Stochastic Minimum Norm Combinatorial Optimization

by

Sharat Ibrahimpur

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Sanjeev Khanna
Professor, Department of Computer and Information Science
University of Pennsylvania

Supervisor(s): Chaitanya Swamy
Professor, Department of Combinatorics and Optimization
University of Waterloo

Internal Members: Joseph Cheriyan
Professor, Department of Combinatorics and Optimization
University of Waterloo

Jochen Koenemann
Professor, Department of Combinatorics and Optimization
University of Waterloo

Internal-External Member: Eric Blais
Associate Professor, Cheriton School of Computer Science
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The main results of this thesis are based on the following papers that I have coauthored.

1. [17]: *Approximation Algorithms for Stochastic Minimum-Norm Combinatorial Optimization*. Joint work with Chaitanya Swamy. In Proceedings of the 61st Foundations of Computer Science (FOCS 2020).

2. [19]: *Minimum-Norm Load Balancing Is (Almost) as Easy as Minimizing Makespan*. Joint work with Chaitanya Swamy. In Proceedings of the 48th International Colloquium on Automata, Languages, and Programming (ICALP 2021).

3. [20]: *A Simple Approximation Algorithm for Vector Scheduling and Applications to Stochastic Min-Norm Load Balancing*. Joint work with Chaitanya Swamy. In Proceedings of the 5th Symposium on Simplicity in Algorithms (SOSA 2022).

I understand that my thesis may be made electronically available to the public.

Abstract

Motivated by growing interest in *optimization under uncertainty*, we undertake a systematic study of designing approximation algorithms for a wide class of 1-stage stochastic-optimization problems with norm-based objective functions. We introduce the model of *stochastic minimum norm combinatorial optimization*, denoted StochNormOpt. We have a combinatorial-optimization problem where the costs involved are random variables with given distributions, and we are given a monotone, symmetric norm f. Each feasible solution induces a random multidimensional cost vector whose entries are *independent random variables*, and the goal is to find an *oblivious* solution (i.e., one that does not depend on the realizations of the costs) that minimizes the expected f-norm of the induced cost vector.

We consider two concrete problem settings. In *stochastic load balancing*, jobs with random processing times need to be assigned to machines, and the induced cost vector is the machine-load vector, where the load on a machine is given by the sum of job random variables that are assigned to it. In *stochastic spanning tree*, we have a graph whose edges have stochastic weights, and the induced cost vector consists of edge-weight variables of edges that belong to the spanning tree.

The class of monotone, symmetric norms is broad: it includes frequently-used objectives such as max-cost (ℓ_∞-norm) and sum-of-costs (ℓ_1-norm), and more generally all ℓ_p-norms and Top_ℓ-norms (sum of ℓ largest coordinates in absolute value). Closure properties under taking nonnegative linear combinations and pointwise maximums offer versatility to this class of norms. In particular, the latter closure-property can be used to incorporate *multiple* norm budget constraints $f_\ell(x) \leq B_\ell$, $\ell = 1, \ldots, k$ through a single norm-minimization objective.

Our chief contribution is a framework for designing approximation algorithms for stochastic minimum norm optimization, a significant generalization of the framework of Chakrabarty and Swamy [5] for the deterministic version of StochNormOpt. Our framework has two key components:

(i) A reduction from the minimization of expected f-norm to the simultaneous minimization of a (*small*) collection of expected Top_ℓ-norms; and

(ii) Showing how to tackle the minimization of a single expected Top_ℓ-norm by leveraging techniques used to deal with minimizing the expected maximum, circumventing the difficulties posed by the non-separable nature of Top_ℓ norms.

We apply our framework to obtain approximation algorithms for stochastic min-norm versions of load balancing (StochNormLB) and spanning tree (StochNormTree) problems. We highlight the following approximation guarantees.

- An $O(1)$-approximation for StochNormLB on unrelated machines with: (i) arbitrary monotone symmetric norms and job sizes that are weighted Bernoulli random variables; and (ii) Top_ℓ norms and arbitrary job-size distributions.
- An $O(\log\log m/\log\log\log m)$-approximation for general StochNormLB, where m is the number of machines.
- For identical machines, the above approximation guarantees are in fact simultaneous approximations that hold with respect to *every* monotone, symmetric norm.
- An $O(1)$-approximation for StochNormTree with an arbitrary monotone, symmetric norm and arbitrary edge-weight distributions; this guarantee extends to stochastic minimum-norm matroid basis.

We also consider the special setting of StochNormOpt when the underlying random variables follow Poisson distributions. Our main result here is a novel and powerful reduction showing that, in essence, the stochastic minimum-norm problem can be reduced to a deterministic min-norm version of the same problem. Applying this reduction to (Poisson versions of) spanning tree and load balancing problems yields: (i) an optimal algorithm for StochNormTree; (ii) a $(2+\varepsilon)$-approximation for StochNormLB when the machines are unrelated, and (iii) a PTAS for StochNormLB when the machines are identical. Results (ii) and (iii) utilize approximation algorithms for (deterministic) min-norm load balancing from the work of Ibrahimpur and Swamy [19] in a black-box fashion.

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my advisor Chaitanya Swamy for providing a fulfilling grad school experience over the last seven years. I would not have converged to this wonderful thesis topic without his guidance and patience as I explored a variety of problems. Our weekly meetings provided the constant support that carried me through the highs and lows of research across the years. Swamy's thorough feedback on both my writing and presentations were invaluable in improving my communication skills. I could not have wished for a better advisor. Thank you Swamy.

I thank my thesis committee members, Sanjeev Khanna, Joseph Cheriyan, Jochen Koenemann, and Eric Blais for reading my thesis and providing valuable comments and questions.

I am fortunate to have found great collaborators through my other research projects. Thank you Sylvia Boyd, Arash Haddadan, Ishan Bansal, Logan Grout, Manish Purohit, Erik Vee, Zoya Svitkina, and Joshua R. Wang for all the insightful discussions we have had over the years. Special thanks to Joseph Cheriyan for being so generous with his time and for working with me on several fruitful projects.

I appreciate all the hard work of the staff in the C&O department, particularly Melissa, Carol, and Chloe.

Shout-out to Williams Cafe and Math C&D for creating an ambience that was conducive to obtaining many of the results in this work.

My time in Waterloo was filled with a lot of joy all thanks to the friends that I made here. I will fondly remember our hangouts and academic discussions, and camping trips and conference visits. I will sorely miss our days of shared suffering, be it at the bouldering gym or while taking grad courses. Thank you Hemant, Priya, Cedric, Abhinav, Dhinakaran, Anirudh, Jimit, Royal, Retnika, Harshita, Archana, Nathan, Hari, Justin, Akshay, Madison and Harry for your friendship. Thank you Nishad and Andre for offering advice on various matters.

To Julie, thank you for your love and support. Your silliness filled my life with a lot of laughter, and I will always cherish this.

Last and most importantly, I am deeply indebted to my parents, Lalita and Shivalingappa, and my sisters, Shilpa and Sanjana, for their unconditional love, support and patience through these many years while I was away pursuing my passion.

Dedication

To the curiosity of the human mind.

Table of Contents

List of Figures

Chapter 1

Introduction

Uncertainty is ubiquitous in real-world optimization problems. *Optimization under uncertainty* refers to the branch of mathematical optimization where a decision-maker is faced with taking actions without complete knowledge of the underlying data or the model. In this thesis, we focus on a well-studied optimization-under-uncertainty paradigm called 1-*stage stochastic optimization*, where the randomness follows a known probability distribution and the decision-maker is only allowed to take oblivious actions; in other words, the randomness is *stochastic* and the action is fixed before knowing the actual realizations of the underlying random variables. Each action induces a multidimensional random cost vector, and the decision maker wants to find an "optimal" action that minimizes (or maximizes) the expected objective value of the induced cost vector, for some given objective function. Here, the expectation is over the known underlying distribution.

Stochastic load-balancing and scheduling problems, where jobs have uncertain sizes (a.k.a processing times), constitute a prominent class of stochastic-optimization problems (see, e.g., [33, 23, 9, 36, 22, 10, 12]). In *stochastic load balancing*, we have stochastic jobs that need to be distributed among m machines. An assignment of jobs to machines induces an m-dimensional random load vector on the machines, where the load on a machine is given by the sum of processing-time random variables of the jobs that are assigned to it. The most-popular objective in load balancing literature is minimizing the maximum load-vector entry, often referred to as the *makespan* objective, and this gives rise to the stochastic makespan minimization problem, where we want to find an assignment that minimizes the expectation of the maximum load across all machines. Kleinberg, Rabani and Tardos [23] were the first to devise approximation algorithms for this problem, and they obtained an $O(1)$-approximation in the setting of identical machines. Gupta, Kumar, Nagarajan and Shen [10] later generalized this result and obtained an $O(1)$-approximation in the unrelated-machines setting. More general objective functions, such as minimizing the expected ℓ_p norm of the load vector, have been considered by [10, 34] as a means to capture solutions that trade off the cost incurred by the worst off machine (i.e., adopting an *egalitarian* view) and the total cost incurred across all machines (i.e., adopting a *utilitarian* view); the latter work by Molinaro [34] gives an $O(1)$-approximation for stochastic load balancing with any ℓ_p norm.

As another example in this setup, consider the *stochastic spanning tree* problem, which is the following basic stochastic network-design problem: we have a graph with stochastic edge-weights and we seek a spanning tree of low expected objective value, where the objective is applied to the cost vector that consists of edge-weight random variables that are part of the tree. For the deterministic version of this problem, where edge e has a given weight $w_e \in \mathbb{R}_{\geq 0}$, it is well-known that a minimum-weight spanning tree T simultaneously minimizes the ℓ_p-norm of the induced cost vector $(w_e)_{e \in T}$ for every $p \geq 1$; this is, however, not true for the stochastic setting.

More often than not, in both deterministic and stochastic settings, the choice of an objective function has a significant influence on the optimal solution with respect to that objective function. Moreover, the algorithms designed for different objectives can be fairly different. Since objective functions are usually a means to an end, it is desirable to build algorithmic frameworks that allow one to choose an objective function from a wide class of functions. Motivated by this consideration, Chakrabarty and Swamy [5] (see also [6]) introduced a general model to unify various (deterministic) optimization problems, that they call *minimum-norm optimization*: given an *arbitrary* monotone, symmetric norm f, find a solution that minimizes the f-norm of the induced cost vector. They give an algorithmic framework for minimum-norm optimization, and use it to obtain $O(1)$-approximation algorithms for minimum-norm versions of load balancing and k-clustering problems. Inspired by their success in the deterministic setting, we naturally ask:

> *"Is there a principled approach to design algorithms for stochastic-optimization problems with objective functions given by monotone, symmetric norms?"*

In this thesis, we introduce and study *stochastic minimum norm combinatorial optimization*, the stochastic generalization of minimum norm optimization. Before getting into the details of our model, we briefly discuss the motivation for working at the level of generality afforded by monotone symmetric norms. We need the following definition. Recall that a function $f : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ is a norm if it satisfies: (i) $f(x) = 0$ iff $x = 0$; (ii) $f(x + y) \leq f(x) + f(y)$ for all $x, y \in \mathbb{R}^m$, and (iii) $f(\theta x) = |\theta| f(x)$ for all $x \in \mathbb{R}^m, \theta \in \mathbb{R}$. A monotone norm f satisfies $f(x) \geq f(y)$ for all $x \geq y \geq 0$, and f is said to be symmetric if $f(x)$ is invariant under permutations of the coordinates of x. We briefly highlight three benefits of working with the class of monotone symmetric norms. First, this class is quite rich and broad. In particular, it contains all ℓ_p-norms, as also another fundamental class of norms called Top_ℓ-*norms*: $\mathsf{Top}_\ell(x)$ is the sum of the ℓ largest coordinates of x (in absolute value). Notice that Top_ℓ norms provide another means of interpolating between the min-max (Top_1) and min-sum (Top_m) problems (where m is dimension of the cost vector). Second, as noted in [5], from a theoretical viewpoint, an algorithmic framework for min-norm optimization gives a unified set of techniques for handling various optimization problems under one umbrella. Lastly, this class is closed under taking nonnegative linear combinations and pointwise maximums. The latter closure property can be used to incorporate *multiple* norm budget constraints $\{f_\ell(x) \le B_\ell\}_{\ell=1,\ldots,k}$ via a single norm-minimization objective by considering the norm $f(x) := \max_{\ell\in[k]} f_\ell(x)/B_\ell$. The above benefits also apply in the stochastic setting, making stochastic minimum-norm optimization an appealing model to study.

