(1 \pm \eta) \) we have \( \text{opt}(\mathcal{L}(b)^I) \leq (1 + \eta)\text{opt}(\mathcal{L}^I) \) (since scaling a solution for \( \mathcal{L}(b)^I \) by \( \frac{1}{1 + \eta} \)-factor gives a feasible solution for \( \mathcal{L}^I \)).
Online LP

1 Notation

A random LP is an LP of the form

\[
\begin{align*}
\max & \sum_{t=1}^{n} \pi_t x_t \\
\sum_{t=1}^{n} a^t x_t & \leq b \\
x & \in [0,1]^n,
\end{align*}
\]

where the sequence \( (\pi^t, a^t)_{t=1}^{n} \) is sampled without replacement from a collection \( \{(\pi_t, a_t)\}_{t=1}^{n} \). We assume \( a^t \in [0,1] \) throughout.

We use \( RHS(LP) \in \mathbb{R}^m \) to denote the right-hand sides of such LP, and \( OPT(LP) \) as the random variable for the scenario-wise optimal value. ≪Anupam 1.1: Actually \( OPT(LP) \) is a deterministic value.≫

For a set of time steps \( I \subseteq [n] \), let \( LP^I \) denote the LP that only considers columns \( (\pi^t, a^t)_{t \in I} \) and has RHS equal to \( \frac{|I|}{n} RHS(LP) \).

Given a threshold \( \tau \), we define \( LP^I(val \leq \tau) \) as the residual LP of \( LP^I \) after we select the items with value larger than \( \tau \) and \( (0,0) \) otherwise; then \( LP^I(val \leq \tau) \) is

\[
\begin{align*}
\max & \sum_{t \in I} \pi_t x_t \\
\sum_{t \in I} a^t x_t & \leq b - \sum_{t \in I} (a^t - \bar{a}^t) \\
x & \in [0,1]^n.
\end{align*}
\]

Notice that the right-hand side of this LP is random, and the reduction \( a^t - \bar{a}^t \) equals the budget occupation of high-value items that land in \( I \).

2 Solving LPs

2.1 Reduction of LP to Load-balancing

Lemma 1. Let \( LP \) be a random LP. Consider \( \epsilon > 0 \) such that all right-hand sides \( RHS(LP) \) are at least \( \Omega(\frac{\log m}{\epsilon^2}) \). Suppose we are given \( \bar{b} = (1 \pm \epsilon)RHS(LP), \) \( OPT = OPT(LP) \pm \epsilon o \) (for some \( o \geq 0 \)) and \( n \). Then with probability at least \( 1 - \epsilon \), we can obtain an online solution for \( LP \) with value at least

\[
\left(1 - \sqrt{\gamma \log m} - \epsilon\right) OPT(LP) - \epsilon o,
\]

where \( \gamma = \frac{\max_t \pi_t}{OPT(LP)} \).
2.2 Solving Half of an LP

In this section we consider the following problem: \( LP \) is a random LP with known right-hand sides, each satisfying \( \text{RHS}(LP)_i \geq B \triangleq \frac{\log m}{\epsilon^2} \) for some \( \epsilon > 0 \). The number of columns \( n \) in \( LP \) is also known, but \( \text{OPT}(LP) \) is unknown.

Our goal is to obtain an online solution for the second half \( LP^{> n/2} \) with value at least \( (1 - \epsilon)\frac{1}{2}\text{OPT}(LP) \).

**Definition 1** (Good threshold). We say that (a possibly random) \( \tau \) is a good threshold for \( I \subseteq [n] \) if with probability at least \( 1 - \epsilon \):

1. \( LP^{> n/2}(\text{val} \leq \tau) + \text{val}(\{\text{val} > \tau\}) \geq (1 - \epsilon)\text{OPT}(LP^{> n/2}) \)
2. For all \( i \), \( \text{RHS}(LP^{> n/2}(\text{val} \leq \tau))_i \geq \Omega(B) \)
3. \( \tau \leq \frac{10\text{OPT}(LP)}{B} \)

Notice that if we can obtain a good threshold \( \tau \) from the first \( n/2 \) sample columns of the LP, and also obtain a good estimate of \( \text{RHS}(LP^{> n/2}(\text{val} \leq \tau)) \) and \( \text{OPT}(LP^{> n/2}(\text{val} \leq \tau)) \), then the following algorithm returns a \((1 - \epsilon)\)-approximation to \( LP^{> n/2} \): pick the items of value \( > \tau \) that come at times \( > n/2 \), approximate \( LP^{> n/2}(\text{val} \leq \tau) \) using the result from the previous section, and combine the two solutions.