1.1 Stochastic Minimum Norm Combinatorial Optimization

In an instance of *stochastic minimum norm combinatorial optimization* ($\mathsf{StochNormOpt}$), we have a combinatorial-optimization problem where the costs involved are nonnegative random variables with given distributions. Each feasible solution s induces a random m-dimensional cost vector Y^s whose entries are obtained by aggregating the underlying cost random variables in some fashion; we drop the superscript s if the solution is clear from the context. The objective function is implicitly specified by a monotone, symmetric norm $f : \mathbb{R}^m \to \mathbb{R}_{\ge 0}$. We assume access to a *value oracle* for f, that is, an oracle that given x, returns $f(x)$. The goal in $\mathsf{StochNormOpt}$ is to find a solution s that minimizes $\mathbf{E}[f(Y^s)]$, the expected f-norm of the cost vector induced by s. Here, the expectation is with respect to the distributions of the underlying random variables.

We make a few remarks. The instances of $\mathsf{StochNormOpt}$ that we consider will have a certain degree of independence in the underlying costs, so that the components of the induced cost vector Y^s are always *independent* nonnegative random variables. For brevity, we say that a random vector Y *follows a product distribution on* $\mathbb{R}^m_{\ge 0}$ if $Y_1, \ldots, Y_m$ are independent nonnegative random variables. Second, our assumption that we only have a value oracle for the norm f is weaker than the optimization-oracle and first-order-oracle access required to f in [5] and [6] respectively. Lastly, we assume that the probability distributions of the underlying cost variables are given to us in a succinct form.

We consider two concrete settings of $\mathsf{StochNormOpt}$ arising from load balancing and spanning tree applications.

1.1.1 Stochastic Minimum Norm Load Balancing

In an instance of *stochastic minimum norm load balancing* (denoted $\mathsf{StochNormLB}$), we are given n stochastic jobs that need to be processed on exactly one of m *unrelated machines*. We use J and $[m]$ to denote the set of jobs and machines respectively; we use j to index jobs, and i to index machines. For each job j and machine i, we are given a nonnegative random variable X_{ij} that denotes the processing time of job j on machine i. Jobs are independent, so X_{ij} and $X_{i'j'}$ are independent whenever $j \neq j'$; however, X_{ij} and $X_{i'j}$ could be correlated. Any assignment $\sigma : J \to [m]$ of jobs to machines induces a random load vector $\overrightarrow{\mathsf{load}}^\sigma$ on the machines, where the load on machine i is given by $\overrightarrow{\mathsf{load}}^\sigma(i) := \sum_{j:\sigma(j)=i} X_{ij}$; note that $\overrightarrow{\mathsf{load}}^\sigma$ follows a product distribution on $\mathbb{R}^m_{\geq 0}$. The goal is to find an assignment σ that minimizes $\mathbf{E}[f(\overrightarrow{\mathsf{load}}^\sigma)]$ for a given monotone, symmetric norm $f : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$. Here, the expectation is over the randomness in $\{X_{ij}\}_{i,j}$.

There are three sources of generality in $\mathsf{StochNormLB}$: the generality of monotone symmetric norms, the generality of the unrelated-machines environment, and the generality of job-size distributions. Limiting the level of generality in each of these leads to the following important special cases. We use $\mathsf{StochTop}_\ell\mathsf{LB}$ to refer to $\mathsf{StochNormLB}$ when f is a Top_ℓ norm. In the setting of identical machines, we have $X_{ij} = X_j$ for any job j and machine i. We use $\mathsf{MinNormLB}$ to refer to the deterministic version of $\mathsf{StochNormLB}$ where X_{ij} takes value $p_{ij} \in \mathbb{R}_{\geq 0}$ with probability 1. In $\mathsf{BerNormLB}$, each X_{ij} is a weighted Bernoulli trial that takes size $s_{ij} \in \mathbb{R}_{\geq 0}$ with probability $q_{ij} \in [0, 1]$, and size 0 otherwise; note that $\mathsf{BerNormLB}$ generalizes $\mathsf{MinNormLB}$ and is, in some sense, the simplest non-deterministic version of $\mathsf{StochNormLB}$. Lastly, we use $\mathsf{PoisNormLB}$ to refer to $\mathsf{StochNormLB}$ with Poisson jobs. In this setting, the job-size random variable X_{ij} follows a Poisson distribution with a given mean $\lambda_{ij} \in \mathbb{R}_{\geq 0}$. That is, for any nonnegative integer k, $\mathbf{Pr}[X_{ij} = k] = e^{-\lambda_{ij}} \lambda_{ij}^k / k!$.

1.1.2 Stochastic Minimum Norm Spanning Tree

In an instance of *stochastic minimum norm spanning tree* (denoted $\mathsf{StochNormTree}$), we are given an undirected graph $G = (V, E)$ with stochastic edge-weights $\{X_e\}_{e \in E}$. Edge weights are independent. Any spanning tree $T \subseteq E$ of G induces a random weight vector $Y^T = (X_e)_{e \in T}$; note that Y^T follows a product distribution on $\mathbb{R}^{n-1}_{\geq 0}$ where $n := |V|$. The goal in $\mathsf{StochNormTree}$ is to find a spanning tree T that minimizes $\mathbf{E}[f(Y^T)]$ for a given monotone, symmetric norm $f : \mathbb{R}_{\geq 0}^{n-1} \to \mathbb{R}_{\geq 0}$.

We use $\mathsf{StochTop}_\ell\mathsf{Tree}$ to refer to $\mathsf{StochNormTree}$ when f is a Top_ℓ norm.

Notation and convention. For the rest of this chapter, Y, W denote random vectors that follow a product distribution on $\mathbb{R}_{\geq 0}^m$, and f to denote an arbitrary monotone, symmetric norm. In particular, we reserve W to refer to the cost vector induced by an optimal solution to the given instance of $\mathsf{StochNormOpt}$. For any $x \in \mathbb{R}_{\geq 0}^m$, we use $x^\downarrow$ to denote x with its coordinates sorted in nonincreasing order.

We will adopt the convention that whenever we talk about a stochastic min-norm problem without additional qualifiers, we mean the setting where we have arbitrary distributions and an arbitrary monotone, symmetric norm.

1.1.3 Our Contributions

We now state the main contributions of this thesis along with a brief overview of our techniques. Most of these results have appeared in the papers [17, 19, 20].

Our chief contribution is a framework that we develop for designing algorithms for stochastic minimum-norm combinatorial optimization, using which we devise approximation algorithms for the stochastic minimum-norm versions of load balancing and spanning tree problems.

Our framework has two key components that each address a distinct challenge that arises in dealing with stochastic min-norm optimization. First, how do we reason about the expectation of an *arbitrary* monotone, symmetric norm? In the deterministic setting, the classical majorization inequality for monotone symmetric norms (see [13]) shows that controlling all Top_ℓ norms yields a control on the f-norm. It is not clear whether the above strategy generalizes to the stochastic setting: does controlling all expected Top_ℓ norms of Y yield a control on its expected f-norm? One of our main insights, which forms the foundation of our framework, is to show that this works in an approximate sense.

Theorem 1.1 (See Theorem 4.1)**.** *We have* $f(\mathbf{E}[Y^\downarrow]) \leq \mathbf{E}[f(Y)] \leq O(f(\mathbf{E}[Y^\downarrow]))$.

The above result suggests that if we consider the (deterministic) nonincreasing vector $\mathbf{E}[W^\downarrow] \in \mathbb{R}_{\geq 0}^m$, consisting of the *order statistics* of the optimal solution's cost vector, then some of the ideas from Chakrabarty and Swamy's framework for min-norm optimization can be useful for the stochastic setting. In particular, we obtain an *approximate majorization* inequality, which implies that minimizing the expected f-norm reduces to the simultaneous minimization of a (*small*) collection of expected Top_ℓ-norms. This is the first component of our framework.

Second, how do we deal with the problem of minimizing the expected Top_ℓ norm, even for a single ℓ? Although Top_ℓ norms are conceptually much simpler than arbitrary monotone symmetric norms (see Theorem 2.6), their non-separable nature introduces considerable difficulty in reasoning about $\mathbf{E}[\mathsf{Top}_\ell(Y)]$. To the best of our knowledge, other than the special case of $\ell = 1$, there is no prior work on any stochastic Top_ℓ-norm minimization problem. Here, our approach is based on carefully identifying certain order statistics of the random vector Y that provide a convenient handle on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ (see Sections 3.1 and 3.2); these statistics also play a role in establishing Theorem 1.1. The second component of our framework involves developing mathematical and algorithmic tools for controlling $\mathbf{E}[\mathsf{Top}_\ell(Y)]$; these results can be found in Chapter 3.

The high-level idea of using our framework for specific applications, e.g., StochNormLB and StochNormTree, involves formulating an LP that roughly encodes that the statistics of our random cost vector match the corresponding statistics derived from the cost vector of an optimal solution. Then the remaining technical challenge is to devise a rounding algorithm that rounds the LP solution while losing only a small factor in these statistics. To achieve this, we utilize certain *iterative-rounding* and *randomized-rounding* results from prior work [30], [14].

The above approximation strategy works readily for StochNormTree, and yields the following approximation.

Theorem 1.2 (See Theorems 10.1 and 10.2)**.** *There is a constant-factor approximation algorithm for* StochNormTree*. Furthermore, for* $\mathsf{StochTop}_\ell\mathsf{Tree}$*, we obtain improved approximation guarantees of* $(2+\varepsilon)$ *for general* ℓ*, and* $\big(e/(e-1)+\varepsilon\big)$ *for* $\ell = 1$*, where* $\varepsilon > 0$ *is a constant.*

Theorem 1.2 extends quite seamlessly to the *stochastic minimum norm matroid basis* problem, which is the generalization of StochNormTree with spanning trees replaced by bases of an arbitrary matroid.

There is an added challenge in applying our framework to obtain approximation algorithms for StochNormLB. Our algorithmic tools from Chapters 3 and 4 are based on the assumption that we have distributional information about the components of the random vector Y. This is not true, in general, for stochastic load balancing (unlike StochNormTree) because we do not have direct access to the distribution of the load on a machine: it is implicitly defined as the sum of job-size random variables that are assigned to it; note that we do have direct access to the job-size distributions. In Chapter 5, we discuss a sophisticated notion called *effective size* of a random variable, due to [16] (see also [23]), that will be useful in mitigating some of the difficulties that arise in StochNormLB. Due to the many technical challenges involved in approximating StochNormLB, our results in this setting are the most-sophisticated, and they take up the bulk of this thesis.

Our strongest approximation guarantees are for $\mathsf{StochTop}_\ell\mathsf{LB}$ and BerNormLB, where the latter problem is stochastic load balancing with Bernoulli job-size distributions.

Theorem 1.3 (See Theorems 6.1 and 6.4)**.** *There is a constant-factor approximation algorithm for* $\mathsf{StochTop}_\ell\mathsf{LB}$*.*

The above result considerably generalizes the results of Kleinberg et al. [23] and Gupta et al. [10], which obtained $O(1)$-approximation algorithms for identical and unrelated machines respectively, for the special case of $\ell = 1$, i.e., stochastic makespan minimization.

Theorem 1.4 (See Theorem 8.1)**.** *There is a constant-factor approximation algorithm for* BerNormLB*.*

Since MinNormLB is a trivial case of BerNormLB, modulo constant factors, Theorem 1.4 strictly generalizes the $O(1)$-approximation algorithm (for MinNormLB) in [5, 6].

We have a weaker approximation guarantee for general StochNormLB.

Theorem 1.5 (See Theorem 7.6)**.** *There is an* $O(\log\log m/\log\log\log m)$*-approximation algorithm for* StochNormLB*.*

When the machines are identical, i.e., $X_{ij} = X_j$ for all machines i, the description and analysis of our algorithms in Theorems 1.3, 1.4 and 1.5 become considerably simpler. Moreover, in BerNormLB (and StochNormLB), our algorithm finds an assignment σ that is *simultaneously* $O(1)$-approximate (respectively, $O(\log\log m/\log\log\log m)$-approximate) for *every* monotone, symmetric norm. Our approximation guarantees have an $O(\log\log m/\log\log\log m)$ term because we approximate StochNormLB by relating it to the d-dimensional vector scheduling problem with $d = O(\log m)$, and the current best approximation algorithms for the latter problem have a $O(\log d/\log\log d)$ dependence on the dimension d.

Lastly, for PoisNormLB (i.e., stochastic load balancing with Poisson job-size distributions), we obtain quite strong approximation guarantees via an entirely different approach.

Our main result here is a novel, clean, and versatile *black-box reduction* from PoisNormLB to MinNormLB that loses at most a $(1+\varepsilon)$-factor in approximation, for some $\varepsilon > 0$. We remark that the reduction preserves the machine environment, but the norm in the reduced MinNormLB instance may be different from the norm in the original PoisNormLB instance. We also remark that this reduction is not limited to load-balancing but applies to any stochastic min-norm combinatorial optimization problem, and shows that this can be reduced to a deterministic min-norm version of the *same* combinatorial-optimization problem. Using our results on MinNormLB from [19], we obtain the following concrete approximations for PoisNormLB.

Theorem 1.6 (See Theorems 9.2 and 9.3)**.** *For any $\varepsilon > 0$, there is a randomized $\big(2+O(\varepsilon)\big)$-approximation algorithm for* PoisNormLB*. Furthermore, if the machines are identical, we have a randomized PTAS.*

1.2 Related Work

Stochastic load balancing is a prominent combinatorial-optimization problem that has been investigated in the stochastic setting under various ℓ_p norms. Kleinberg et al. [23] were the first to investigate stochastic makespan minimization: they considered this problem on identical machines and gave an $O(1)$-approximation algorithm. Almost two decades later, Gupta et al. [10] (see [11] for a full version) generalized this result to the unrelated-machines setting, and also gave an $O(p/\log p)$-approximation for ℓ_p norms. The latter guarantee was improved to a constant by Molinaro [34] via a sophisticated tool called the L-function method (see [25]). A common shortcoming of the above works (including our results on StochNormLB) is that the approximation ratios are at least in the hundreds. With an eye towards obtaining small approximation factors, Goel and Indyk [9] considered stochastic makespan minimization (on identical machines) when job-sizes follow a structured distribution. Among other results, they obtained a simple 2-approximation when jobs are Poisson-distributed. Very recently, De et al. [8] improved this approximation guarantee to a PTAS.

Examples of other well-known combinatorial optimization problems that have been investigated in the stochastic setting include stochastic knapsack and bin packing [23, 9, 26, 27], stochastic shortest paths [27]. The works of [27, 29, 28] consider expected-utility-maximization versions of various combinatorial optimization problems. In a sense, this can be viewed as a counterpart of stochastic min-norm optimization, where we have a *concave* utility function, and we seek to maximize the expected utility of the underlying random value vector induced by our solution. Their results are obtained by a clever discretization of the probability space; this does not seem to apply to stochastic minimization problems.

Top_ℓ- and ordered-norms, which are nonnegative linear combinations of Top_ℓ norms, have been proposed in the location-theory literature, as a means of interpolating between the k-center and k-median clustering problems, and have been studied in the Operations Research literature [35, 24], but largely from a modeling perspective. There has been some renewed interest in the algorithms and optimization communities—partly, because Top_ℓ norms yield an alternative way of interpolating between the ℓ_∞- and ℓ_1- objectives—and this work has led to strong algorithmic results for Top_ℓ-norm- and ordered-norm-minimization in deterministic settings [2, 1, 3, 4, 5, 6].

Our approximation algorithms for (general) StochNormLB are obtained by relating it to the d-dimensional vector scheduling problem with $d = O(\log m)$. We mention some related work on this topic in the context of our work. The d-dimensional vector scheduling problem was first considered by Chekuri and Khanna in [7] where they gave an $O(\log^2 d)$-approximation algorithm and showed that the problem is NP-hard to approximate within any constant factor. Meyerson, Roytman and Tagiku [32] gave an improved $O(\log d)$-approximation and the current best approximation factor of $O(\log d/\log\log d)$ is due to Harris and Srinivasan [14] and Im, Kell, Kulkarni and Panigrahi [21]. The result of [14] works even in the unrelated-machines setting where the size-vector of a job j can depend on the machine that it is processed on. Recently, Sai Sandeep [38] gave very strong inapproximability results for vector scheduling indicating that the current best results are almost optimal: under some complexity-theoretic assumptions, they rule out an $O((\log d)^{1-\epsilon})$-approximation for any $\epsilon > 0$.

1.3 Organization

In Chapter 2, we introduce the necessary preliminary concepts that will be used in this thesis.

In Chapters 3 and 4, we develop the mathematical tools that will form the backbone of our framework for StochNormOpt. In Chapter 3, we identify two order statistics of a product distribution Y that provide a convenient handle on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$. In Chapter 4, we prove our main theorem on the expectation of a (monotone, symmetric) norm of a product distribution (Theorem 4.1).

Chapters 5–8 are dedicated for our results on $\mathsf{StochTop}_\ell\mathsf{LB}$, StochNormLB and BerNormLB. Chapter 5 is an introductory chapter where we develop some basic tools that are used in our load-balancing algorithms. Using our framework and the tools developed in Chapter 5, we design an $O(1)$-approximation algorithm for $\mathsf{StochTop}_\ell\mathsf{LB}$ in Chapter 6, an $O(\log\log m/\log\log\log m)$-approximation algorithm for StochNormLB in Chapter 7, and an $O(1)$-approximation algorithm for BerNormLB in Chapter 8.

In Chapter 9, we consider PoisNormLB (i.e., StochNormLB with Poisson jobs). We give a randomized reduction from PoisNormLB to MinNormLB which incurs at most a $(1+\varepsilon)$-factor loss in approximation. This chapter can be read independently of all other chapters.

In Chapter 10, we use our framework to obtain an $O(1)$-approximation algorithm for StochNormTree. Our results for StochNormTree are conceptually much simpler than the corresponding results on StochNormLB. For a reader interested in getting a full picture of all the ingredients that go into designing an algorithm for a concrete setting of StochNormOpt, we suggest reading this chapter right after reading Chapters 3 and 4, and this Chapter can be read independently of Chapters 5 – 9.

We conclude the thesis in Chapter 11 with a brief summary and discussion of future directions.

Chapter 2

Preliminaries

In this chapter, we discuss tools and techniques that are available from prior work.

Notation. The following notation will be frequently used in the rest of the thesis. For an integer $m \geq 1$, we use $[m]$ to denote the set $\{1, 2, \ldots, m\}$. For $z \in \mathbb{R}$, define $z^+ := \max\{z, 0\}$. For an event $\mathcal{E}$ we use the indicator random variable $\mathbb{1}_{\mathcal{E}}$ to denote if the event $\mathcal{E}$ happens. We use $e \approx 2.71828$ to denote the base of the natural logarithm. For a vector $x \in \mathbb{R}^m_{\geq 0}$ we use $x^{\downarrow}$ to denote x with its coordinates sorted in nonincreasing order; i.e., we have $x^{\downarrow}_i = x_{\pi(i)}$, where π is a permutation of $[m]$ such that $x_{\pi(1)} \geq x_{\pi(2)} \geq \cdots \geq x_{\pi(m)}$. We say that x is nonincreasing if $x = x^{\downarrow}$. We use $\mathbf{Pr}[\cdot]$ for the probability of an event, and $\mathbf{E}[\cdot]$ for the expectation of a random variable (possibly vector-valued). For vectors $x, y \in \mathbb{R}^m_{\geq 0}$, we use the relation $x \leq y$ if $x_i \leq y_i$ holds for all $i \in [m]$. For a vector $x \in \mathbb{R}^m$ and $\theta \in \mathbb{R}$, we define $N^{>\theta}(x)$ to be the number of coordinates of x that exceed θ, i.e., $N^{>\theta}(x) := |\{i \in [m] : x_i > \theta\}|$. Similarly, we define $N^{\geq\theta}(x)$ to be the number of coordinates of x that are at least θ. We reserve Y to denote a random vector that follows a product distribution on $\mathbb{R}^m_{\geq 0}$, i.e., $\{Y_i\}_{i\in[m]}$ forms an *independent* collection of nonnegative random variables.

2.1 Monotone Symmetric Norms

We give a primer on the class of monotone symmetric norms and some associated results.

Definition 2.1 (Monotone Symmetric Norms).
A function $f : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ *is said to be a* norm *if it satisfies the following:*

(i) $f(x) = 0$ *if and only if* $x = 0$.

(ii) *(Triangle Inequality)* $f(x + y) \leq f(x) + f(y)$ *for all* $x, y \in \mathbb{R}^m$.

(iii) *(Homogeneity)* $f(\theta x) = |\theta| f(x)$ *for all* $x \in \mathbb{R}^m, \theta \in \mathbb{R}$.

A norm f *is said to be monotone if for every* $x \geq y \geq 0$*, we have* $f(x) \geq f(y)$.
A norm f *is said to be symmetric if* $f(x)$ *is invariant over permutations of the coordinates of* x*; in other words,* $f(x) = f(x^{\downarrow})$.

The cost vectors that arise in our problems are nonnegative, so to keep the notation simple, we restrict the domain of f to the nonnegative orthant $\mathbb{R}_{\geq 0}^{\dim(f)}$ for the rest of the thesis. Without loss of generality, we also assume that the (monotone, symmetric) norms that we consider in this work are *normalized* i.e., $f(1, 0, \ldots, 0) = 1$ holds.

The following result yields simple upper and lower bounds on the norm of a vector. The fact that the multiplicative gap between these bounds is at most m is used often in our algorithms.

Lemma 2.2. *Let* $f : \mathbb{R}_{\geq 0}^m \to \mathbb{R}_{\geq 0}$ *be a normalized, monotone, symmetric norm, and* $x \in \mathbb{R}_{\geq 0}^m$*. We have:*

$$\max_{i \in [m]} x_i \leq f(x) \leq \sum_{i \in [m]} x_i.$$

Proof. Since f is a normalized, monotone, symmetric norm, for any $i \in [m]$ we get: $f(x) \geq f(0, \ldots, 0, x_i, 0, \ldots, 0) = x_i$. The lower bound follows. We use triangle inequality for the upper bound: $f(x) \leq \sum_{i \in [m]} f(0, \ldots, 0, x_i, 0, \ldots, 0) = \sum_{i \in [m]} x_i$. ∎

2.1.1 Top_ℓ Norms and Ordered Norms

The class of monotone symmetric norms is broad and versatile. It contains all ℓ_p norms (defined as $\|x\|_p := \left(\sum_i |x_i|^p\right)^{1/p}$ for every $p \geq 1$), and is closed under taking nonnegative linear combinations and pointwise maximums. For a more elaborate discussion on the versatility of monotone, symmetric norms see [5]. The following definitions and results give a more concrete picture of this class. We start with the basic building blocks.

Definition 2.3 (Top_ℓ Norms)**.** *For any $\ell \in [m]$, the Top_ℓ norm is defined as follows: for $x \in \mathbb{R}^m_{\geq 0}$, $\mathsf{Top}_\ell(x)$ is the sum of the ℓ largest coordinates of x, i.e., $\mathsf{Top}_\ell(x) = \sum_{i=1}^{\ell} x_i^\downarrow$. For notational convenience, we define $\mathsf{Top}_\ell(x)$ to be 0 if $\ell = 0$, and $\mathsf{Top}_m(x)$ if $\ell > m$.*

Observe that, like ℓ_p norms, Top_ℓ norms provide another way of interpolating between the max- (Top_1) and sum- (Top_m) norms. However, there are a few key differences between these two important sub-families of monotone symmetric norms. Since Top_ℓ norms are non-separable, the contribution of an individual coordinate x_i to $\mathsf{Top}_\ell(x)$ is not as apparent. More concretely, associating changes in a single coordinate x_i to changes in $\mathsf{Top}_\ell(x)$ requires some knowledge about the rank of x_i in the ordering given by $x^\downarrow$. On the bright side, the norm-ball $\{x \in \mathbb{R}^m : \mathsf{Top}_\ell(x) \leq 1\}$ of a Top_ℓ norm is polyhedral, so linear programming based techniques are more easily applicable to optimization problems with Top_ℓ objectives.