### 2.2.1 Obtaining a Good Threshold

We want a good threshold for \( \{\frac{n}{2} + 1, \ldots, n\} \), but we prove something slightly more general.

**Lemma 2.** Consider \( I \subseteq \{\frac{n}{4} + 1, \ldots, n\} \) of size at least \( n/2 \). Then with probability at least \( 1 - \epsilon \), \( \tau = \frac{10}{B}\text{OPT}(LP^{\leq n/4}) \) is a good threshold for \( I \).

We start with the following, which is a generalization of the “value/weight” rule for solving single knapsack problems.

**Lemma 3.** Consider a deterministic \( LP' \) given by \( \max \{ \pi' x : A' x \leq b', x \in [0,1]^n \} \) with \( A' \) having entries in \([0,1]\). Then for every \( \tau \geq 10\text{OPT}(LP') \min_i b'_i \), we have

\[
LP'(\text{val} \leq \tau) + \sum_{t: \pi'_t > \tau} \pi'_t = \text{OPT}(LP')
\]

**Claim 1.** With probability at least \( 1 - \epsilon \), \( \text{OPT}(LP^{\leq n/4}) \geq \frac{1}{2}\text{OPT}(LP) - \sum_{k=1}^{\epsilon B} \pi(k) \).

**Proof of Lemma.** [TODO]
2.2.2 Estimating RHS and OPT of $LP^{> n/2}(val \leq \tau)$

We are now going to use the columns in time steps $\{\frac{n}{4} + 1, \ldots, n\}$ to provide such estimates. To simplify the notation, let $T = \{\frac{n}{4} + 1, \ldots, n\}$ and $K = \{\frac{n}{4} + 1, \ldots, \frac{n}{2}\}$.

Lemma 4. Let $\tau$ be defined as in Lemma 2, and let $I$ be a subset of $T$ of size at least $n/4$. Then with probability at least $1 - \epsilon$, $\text{RHS}(LP^I(val \leq \tau)) = \frac{|I|}{|T|} \text{RHS}(LP^T(val \leq \tau)) \pm \epsilon B$.

Now we focus on estimates for OPT.

Lemma 5. Let $\tau$ be defined as in Lemma 2, and let $I$ be a subset of $T$ of size at least $n/4$. Then with probability at least $1 - \epsilon$, $\text{OPT}(LP^I(val \leq \tau)) = \frac{|I|}{|T|} \text{OPT}(LP^T(val \leq \tau)) \pm \epsilon \text{OPT}(LP)$.

Corollary 1. Let $\tau$ be defined as in Lemma 2. Then with probability at least $1 - 2\epsilon$ we have

\[
2\text{RHS}(LP^K(val \leq \tau)) = \text{RHS}(LP^{> n/2}(val \leq \tau)) \pm \epsilon B
\]

\[
2\text{OPT}(LP^K(val \leq \tau)) = \text{OPT}(LP^{> n/2}(val \leq \tau)) \pm \epsilon \text{OPT}(LP).
\]

2.2.3 Putting Things Together

Theorem 1. With probability at least $1 - 3\epsilon$ we can find an online feasible solution for $LP^{> n/2}$ with value at least $(1 - \epsilon)^2 \text{OPT}(LP)$.

2.3 Solving a Whole LP

Again consider the following problem: $LP$ is a random LP with known right-hand sides, each satisfying $\text{RHS}(LP)_i \geq B \triangleq \frac{\log m}{\epsilon}$ for some $\epsilon > 0$. The number of columns $n$ in $LP$ is also known, but $\text{OPT}(LP)$ is unknown. Our goal is to find an online solution feasible for this LP with value at least $(1 - \epsilon)\text{OPT}(LP)$.

Lemma 6. With probability at least $1 - \epsilon \log(1/\epsilon)$ we can an online solution feasible for $LP$ with value at least $(1 - \epsilon)\text{OPT}(LP) - \text{OPT}(LP^{\leq \epsilon n})$.

Proof sketch. For simplicity assume that $\log(1/\epsilon)$ is an integer. Let $LP_0 = LP^{\leq \epsilon n}$ and for $j \in \{1, \ldots, \log(1/\epsilon)\}$ let $LP_j = LP^{[2^{j-1} \epsilon n + 1, \ldots, 2^j \epsilon n]}$. Also let $LP_{\leq j} = LP^{[1, \ldots, 2^j \epsilon n]}$.

Let $\epsilon_0 = \sqrt{\epsilon}$ and $\epsilon_j = \sqrt{\epsilon/2^j - 1}$. By definition of $\epsilon$ and the $\epsilon_j$’s we have that all the coordinates of $\text{RHS}(LP_j)$ are at least $\frac{\log m}{\epsilon_j}$.

[TODO]