In Lemma 2.4, we give alternative characterizations of Top_ℓ norms that give us a better sense of the behavior of $\mathsf{Top}_\ell(x)$. Recall that for a vector $x \in \mathbb{R}^m$ and $\theta \in \mathbb{R}$, we define $N^{>\theta}(x) := |\{i \in [m] : x_i > \theta\}|$.

Lemma 2.4. *The following holds for any $x \in \mathbb{R}^m_{\geq 0}$ and any $\ell \in [m]$:*

$$\mathsf{Top}_\ell(x) = \min_{\theta \geq 0}\Big\{\ell\theta + \sum_{i \in [m]} (x_i - \theta)^+\Big\} = \ell x_\ell^\downarrow + \sum_{i \in [m]} (x_i - x_\ell^\downarrow)^+ = \int_0^\infty \min\big(\ell, N^{>\theta}(x)\big)\, d\theta.$$

Lemma 2.4 motivates the order statistic $\tau_\ell(x)$, which tracks the ℓ^{th} largest coordinate of $x \in \mathbb{R}^m_{\geq 0}$.

$$\tau_\ell(x) := x_\ell^\downarrow = \inf\big\{\theta \in \mathbb{R}_{\geq 0} : N^{>\theta}(x) < \ell\big\} \tag{2.1}$$

We further define $\tau_0(x) := \infty$ and $\tau_\ell(x) = 0$ for $\ell > m$. While the expression on the right hand side of (2.1) seems excessively complex for describing $y_\ell^\downarrow$, its utility will be clear in Chapter 3 when we work with random vectors Y and need a notion of ℓ^{th} largest coordinate of Y.

We now define ordered norms, which are simply nonnegative linear combinations of Top_ℓ norms.

Definition 2.5 (w-Ordered Norms)**.** *Let $w \in \mathbb{R}^m_{\geq 0}$ be a nonincreasing vector, and let $w_{m+1} := 0$. The w-ordered norm is defined as follows: for $x \in \mathbb{R}^m_{\geq 0}$,*

$$\|x\|_w := w^T x^\downarrow = \sum_{\ell \in [m]} w_\ell x_\ell^\downarrow = \sum_{\ell \in [m]} (w_\ell - w_{\ell+1})\, \mathsf{Top}_\ell(x).$$

The following structural result from Chakrabarty and Swamy [5] gives a characterization of monotone symmetric norms in terms of ordered norms. This viewpoint is quite helpful in understanding the complexity of controlling the expected f-norm of a random vector.

Theorem 2.6. *For any monotone, symmetric norm $f : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$, there exists a (possibly uncountable) collection $\mathcal{C} \subseteq \mathbb{R}^m_{\geq 0}$ of nonincreasing vectors such that for any $x \in \mathbb{R}^m_{\geq 0}$, we have $f(x) = \max_{w \in \mathcal{C}} w^T x^{\downarrow}$.*

2.1.2 Majorization Inequality for Monotone Symmetric Norms

The following result is a specialization of the classical majorization inequality for *Schur convex* functions (see [31] for relevant definitions). The original result is due to Hardy, Littlewood and Pólya [13], and can also be derived from Theorem 2.6.

Theorem 2.7. *Let $x, y \in \mathbb{R}^m_{\geq 0}$ be such that $\mathsf{Top}_\ell(x) \leq \mathsf{Top}_\ell(y)$ holds for all $\ell \in [m]$. Then $f(x) \leq f(y)$ for any monotone, symmetric norm $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$.*

Theorem 2.7 forms the backbone of Chakrabarty and Swamy's framework in [5] for (deterministic) minimum-norm optimization. The following result is a corollary of Theorem 2.7, and gives a convenient way to compare norms of two vectors.

Lemma 2.8. *Let $x, y \in \mathbb{R}^m_{\geq 0}$ and $\alpha, \beta \in \mathbb{R}_{\geq 0}$. Suppose that $\mathsf{Top}_\ell(x) \leq \alpha \mathsf{Top}_\ell(y) + \beta$ holds for all $\ell \in [m]$. Then, $f(x) \leq \alpha f(y) + \beta$ for any (normalized) monotone, symmetric norm $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$.*

Proof. Fix a monotone, symmetric norm f. By symmetry of f, we may assume that $x = x^{\downarrow}$ and $y = y^{\downarrow}$. Consider the vector $y' := \alpha y + \beta \mathbf{e}_1$, where $\mathbf{e}_1 \in \{0, 1\}^m$ has a single 1 at the first coordinate. By our assumption $\mathsf{Top}_\ell(x) \leq \mathsf{Top}_\ell(y')$ for all $\ell \in [m]$. By Theorem 2.7, we have

$$f(x) \leq f(y') = f(\alpha y + \beta \mathbf{e}_1) \leq f(\alpha y) + f(\beta \mathbf{e}_1) = \alpha f(y) + \beta,$$

where we use triangle inequality and homogeneity of norms. ∎

Our algorithms will "guess" (i.e., enumerate) the Top_ℓ norms (or certain associated quantities) of the cost vector induced by an optimal solution, and will aim to obtain a solution whose induced cost vector has comparable Top_ℓ norms. However, to make this approach polynomial time, we will only be able to enumerate the Top_ℓ norms for a certain *sparse* subset of indices in $[m]$. The next few definitions and results make this precise, and show that the move to this sparse subset only leads to a small loss in approximation. To this end, we define $\mathsf{POS}_m := \{1, 2, 4, \dots, 2^{\lfloor \log_2 m \rfloor}\}$ to be the set of powers of 2 that are at most m; we drop the subscript in POS_m when it is clear from the context.

Lemma 2.9. *Let $x, y \in \mathbb{R}^m_{\geq 0}$ be such that $\mathsf{Top}_\ell(x) \leq \mathsf{Top}_\ell(y)$ for all $\ell \in \mathsf{POS}$. Then we have $f(x) \leq 2f(y)$ for every monotone, symmetric norm $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$.*

Proof. Fix some $\ell \in [m]$, and define $\ell' := 2^{\lfloor \log_2 \ell \rfloor}$. Note $\ell \in \mathsf{POS}$ and $\ell' \leq \ell \leq 2\ell'$. By definition of Top_ℓ norms, we have $\mathsf{Top}_\ell(x) \leq 2\,\mathsf{Top}_{\ell'}(x) \leq 2\,\mathsf{Top}_{\ell'}(y) \leq 2\,\mathsf{Top}_\ell(y)$. The desired conclusion follows from Theorem 2.8 by taking $\alpha = 2$ and $\beta = 0$. ∎

We generalize Lemma 2.9 by considering a generalization of POS that, roughly speaking, consists of powers of $1+\delta$ for some $\delta > 0$. Fix a parameter $\delta > 0$. We define $\mathsf{POS}_{m,\delta} \subseteq [m]$ iteratively as follows: include the index 1 in $\mathsf{POS}_{m,\delta}$; as long as the largest index $\ell \in \mathsf{POS}_{m,\delta}$ is such that $\lceil (1+\delta)\ell \rceil \leq m$, include $\lceil (1+\delta)\ell \rceil$ (which is larger than ℓ) in $\mathsf{POS}_{m,\delta}$. We remark that this definition is mathematically slightly more convenient to work with than the one in [5], where $\mathsf{POS}_{m,\delta}$ is defined as $\big\{\min\{\lceil (1+\delta)^s \rceil, m\} : s \in \mathbb{Z}_{\geq 0}\big\}$, but this change is not crucial. Also note that $\mathsf{POS}_m = \mathsf{POS}_{m,1}$; we drop the subscript δ in $\mathsf{POS}_{m,\delta}$ if $\delta = 1$.

Claim 2.10. *We have $|\mathsf{POS}_{m,\delta}| \leq 1 + \log_{1+\delta} m = O\big((\log m)/\delta\big)$.*

Fix some m and $\delta > 0$. For $i \in [m]$, let $\mathsf{next}(i)$ be the smallest index in POS (strictly) larger than i; if no such index exists, then we define $\mathsf{next}(i) := m + 1$ for notational convenience. Similarly, let $\mathsf{prev}(i)$ be the largest index in POS (strictly) smaller than i; set $\mathsf{prev}(1) := 0$. By definition, we trivially have $\mathsf{next}(\ell) - 1 \leq (1+\delta)\ell$ for any $\ell \in \mathsf{POS}_{m,\delta}$. The following claim is also immediate.

Claim 2.11. *We have $\mathsf{next}(i) - 1 \leq (1+\delta)i$ for all $i \in [m]$.*

Lemma 2.12 shows that focusing on only the indices in $\mathsf{POS}_{m,\delta}$ results in at most a $(1+\delta)$-factor loss.

Lemma 2.12. *Let $x, y \in \mathbb{R}^m_{\geq 0}$ be such that $\mathsf{Top}_\ell(x) \leq \mathsf{Top}_\ell(y)$ for all $\ell \in \mathsf{POS}_{m,\delta}$. Then we have $f(x) \leq (1+\delta)f(y)$ for every monotone, symmetric norm $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$.*

Proof. Fix some index $i \in [m] \setminus \mathsf{POS}_{m,\delta}$, and let $\ell := \mathsf{prev}(i) \in \mathsf{POS}_{m,\delta}$. By Claim 2.11, we have $i \leq (1+\delta)\ell$. Therefore $\mathsf{Top}_i(x) \leq (1+\delta)\mathsf{Top}_\ell(x) \leq (1+\delta)\mathsf{Top}_\ell(y) \leq (1+\delta)\mathsf{Top}_i(y)$. The desired conclusion follows from Lemma 2.8 by taking $\alpha = 1+\delta$ and $\beta = 0$. ∎

2.2 Tools to Control Monotone Symmetric Norms

For the rest of this chapter, we use the term norm to mean a monotone, symmetric norm. In Section 2.1.2 we saw some basic results on comparing norms of vectors via controlling their Top_ℓ norms. We will, however, need more versatile versions of Lemma 2.12 to compare norms of vectors that are produced by our algorithms.

2.2.1 τ_ℓ-Based Majorization Inequality

In this section, we show that if the τ_ℓ statistic of x is comparable to the $\tau_{\ell/\alpha}$ statistic of y, for some $\alpha \geq 1$, then $f(x)$ is comparable to $\alpha f(y)$. The following results will be useful in Chapters 4 and 10.

Lemma 2.13. *Let $x, y \in \mathbb{R}^m_{\geq 0}$, $\alpha \in \mathbb{R}_{\geq 1}$ be a scalar, and $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotone symmetric norm. Suppose that $x^\downarrow_\ell \leq y^\downarrow_{\lceil \ell/\alpha \rceil}$ for all $\ell \in [m]$. Then, $f(x) \leq \alpha f(y)$.*

Proof. Without loss of generality, suppose $x = x^\downarrow$ and $y = y^\downarrow$. We start with some intuition for the proof by considering the easy case when α is an integer. First, suppose $\alpha = 1$. Clearly, $\mathsf{Top}_\ell(x) \leq \mathsf{Top}_\ell(y)$ for all ℓ, so $f(x) \leq f(y)$ follows. Next, suppose $\alpha > 1$. Now consider the vector $y' \in \mathbb{R}^{\alpha m}_{\geq 0}$ obtained by concatenating α copies of y, and let $y'' \in \mathbb{R}^m_{\geq 0}$ be the vector consisting of the m largest coordinates of y'. Since $\tau_\ell(x) \leq \tau_\ell(y'')$, the result for the $\alpha = 1$ case implies $f(x) \leq f(y'')$. It is not hard to see that triangle inequality implies $f(y'') \leq \alpha f(y)$, which yields the claimed bound. We now extend the above argument for any scalar $\alpha \geq 1$.

Since f is monotone, we may assume that $x_\ell = y_{\lceil \ell/\alpha \rceil}$ for all $\ell \in [m]$. We prove the lemma by using Theorem 2.8: we show that for any $\ell \in [m]$, $\mathsf{Top}_\ell(x) \leq \alpha\, \mathsf{Top}_\ell(y)$ holds. In fact, we show something stronger: $\mathsf{Top}_\ell(x) \leq \alpha\, \mathsf{Top}_{\lceil \ell/\alpha \rceil}(y)$ for any $\ell \in [m]$.

Fix some $k \in \{1, \ldots, \lceil m/\alpha \rceil\}$ and define $\ell := \min(\lfloor \alpha k \rfloor, m)$. It suffices to focus only on this choice of ℓ because $\lceil \ell'/\alpha \rceil = k$ for all $\ell' \in \{\lfloor \alpha(k-1) \rfloor + 1, \ldots, \min(\lfloor \alpha k \rfloor, m)\}$. We have:

$$
\begin{aligned}
\mathsf{Top}_\ell(x) &= \left\{ \sum_{k'=1}^{k-1} (\lfloor \alpha k' \rfloor - \lfloor \alpha(k'-1) \rfloor) \cdot y_{k'} \right\} + (\min(\lfloor \alpha k \rfloor, m) - \lfloor \alpha(k-1) \rfloor) y_k \\
&\leq \sum_{k'=1}^{k} (\lfloor \alpha k' \rfloor - \lfloor \alpha(k'-1) \rfloor) y_{k'}
\end{aligned}
$$

$$= \sum_{k'=1}^{k} \left\{ \alpha + \big(\alpha(k'-1) - \lfloor \alpha(k'-1) \rfloor\big) - \big(\alpha k' - \lfloor \alpha k' \rfloor\big) \right\} \cdot y_{k'}$$

$$= \alpha \mathsf{Top}_k(y) + \left\{ \sum_{k'=1}^{k-1} (\alpha k' - \lfloor \alpha k' \rfloor) \cdot (y_{k'+1} - y_{k'}) \right\} - (\alpha k - \lfloor \alpha k \rfloor) y_k$$

$$\leq \alpha \mathsf{Top}_k(y).$$ ■

The following lemma is the contrapositive of Lemma 2.13.

Lemma 2.14. *Let $x, y \in \mathbb{R}^m_{\geq 0}$, $\alpha \in \mathbb{R}_{\geq 1}$ be a scalar, and $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotone symmetric norm. If $f(x) > \alpha f(y)$, then there exists an index $k \in \{1, 2, \ldots, \lceil m/\alpha \rceil\}$ such that $x^{\downarrow}_{\lfloor \alpha(k-1) \rfloor + 1} > y^{\downarrow}_k$ holds.*

Proof. If $f(x) > \alpha f(y)$, then by Lemma 2.13 we have $x^{\downarrow}_{\ell} > y^{\downarrow}_{\lceil \ell/\alpha \rceil}$ for some $\ell \in [m]$. Let $k := \lceil \ell/\alpha \rceil \in \{1, 2, \ldots, \lceil m/\alpha \rceil\}$. Observe that $\ell \in \{\lfloor \alpha(k-1) \rfloor + 1, \ldots, \lfloor \alpha k \rfloor$. Since $x^{\downarrow}$ is nonincreasing, we have $x^{\downarrow}_{\lfloor \alpha(k-1) \rfloor + 1} > y^{\downarrow}_k$. ■

2.2.2 Budgeted Version of Majorization Inequality

Suppose that we are given some (upper or lower) bounds B_ℓ on $\mathsf{Top}_\ell(x)$ for every $\ell \in \mathsf{POS}$. Our main result in this section, Theorem 2.15, is a budgeted version of the majorization inequality: we prove that $f(x)$ can be bounded in terms of the f-norm of a vector that is induced by the B_ℓ's. At a high level, our proof of Theorem 2.15 is based on a back-calculation of the vector y whose Top_ℓ norms are comparable to the budgets B_ℓ, so that Theorem 2.8 can yield bounds on $f(x)$ in terms of $f(y)$. We state Theorem 2.15 below, but defer its proof to the end of the section. The definition of upper envelope curve will be given shortly.

Theorem 2.15. *Let $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotone symmetric norm, and $y \in \mathbb{R}^m_{\geq 0}$ be a nonincreasing vector. Let $(B_\ell)_{\ell \in \mathsf{POS}_m}$ be a nondecreasing, nonnegative sequence, and $b : [0, m] \to \mathbb{R}_{\geq 0}$ denote its upper envelope curve. Define $\vec{b} := (b(i) - b(i-1))_{i \in [m]} \in \mathbb{R}^m_{\geq 0}$.*

(a) *If $\mathsf{Top}_\ell(x) \geq B_\ell$ for all $\ell \in \mathsf{POS} = \{1, 2, \ldots, 2^{\lfloor \log_2 m \rfloor}\}$, then $f(x) \geq f(\vec{b})$.*

(b) *If $\mathsf{Top}_\ell(x) \leq B_\ell$ for all $\ell \in \mathsf{POS}$, then $f(x) \leq 2f(\vec{b})$.*

Definition 2.16 (Upper Envelope Curve)**.** *Let $\mathcal{I} \subseteq [m]$ be a subset of indices and $(B_\ell)_{\ell\in\mathcal{I}}$ be a nondecreasing (over $\mathcal{I}$), nonnegative sequence. The* upper envelope curve $b : [0, m] \to \mathbb{R}_{\geq 0}$ *of $(B_\ell)_{\ell\in\mathcal{I}}$ is defined as follows: for any $k \in [0, m]$,*

$$b(k) := \max \left\{\theta \in \mathbb{R}_{\geq 0} : (k, \theta) \in \text{conv}(S)\right\},$$

where $S = \{(\ell, B_\ell) : \ell \in \mathcal{I}\} \cup \{(0, 0), (m, \max_{\ell\in\mathcal{I}} B_\ell)\}$ and $\text{conv}(S)$ denotes the convex hull of S.

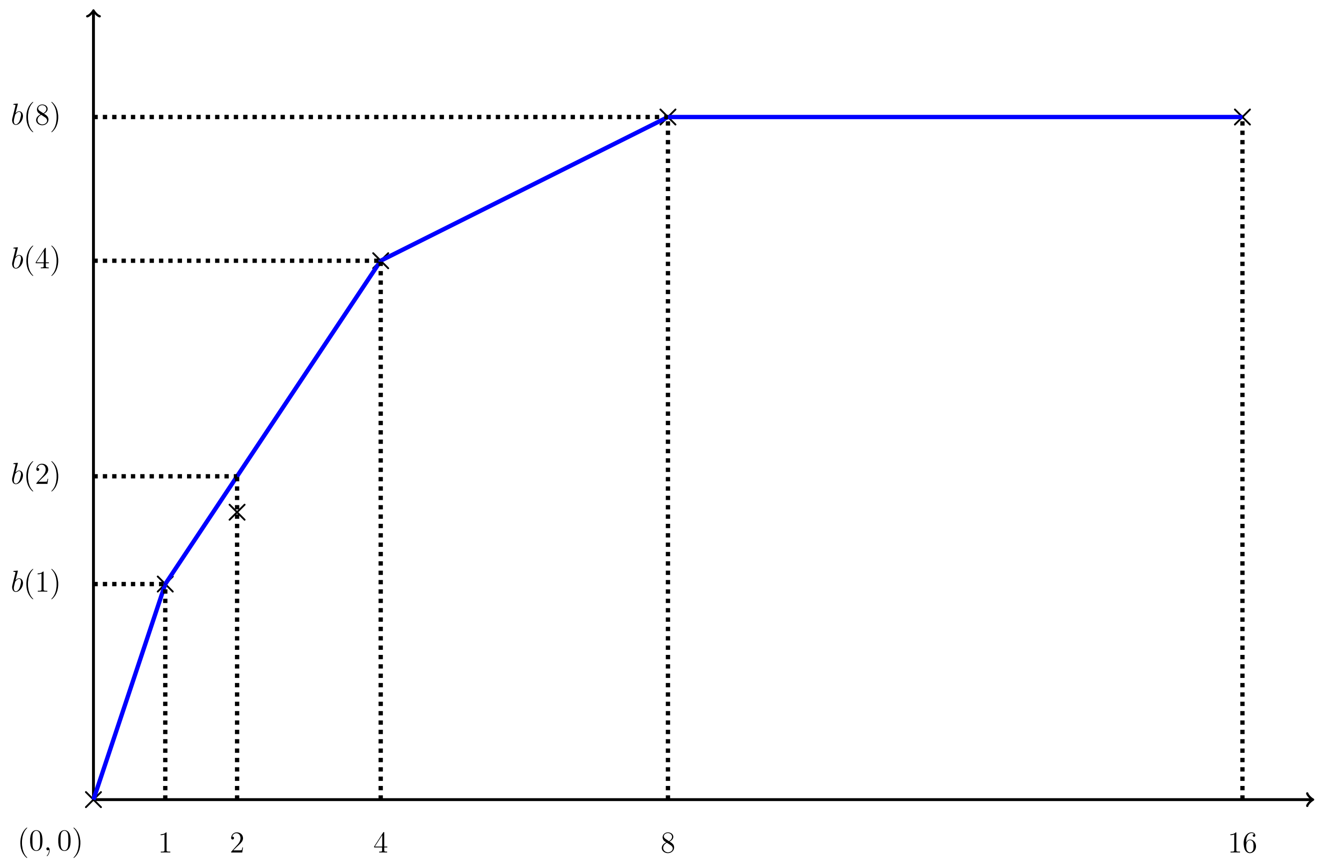

Figure 2.1: The upper envelope curve for $(B_1 = 3, B_2 = 4, B_4 = \frac{15}{2}, B_8 = \frac{19}{2})$ is shown in blue. We have $b(\ell) = B_\ell$ for $\ell = 1, 4, 8,$ and $b(2) = \frac{9}{2} > B_2$. The induced $\vec{b}$ is $\left(3, \frac{3}{2}, \frac{3}{2}, \frac{3}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, \ldots, 0\right)$

Lemma 2.17. *Let $\mathcal{I} \subseteq [m]$ be a subset of indices, $(B_\ell)_{\ell\in\mathcal{I}}$ be a nondecreasing, nonnegative sequence, and b denote the corresponding upper envelope curve. The following are true:*

(a) *The function b is nondecreasing over $[0,m]$ and satisfies $b(\ell) \geq B_\ell$ for all $\ell \in \mathcal{I}$.*

(b) *The function b is concave over $[0,m]$.*

(c) *Define $\vec{b} \in \mathbb{R}^m$ as follows: $\vec{b}_i := b(i) - b(i-1)$ for any $i \in [m]$. The vector $\vec{b}$ is nonincreasing and nonnegative. So, for any $\ell \in [m]$ we have $\mathsf{Top}_\ell(\vec{b}) = b(\ell)$.*

Proof. We first show that (c) follows from (a) and (b). The vector $\vec{b}$ is nonnegative because b is nondecreasing. Next, $\vec{b}$ is nonincreasing because for any $i \in [m-1]$, $b(i) - b(i-1) \geq b(i+1) - b(i)$ if and only if $b(\frac{1}{2} \cdot (i-1) + \frac{1}{2} \cdot (i+1)) \geq \frac{1}{2} \cdot b(i-1) + \frac{1}{2} \cdot b(i+1)$. The latter statement follows from concavity of b.

For notational convenience, let $B_0 := 0$ and $B_m := \max_{\ell \in \mathcal{I}} B_\ell$. Since S (see Definition 2.16) has finitely many points in $\mathbb{R}^2_{\geq 0}$, $\mathrm{conv}(S)$ is a polygon in the nonnegative quadrant. By definition of the upper envelope curve, the function b is piecewise linear. Suppose that the upper envelope curve is composed of n lines $L_1, \ldots, L_n$, where for $j \in [n]$ the line L_j connects extreme points $(\ell_{j-1}, b(\ell_{j-1}))$ and $(\ell_j, b(\ell_j))$. Assume that $0 = \ell_0 < \ell_1 < \cdots < \ell_n = m$ holds. Also, observe that we have $b(\ell_j) = B_{\ell_j}$ for any $j \in \{0, 1, \ldots, n\}$.

We prove (a) and (b) using the above observations. The function b is nondecreasing over $[0,m]$ because the sequence $(B_{\ell_j})_{0 \leq j \leq n}$ is nondecreasing, and it is concave because it is the upper envelope (a.k.a. upper hull) of a convex body. Lastly, for all $\ell \in \mathcal{I}$ we have $b(\ell) \geq B_\ell$ because $(\ell, B_\ell) \in \mathrm{conv}(S)$. ∎

Lemma 2.18. *Let $y \in \mathbb{R}^m_{\geq 0}$ be a nonincreasing vector. Let $b : [0,m] \to \mathbb{R}_{\geq 0}$ denote the upper envelope curve for $(\mathsf{Top}_\ell(y))_{\ell \in [m]}$. Then, $b(\ell) = \mathsf{Top}_\ell(y)$ for all $\ell \in [m]$.*

Proof. Suppose for the sake of contradiction that there exists $\ell \in [m]$ such that $b(\ell) > \mathsf{Top}_\ell(y)$. Consider the smallest such ℓ, so that $b(\ell') = \mathsf{Top}_{\ell'}(y)$ for all $1 \leq \ell' < \ell$. By definition of the upper envelope curve, there exist $\ell_1, \ell_2 \in [m]$ and $\alpha \in (0,1)$ satisfying: (i) $\ell_1 < \ell_2$; (ii) $\ell = \alpha \ell_1 + (1-\alpha)\ell_2$; and (iii) $b(\ell) = \alpha \mathsf{Top}_{\ell_1}(y) + (1-\alpha)\mathsf{Top}_{\ell_2}(y)$. In other words, the point $(\ell, b(\ell))$ lies on the line L joining $(\ell_1, \mathsf{Top}_{\ell_1}(y))$ and $(\ell_2, \mathsf{Top}_{\ell_2}(y))$. By using convexity of $\mathrm{conv}(S)$, we may take $\ell_1 = \ell - 1$. This is because the point $(\ell_1 + 1, b(\ell_1 + 1))$, which equals $(\ell_1 + 1, \mathsf{Top}_{\ell_1+1}(y))$ by our assumption on ℓ, is either on or above the line L.

In either case it can be used to obtain a convex combination for $(\ell, b(\ell))$. Observe that:

$$\begin{aligned}
\mathsf{Top}_\ell(y) < b(\ell) &= \alpha \mathsf{Top}_{\ell-1}(y) + (1-\alpha)\mathsf{Top}_{\ell_2}(y) \\
&= \mathsf{Top}_{\ell-1}(y) + (1-\alpha) \cdot \sum_{i=\ell}^{\ell_2} y_i \\
&\leq \mathsf{Top}_{\ell-1}(y) + (1-\alpha) \cdot (\ell_2 - \ell + 1) \cdot y_\ell && (y \text{ is nonincreasing}) \\
&\leq \mathsf{Top}_{\ell-1}(y) + (1-\alpha) \cdot \left(\frac{\alpha}{1-\alpha} + 1\right) \cdot y_\ell && (\text{as } \ell = \alpha(\ell-1) + (1-\alpha)\ell_2) \\
&= \mathsf{Top}_\ell(y),
\end{aligned}$$

which leads to a contradiction. Hence, the statement is true. ■

Lemma 2.19. *Let $\mathcal{I} \subseteq \mathcal{I}' \subseteq [m]$ be two index sets. Let $(B_\ell)_{\ell \in \mathcal{I}}$ and $(B'_\ell)_{\ell \in \mathcal{I}'}$ be two nondecreasing, nonnegative sequences, and let b and b' denote their respective upper envelope curves. If for every $\ell \in \mathcal{I}$ we have $B_\ell \leq B'_\ell$, then for any $k \in [0, m]$ we have $b(k) \leq b'(k)$.*

Proof. Follows from the definition. ■

We now prove Theorem 2.15 by combining the above lemmas.

Proof of Theorem 2.15. For the first part, let b' denote the upper envelope curve for the sequence $(\mathsf{Top}_\ell(x))_{\ell \in [m]}$. By Lemma 2.18, $b'(\ell) = \mathsf{Top}_\ell(x)$ for all $\ell \in [m]$. Since the hypothesis of Lemma 2.19 holds, we have $b'(\ell) \geq b(\ell)$ for every $\ell \in [m]$, and by definition we have $b(\ell) = \mathsf{Top}_\ell(\vec{b})$. Therefore, Theorem 2.7 yields $f(y) \geq f(\vec{b})$.

We prove part (b) by using Lemma 2.17. To show that $f(y) \leq 2f(\vec{b})$, it suffices to show that for all $k \in [m]$, we have $\mathsf{Top}_k(y) \leq 2\mathsf{Top}_k(\vec{b})$. By our assumption, this is true for $k \in \mathsf{POS}$ because we have $\mathsf{Top}_k(y) \leq B_k \leq b(k)$, where we use Lemma 2.17(i). For $k \in [m] \setminus \mathsf{POS}$, let $\ell := 2^{\lfloor \log_2 k \rfloor} \in \mathsf{POS}$ denote the largest index in POS that is at most k. Note $\ell \leq k < 2\ell$. Now, $\mathsf{Top}_k(y) \leq 2\mathsf{Top}_\ell(y) \leq 2b(\ell) \leq 2b(k)$ holds because b is nondecreasing. Therefore, $f(y) \leq 2f(\vec{b})$ follows. ■

2.3 Enumeration Tools for Guessing a Sorted Vector

Our algorithms will often need to estimate a nonincreasing (or nondecreasing) vector x whose coordinates are some statistics of the optimal cost vector. We show that if we have suitable bounds on the coordinates of x, then we can identify a (polynomially bounded) set containing a vector close to x. The entries of the "guess vector" for x will usually consist of powers of 2 (or powers of $(1+\varepsilon)$ for some $\varepsilon > 0$), so our enumeration of the guess vector will involve enumerating over vectors with monotone integer coordinates. The following result, lifted from [5], gives a convenient bound for this task and will be useful in the proof of Lemma 2.21.

Claim 2.20. *There are at most $(2e)^{\max(M,k)}$ nonincreasing sequences of k integers chosen from $\{0, \dots, M\}$.*

Proof. We reproduce the argument from [5]. A nondecreasing sequence $a_1 \geq a_2 \geq \cdots \geq a_k$, where $a_i \in \{0\} \cup [M]$ for all $i \in [k]$, can be mapped bijectively to the set of $k+1$ integers $M - a_1, a_1 - a_2, \dots, a_{k-1} - a_k, a_k$ from $\{0\} \cup [M]$ that add up to M. The number of such sequences of $k+1$ integers is equal to the coefficient of x^M in the generating function $(1 + x + \cdots + x^M)^k$. This is equal to the coefficient of x^M in $(1-x)^{-k}$, which is $\binom{M+k-1}{M}$ using the binomial expansion. Let $U = \max\{M, k-1\}$. We have

$$\binom{M+k-1}{M} = \binom{M+k-1}{U} \leq \Big(\frac{e(M+k-1)}{U}\Big)^U \leq (2e)^U. \qquad \blacksquare$$

Lemma 2.21. *Let $\mathcal{I} \subseteq [m]$ be an index-set. Let $x \in \mathbb{R}^{\mathcal{I}}_{\geq 0}$ be a nonincreasing vector, i.e., $x_\ell \geq x_{\ell'}$ for indices $\ell, \ell' \in \mathcal{I}$, $\ell < \ell'$. Let ub be such that $x_\ell \leq \mathsf{ub}$ for all $\ell \in \mathcal{I}$. Let $\varepsilon, \kappa > 0$, and $\varepsilon' = \min\{1, \varepsilon\}$.*

(a) *We can construct a set $\mathcal{T} \subseteq \mathbb{R}^{\mathcal{I}}_{\geq 0}$ with $|\mathcal{T}| \leq N := (2e)^{|\mathcal{I}|} + \big(\mathsf{ub}/\kappa\big)^{O(1/\varepsilon')}$ in $O(N)$ time, containing a nonincreasing vector $y \in \mathbb{R}^{\mathcal{I}}_{\geq 0}$, such that $x_\ell \leq y_\ell \leq (1+\varepsilon)x_\ell + \kappa$ for all $\ell \in \mathcal{I}$.*
(b) *Suppose that we also have $x_\ell \geq \mathsf{lb}$ for all $\ell \in \mathcal{I}$, where $\mathsf{lb} > 0$. We can construct $\mathcal{T} \subseteq \mathbb{R}^{\mathcal{I}}_{\geq 0}$ with $|\mathcal{T}| \leq N := (2e)^{|\mathcal{I}|} + \big(\mathsf{ub}/\mathsf{lb}\big)^{O(1/\varepsilon')}$ in $O(N)$ time, containing a nonincreasing vector $y \in \mathbb{R}^{\mathcal{I}}_{\geq 0}$, such that $x_\ell \leq y_\ell \leq (1+\varepsilon)x_\ell$ for all $\ell \in \mathcal{I}$.*

Proof. For part (a), consider the set

$$\mathcal{T} := \Big\{\vec{t} \in \mathbb{R}^{\mathcal{I}}_{\geq 0} : \quad \vec{t} \text{ is a nonincreasing vector}, \quad \forall \ell \in \mathcal{I}, \quad t_\ell = \mathsf{ub}/(1+\varepsilon)^k, \text{ where } k \in \mathbb{Z}_{\geq 0}, \quad t_\ell \geq \kappa/(1+\varepsilon)\Big\}.$$

Each $\vec{t} \in \mathcal{T}$ is a nonincreasing vector, and there are

$$U := 1 + \lfloor \log_{1+\varepsilon} \frac{(1+\varepsilon)\mathsf{ub}}{\kappa} \rfloor = O\Big(\frac{\log(\mathsf{ub}/\kappa)}{\varepsilon'}\Big)$$

choices for $\log_{1+\varepsilon}(\mathsf{ub}/t_\ell)$ for every $\ell \in \mathcal{I}$. (Recall that $\varepsilon' = \min\{\varepsilon, 1\}$.) So Claim 2.20 implies that $|\mathcal{T}| \leq (2e)^{\max\{|\mathcal{I}|,U\}}$. We have $(2e)^U \leq \big(\mathsf{ub}/\kappa\big)^{O(1/\varepsilon')}$, so this yields the bound on $|\mathcal{T}|$ and the time to construct $\mathcal{T}$.

Consider the vector $y \in \mathbb{R}^{\mathcal{I}}_{\geq 0}$, where for every $\ell \in \mathcal{I}$, y_ℓ is the smallest number of the form $\mathsf{ub}/(1+\varepsilon)^k$, $k \in \mathbb{Z}_{\geq 0}$ that is at least $\max\{\kappa/(1+\varepsilon), x_\ell\}$. Then, y is a nonincreasing vector, $y \in \mathcal{T}$, and $x_\ell \leq y_\ell \leq (1+\varepsilon)x_\ell + \kappa$ for all $\ell \in \mathcal{I}$.

Part (b) is proved very similarly. We now take $\mathcal{T}$ to be the set of all nonincreasing vectors $\vec{t} \in \mathbb{R}^{\mathcal{I}}_{\geq 0}$ such that for all $\ell \in \mathcal{I}$, we have $t_\ell \geq \mathsf{lb}$, $t_\ell = \mathsf{ub}/(1+\varepsilon)^k$ where $k \in \mathbb{Z}_{\geq 0}$. As before, one can infer that the size of $\mathcal{T}$ and the time taken to construct it are bounded by $(2e)^{|\mathcal{I}|} + \big(\mathsf{ub}/\mathsf{lb}\big)^{O(1/\varepsilon')}$. Now if $y \in \mathbb{R}^{\mathcal{I}}_{\geq 0}$ is such that, for every $\ell \in \mathcal{I}$, y_ℓ is the smallest number of the form $\mathsf{ub}/(1+\varepsilon)^k$, $k \in \mathbb{Z}_{\geq 0}$ that is at least x_ℓ, then, y is a nonincreasing vector, $y \in \mathcal{T}$, and $x_\ell \leq y_\ell \leq (1+\varepsilon)x_\ell$ for all $\ell \in \mathcal{I}$. ■

2.4 Probability Theory

In this section, we discuss some tools from probability theory that are useful in our proofs and analyses. We first define some well-known probability distributions, and state associated facts that are extensively used in this thesis.

Definition 2.22 (Weighted Bernoulli trial)**.** *A discrete random variable B is said to be a weighted Bernoulli trial of type (q, s) if B takes value s with probability q, and 0 otherwise. If $s = 1$, then we call B as a Bernoulli trial with success probability q.*

Definition 2.23 (Poisson random variable)**.** *A discrete random variable Z is said to have a Poisson distribution with parameter $\lambda \geq 0$, denoted $Z \sim \mathsf{Pois}(\lambda)$, if for all $k \in \mathbb{Z}_{\geq 0}$ we have $\mathbf{Pr}[Z = k] = e^{-\lambda}\lambda^k/k!$*

Fact 2.1. *The mean and variance of $\mathsf{Pois}(\lambda)$ are both equal to λ.*

Fact 2.2. *Let $\{Z_j\}_j$ be a collection of independent Poisson variables with parameters $\{\lambda_j\}_j$. Then, $S = \sum_j Z_j$ is distributed as $\mathsf{Pois}(\sum_j \lambda_j)$.*

Definition 2.24 (Binomial random variable)**.** *Let $n \in \mathbb{Z}_{>0}$ and $p \in [0,1]$. The binomial distribution with parameters n and p, denoted $\mathsf{Bin}(n,p)$, is the sum of n independent Bernoulli trials each with success probability p. The probability distribution function of $\mathsf{Bin}(n,p)$ is given by: for any $k \in \{0,1,\ldots,n\}$,*

$$\mathbf{Pr}[\mathsf{Bin}(n,p) = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Fact 2.3. *For any $n \in \mathbb{Z}_{>0}$ and $p \in [0,1]$, the median of $\mathsf{Bin}(n,p)$ lies in $\{\lfloor np \rfloor, \lceil np \rceil\}$.*

The following result will be useful to us. We thank Daniel Perales Anaya for a proof of this result.

Lemma 2.25. *We have $\mathbf{Pr}[\mathsf{Bin}(n,k/n) = k] \le 1/2$ for integers n,k satisfying $1 \le k < n$.*

Proof. For any $r \in \{0,1,\ldots,n\}$, define $p_r := \mathbf{Pr}[\mathsf{Bin}(n,k/n) = r]$. We prove the lemma by showing that $p_k \le p_{k-1} + p_{k+1}$, which would then imply that $2p_k \le p_{k-1} + p_k + p_{k+1} \le 1$. We now show that $p_k \le p_{k-1} + p_{k+1}$ holds. We have:

$$\begin{aligned}
\frac{p_{k-1}+p_{k+1}}{p_k} &= \frac{\binom{n}{k-1}\cdot\left(\frac{k}{n}\right)^{k-1}\cdot\left(1-\frac{k}{n}\right)^{n-k+1} + \binom{n}{k+1}\cdot\left(\frac{k}{n}\right)^{k+1}\cdot\left(1-\frac{k}{n}\right)^{n-k-1}}{\binom{n}{k}\cdot\left(\frac{k}{n}\right)^{k}\cdot\left(1-\frac{k}{n}\right)^{n-k}} \\
&= \frac{k}{n-k+1}\cdot\frac{n}{k}\cdot\frac{n-k}{n} + \frac{n-k}{k+1}\cdot\frac{k}{n}\cdot\frac{n}{n-k} \\
&= \left(1-\frac{1}{n-k+1}\right) + \left(1-\frac{1}{k+1}\right) \\
&\ge \frac{1}{2}+\frac{1}{2} = 1 \qquad \text{(since } n > k \text{ and } k \ge 1\text{)}
\end{aligned}$$

■

2.4.1 Concentration Inequalities

The following general-purpose concentration inequalities will be useful to us.

Lemma 2.26 (Markov's inequality)**.** *Let Z be a nonnegative random variable, and $\alpha > 0$ be a scalar. Then, $\mathbf{Pr}[Z \ge \alpha] \le \mathbf{E}[Z]/\alpha$.*

Lemma 2.27 (Chebyshev's inequality)**.** *Let Z be a nonnegative random variable with finite mean and variance. Then, for any $\alpha > 0$, $\mathbf{Pr}[|Z - \mathbf{E}[Z]| \ge \alpha] \le \mathrm{Var}[Z]/\alpha^2$.*

The Chernoff tail bounds, due to Herman Rubin, will be useful to us in Chapter 4 to prove a bound on the expected f-norm of a random vector, and in Chapter 7 to give a simple randomized algorithm for vector scheduling on identical machines.

Lemma 2.28 (Chernoff tail bounds)**.** *Let $Z_1, \ldots, Z_k$ be independent $[0,1]$-bounded random variables, and $S = \sum_{j\in[k]} Z_j$ denote their sum. The following tail bounds hold:*

(1) *(Upper Tail) For any $\mu \geq \mathbf{E}[S]$ and $\delta \geq 0$, we have:*

$$\mathbf{Pr}\Big[S \geq (1+\delta)\mu\Big] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}.$$

For $\delta \geq 1$, we have the following simpler bound: $\mathbf{Pr}[S \geq (1+\delta)\mu] \leq e^{-\delta\mu/3}$.

(2) *(Lower Tail) For any $\mu \leq \mathbf{E}[S]$ and $\delta \in [0,1]$, we have:*

$$\mathbf{Pr}[S \leq (1-\delta)\mu] \leq e^{-\delta^2\mu/2}.$$

2.4.2 Hoeffding's Lemma

To optimize constants that appear in some of our important results, such as Theorems 3.9 and 4.1, we have to obtain strong tail-probability bounds on sums of independent Bernoulli trials. General-purpose concentration inequalities, like the ones in Section 2.4.1, are often too wasteful. The following result due to Hoeffding will be useful for our purposes.

Lemma 2.29 (Corollary 2.1 in [15])**.** *Let n be a positive integer, $g : \{0,1,\ldots,n\} \to \mathbb{R}$ be an arbitrary real-valued function, and $\mu \in [0,n]$ be a real number. Let $B_1, \ldots, B_n$ be n independent Bernoulli trials with $\sum_{j\in[n]} \mathbf{Pr}[B_j = 1] = \mu$ such that $\mathbf{E}[g(B_1 + \cdots + B_n)]$ is minimized (or maximized). Then, we may assume that there exists $q \in (0,1)$ such that $\mathbf{Pr}[B_j = 1] \in \{0, q, 1\}$ for all $j \in [n]$.*

The following lemma can be obtained by a straightforward application of Lemma 2.29.

Lemma 2.30. *Let $\{B_j\}_{j\in[n]}$ be a collection of n independent Bernoulli trials, and $S := \sum_{j\in[n]} B_j$ denote their sum. Let $\mu := \mathbf{E}[S] = \sum_{j\in[n]} \mathbf{Pr}[B_j = 1]$. The following are true:*

(i) *If $\mu \leq 1$, then $\mathbf{Pr}[S \geq 1] \geq 1 - e^{-\mu} \geq (1-1/e)\mu$;*

(ii) *If $\mu \geq 1$, then $\mathbf{Pr}[S \geq 1] \geq 1 - 1/e$; and*

(iii) *If* $\mu \geq 1$*, then* $\mathbf{Pr}[S \geq \lfloor \mu \rfloor] \geq 1/2$.

Proof. For convenience, for $j \in [n]$ let $q_j := \mathbf{Pr}[B_j = 1]$. We start with the first claim. Suppose that $\mu \in [0, 1]$. Consider the function $g : \{0, 1, \ldots, n\} \to \{0, 1\}$ that satisfies $g(r) = 0$ if and only if $r = 0$. Since $\mathbf{E}[g(B_1 + \cdots + B_m)]$ is exactly $\mathbf{Pr}[S \geq 1]$, our problem reduces to arguing that $\mathbf{E}[g(B_1 + \cdots + B_m)]$ is lower bounded by $(1 - 1/e)\mu$. By Lemma 2.29, we may further assume that all q_j's take a uniform value μ/n. This is because (i) we can drop variables with $q_j = 0$, and (ii) if any $q_j = 1$, then we must have $n = \mu = 1$, which implies $\mathbf{Pr}[S \geq 1] = 1$. Therefore, we get:

$$\mathbf{Pr}[S \geq 1] = 1 - \mathbf{Pr}[B_j = 0 \, \forall j] = 1 - (1 - \mu/n)^n \geq 1 - e^{-\mu} \geq (1 - 1/e)\mu,$$

where the final inequality uses $\mu \in [0, 1]$.

Next, for the second claim, suppose that $\mu \geq 1$. We reuse the function g defined above. As g is nondecreasing, we may assume that $\mu = 1$. Then, by (i), we have $\mathbf{Pr}[S \geq 1] \geq 1 - 1/e$.

Lastly, for the third claim, suppose that $\mu \geq 1$. Similar to what we did before, consider the 0-1 function g that maps r to 0 if and only if $0 \leq r < \lfloor \mu \rfloor$. Again, $\mathbf{E}[g(B_1 + \cdots + B_n)] = \mathbf{Pr}[S \geq \lfloor \mu \rfloor]$, which is the quantity that we want to lower bound. Since g is nondecreasing, we may assume that $\mu = \lfloor \mu \rfloor$: multiplying every q_j with $\lfloor \mu \rfloor / \mu$ may only lead to a decrease in $\mathbf{Pr}[S \geq \lfloor \mu \rfloor]$. By Lemma 2.29, we may assume that each q_j is either a common value $q \in (0, 1)$ or 1. Note that we can drop variables with $q_j = 0$. Suppose that $k \in \{0, 1, \ldots, \mu\}$ of the B_j's have $q_j = 1$, so that $q = (\mu - k)/(n - k)$. Then, $\mathbf{Pr}[S \geq \mu] = \mathbf{Pr}[\mathsf{Bin}(n - k, (\mu - k)/(n - k)) \geq \mu - k] \geq 1/2$, because the median of $\mathsf{Bin}(n, p)$ is either $\lfloor np \rfloor$ or $\lceil np \rceil$, both of which evaluate to $\mu - k$ in this case. ■

The following result is immediate.

Lemma 2.31. *Let* $S = \sum_{j \in [n]} B_j$ *be a sum of* n *independent Bernoulli trials, and* ℓ *be a positive integer. If* $\mathbf{E}[S] \geq \ell$*, then* $\mathbf{E}[\min(\ell, S)] \geq \ell/2$.

Proof. We use Lemma 2.30(iii): $\mathbf{E}[\min(\ell, S)] \geq \ell \cdot \mathbf{Pr}[S \geq \ell] \geq \ell/2$. ■

2.5 Iterative Rounding Framework

Our algorithms are based on rounding fractional solutions to LP-relaxations that we formulate for the stochastic min-norm versions of load balancing and spanning trees. The rounding algorithm needs to ensure that the various budget constraints that we include in our LP to control quantities associated with expected Top_ℓ norms (for multiple indices ℓ) are roughly preserved. The main technical tool for achieving this is *iterative rounding*, as expressed by the following theorem, which follows from a result in Linhares et al. [30].

Theorem 2.32 (Follows from Corollary 11 in [30]). *Let $\mathcal{M} = (\mathcal{U}, \mathcal{I})$ be a matroid with rank function r (specified via a value oracle), and $\mathcal{Q} := \{z \in \mathbb{R}_{\geq 0}^{\mathcal{U}} : z(\mathcal{U}) = \mathsf{r}(\mathcal{U}), \quad z(F) \leq \mathsf{r}(F)\ \forall F \subseteq \mathcal{U}\}$ be its base polytope. Let $\overline{z}$ be a feasible solution to the following multi-budgeted matroid LP:*

$$\min \quad c^T z \qquad s.t. \qquad Az \leq b, \quad z \in \mathcal{Q}. \tag{Budg-LP}$$

where $A \in \mathbb{R}_{\geq 0}^{k \times U}$ and $c \in \mathbb{R}^{\mathcal{U}}$. Let $\nu \in \mathbb{R}_{\geq 0}$ be a parameter such that for all $e \in \operatorname{supp}(\overline{z})$, $\sum_{i \in [k]} A_{i,e} \leq \nu$ holds. In polynomial time, we can round $\overline{z}$ to obtain a basis B of $\mathcal{M}$ satisfying: (a) $c(B) \leq c^T \overline{z}$; (b) $A\chi^B \leq b + \nu \mathbf{1}$, where $\mathbf{1}$ is the vector of all 1's; and (c) B is contained in the support of $\overline{z}$.

Proof. We first consider a new instance where we move to the support of $\overline{z}$, which will automatically take care of (c). More precisely, let $J = \operatorname{supp}(\overline{z})$. For a vector $v \in \mathbb{R}^U$, let $v_J := (v_e)_{e \in J}$ denote the restriction of v to the coordinates in J. Let $\mathcal{M}_J = (J, \mathcal{I}_J)$ with rank function r_J, be the restriction of the matroid $\mathcal{M}$ to J. Let A_J be A restricted to the columns corresponding to J. Note that $\mathsf{r}_J(J) = \mathsf{r}(J) \geq \overline{z}(J) = \overline{z}(U) = \mathsf{r}(U)$; so we have $\mathsf{r}(J) = \overline{z}(J) = \mathsf{r}(U)$, and therefore a basis of $\mathcal{M}_J$ is also a basis of $\mathcal{M}$. It is easy to see now that $\overline{z}_J$ is a feasible solution to (Budg-LP) where we replace c and A by c_J and A_J respectively, and replace $\mathcal{Q}$ by the base polytope of $\mathcal{M}_J$. It suffices to show how to round $\overline{z}_J$ to obtain a basis B of $\mathcal{M}_J$ satisfying properties (a) and (b) (i.e., $c_J(B) \leq c_J^T \overline{z}_J$ and $A_J \chi^B \leq b + \nu \mathbf{1}$), since, by construction, we have $B \subseteq J$. Note that each column of A_J sums to at most ν.

We now describe how Corollary 11 in [30] yields the desired rounding of $\overline{z}_J$. This result pertains to a more general setting, where we have a fractional point in the the base polytope of one matroid that satisfies matroid-independence constraints for some other matroids, and some knapsack constraints. Translating Corollary 11 to our setting above, where we have only one matroid, yields the following:

> **Corollary 11 in [30] in our setting:** Let A' be obtained from A_J by scaling each row so that $\max_{e \in J} A'_{ie} = 1$ for all $i \in [k]$. Let $p_1, \ldots, p_k \geq 0$ be such that

$\sum_{i\in[k]} \frac{A'_{ie}}{p_i} \leq 1$ for all $e \in J$.[1] Then, we can round $\overline{z}_J$ to obtain a basis B of $\mathcal{M}_J$ such that $c_J(B) \leq c_J^T\overline{z}_J$, and $\sum_{e\in B} A_{ie} \leq b_i + p_i \max_{e\in J} A_{ie}$ for all $i \in [k]$.

Setting $p_i = \frac{\nu}{\max_{e\in J} A_{ie}}$ for all $i \in [k]$ satisfies the conditions above, since $\sum_{i\in[k]} \frac{A'_{ie}}{p_i} = \sum_{i\in[k]} A_{ie}/\nu \leq 1$ for all $e \in J$, and yields the desired rounding of $\overline{z}_J$. ■

[1]In [30], the p_i's are stated to be positive integers, but the proof of Corollary 11 shows that this is not actually needed.

Chapter 3

Order Statistics and Proxy Functions

In this chapter, we devise techniques for getting an approximate handle on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ by only using distributional information about the coordinates of Y. We identify two order statistics of Y that lead to two proxy functions that serve as constant-factor approximations to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$.

At a high level, a proxy function uses one or more parameters to map a vector-valued random variable to a scalar in $\mathbb{R}_{\geq 0}$. The idea here is that for a suitable choice of parameters the proxy function will serve as a constant-factor approximation to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$. An effective proxy function has three desirable properties:

- **Separability among coordinates of Y:** We want the proxy function to have a separable form so that the contribution of any individual coordinate Y_i to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ can be singled out. This is especially tricky to capture for Top_ℓ norms because of their non-separable nature.

- **Linear dependence on underlying cost random variables:** We use two examples to illustrate the benefits of having a proxy function with a linear form. First, in stochastic min-norm load balancing, Y_i corresponds to the load on machine i which is a sum of job random variables that are processed by this machine. Since the optimization decision involves figuring out which jobs to process on machine i, we do not readily have distributional information about Y_i. Thus, it is useful to have a proxy function that can capture the contribution of each Y_i (to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$) as a linear function of its constituent variables. Second, in stochastic min-norm spanning tree, the entries of Y correspond to edge-weight random variables for edges that participate in the tree. Assuming we are using a linear programming approach to solve the problem, we would be interested in rounding a fractional spanning tree that is supported on more than $|V| - 1$ edges to an integral spanning tree that is supported on exactly $|V| - 1$ edges. With linearity, the contribution of an edge variable (to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$) can be captured easily.

- **Simplicity:** A proxy function with a simple form is more likely to be applicable to a wider class of problem settings.

Our order statistics and proxy functions have all these desirable properties.

Throughout this chapter, we use Y to denote an arbitrary random vector that follows a product distribution on $\mathbb{R}^m_{\geq 0}$ (i.e., $\{Y_i\}_i$ are independent nonnegative random variables), and ℓ to denote a fixed integer in $[m]$. We recall some frequently-used notation from Chapter 2. We use $y^\downarrow$ as a shorthand for the vector $y \in \mathbb{R}^m_{\geq 0}$ with its coordinates sorted in nonincreasing order. For an event $\mathcal{E}$, we use $\mathbb{1}_\mathcal{E}$ to denote its indicator random variable, i.e., $\mathbb{1}_\mathcal{E} = 1$ if and only if event $\mathcal{E}$ happens. For any real number z, $z^+ := \max(z, 0)$. The following definition will be useful to us.

Definition 3.1 (Exceptional and Truncated Random Variables).
Let Z be a nonnegative random variable and $\theta \in \mathbb{R}_{\geq 0}$ be a scalar. We decompose $Z = Z^{<\theta} + Z^{\geq\theta}$ at the threshold θ to obtain the truncated random variable $Z^{<\theta} := Z \cdot \mathbb{1}_{Z<\theta}$, whose support lies in $[0, \theta)$, and the exceptional random variable $Z^{\geq\theta} := Z \cdot \mathbb{1}_{Z\geq\theta}$, whose support lies in $\{0\} \cup [\theta, \infty)$.

Our first order statistic and proxy function are based on exceptional variables arising from the coordinates in Y. We define an order statistic $\rho_\ell(Y)$ that is roughly $\mathbf{E}[\mathsf{Top}_\ell(Y)]/\ell$ and show that the proxy function $\ell\theta + \sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}]$ is a constant-factor approximation to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ when θ is roughly $\rho_\ell(Y)$. The ρ_ℓ statistic and the corresponding $\ell\theta + \sum_i \mathbf{E}[Y_i^{\geq\theta}]$ proxy are helpful in settings where each Y_i is a sum of independent random variables, such as stochastic load balancing (Chapters 5 – 8).

Our second order statistic and proxy function are based on the expected number of "large" coordinates of Y. Consider the random variable $N^{>\theta}(Y) := \sum_{i\in[m]} \mathbb{1}_{Y_i>\theta}$ that counts the number of coordinates of Y larger than a threshold θ. Note that $\mathbf{E}[N^{>\theta}(Y)] \in [0, m]$, and it is nonincreasing in θ. We define the order statistic $\tau_\ell(Y)$ to be the threshold θ at which $\mathbf{E}[N^{>\theta}(Y)]$ is roughly ℓ. We show that the proxy function $\ell\theta + \sum_i \mathbf{E}[(Y_i - \theta)^+]$ is a constant-factor approximation to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ when θ is roughly $\tau_\ell(Y)$. The τ_ℓ statistic and the $\ell\theta + \sum_i \mathbf{E}[(Y_i - \theta)^+]$ proxy are helpful in settings where Y_i's are "atomic" random variable with known distributions, such as stochastic spanning tree (Chapter 10).

3.1 Order Statistics and Proxy Function Based on Exceptional Variables

Our main result in this section is that the expression $\ell\theta + \sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}]$ serves as a constant-factor approximation to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ when θ is roughly $\mathbf{E}[\mathsf{Top}_\ell(Y)]/\ell$. We define the ρ_ℓ order statistic and show that $\ell\rho_\ell(Y)$ is roughly $\mathbf{E}[\mathsf{Top}_\ell(Y)]$.

Definition 3.2 (ρ_ℓ Order Statistic)**.**
For an m-dimensional random vector Y and a positive integer $\ell \in [m]$, we define:

$$\rho_\ell(Y) := \inf\Big\{\theta \in \mathbb{R}_{\geq 0} : \sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] \leq \ell\theta\Big\} \tag{3.1}$$

Theorem 3.3.
Let Y be a random vector that follows a product distribution on $\mathbb{R}^m_{\geq 0}$, and $\ell \in [m]$ be a positive integer. We have:

$$\frac{\ell\rho_\ell(Y)}{2} \leq \mathbf{E}[\mathsf{Top}_\ell(Y)] \leq 2\,\ell\rho_\ell(Y).$$

3.1.1 Effectiveness of the ρ_ℓ-Based Proxy Function: Proof of Theorem 3.3

Before delving into the details of Theorem 3.3, we explain the motivation for including the term $\sum_i \mathbf{E}[Y_i^{\geq\theta}]$ as part of our proxy function. Let $y \in \mathbb{R}_{\geq 0}$ be a deterministic vector and $\theta \in \mathbb{R}_{\geq 0}$ be a scalar. We call a scalar z *large* if $z \geq \theta$, and *small* otherwise. Suppose that $\sum_{i\in[m]} y_i^{\geq\theta} \geq \ell\theta$ holds, i.e., the sum of large coordinates of y is at least $\ell\theta$. We claim that $\mathsf{Top}_\ell(y)$ is at least $\ell\theta$. This is because y either has at least ℓ large coordinates, which implies $\mathsf{Top}_\ell(y) \geq \ell\theta$, and if not, the (fewer than ℓ) large coordinates that contribute to $\sum_i y_i^{\geq\theta}$ also contribute to $\mathsf{Top}_\ell(y)$, which again implies $\mathsf{Top}_\ell(y) \geq \ell\theta$. On the other hand, if $\sum_{i\in[m]} y_i^{\geq\theta} \leq \ell\theta$ holds, then we claim that $\mathsf{Top}_\ell(y)$ is at most $2\ell\theta$. This is because small coordinates can contribute at most $\ell\theta$ to $\mathsf{Top}_\ell(y)$ and the sum-total of large coordinates is bounded by $\ell\theta$. Interestingly, this line of reasoning extends to the stochastic setting with only a small constant-factor loss in approximation.

We now delve into the details. The proof of Theorem 3.3 follows from Lemmas 3.5 and 3.6 below. Lemma 3.5 proves an upper bound on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ by $O(\ell\theta)$, assuming a suitable upper bound on $\sum_i \mathbf{E}[Y_i^{\geq\theta}]$, the total expected size of exceptional random variables. Lemma 3.6 complements this and proves a partial converse: it lower bounds $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ by $\Omega(\ell\theta)$ assuming a suitable lower bound on $\sum_i \mathbf{E}[Y_i^{\geq\theta}]$. We begin with the easy upper-bound argument.

Lemma 3.4. *For any scalar $\theta \geq 0$, we have $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq \ell\theta + \sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}]$.*

Proof. By the definition of exceptional and truncated random variables (see Definition 3.1), we have the identity $Y_i = Y_i^{<\theta} + Y_i^{\geq\theta}$ for any $i \in [m]$. Thus,

$$\begin{aligned}
\mathbf{E}[\mathsf{Top}_\ell(Y)] &\leq \mathbf{E}\big[\mathsf{Top}_\ell(Y_1^{<\theta}, \ldots, Y_m^{<\theta})\big] + \mathbf{E}\big[\mathsf{Top}_\ell(Y_1^{\geq\theta}, \ldots, Y_m^{\geq\theta})\big] && \text{(triangle inequality)}\\
&\leq \mathbf{E}[\ell\theta] + \mathbf{E}\big[\mathsf{Top}_m(Y_1^{\geq\theta}, \ldots, Y_m^{\geq\theta})\big] && (Y_i^{<\theta} \leq \theta \text{ holds, } \mathsf{Top}_\ell(\cdot) \leq \mathsf{Top}_m(\cdot))\\
&= \ell\theta + \sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}]. && \text{(linearity of expectation)} \quad \blacksquare
\end{aligned}$$

Lemma 3.5. *If $\sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] \leq \ell\theta$ for some scalar $\theta \in \mathbb{R}_{\geq 0}$, then $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq 2\ell\theta$.*

Proof. Follows immediately from Lemma 3.4. ■

Next, we prove the harder lower-bound argument.

Lemma 3.6. *If $\sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] > \ell\theta$ for some scalar $\theta \in \mathbb{R}_{\geq 0}$, then $\mathbf{E}[\mathsf{Top}_\ell(Y)] > \ell\theta/2$.*

Proof. We prove the lemma via induction on $\ell + m$. The base case is when $\ell = m = 1$, where clearly $\mathbf{E}[\mathsf{Top}_1(Y_1)] = \mathbf{E}[Y_1] \geq \mathbf{E}[Y_1^{\geq\theta}] > \theta$. Another base case that we consider is when $m \leq \ell$: here the Top_ℓ-norm is simply the sum of all the Y_i's and thus,

$$\mathbf{E}[\mathsf{Top}_\ell(Y_1, \ldots, Y_m)] = \mathbf{E}[Y_1 + \cdots + Y_m] \geq \sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] > \ell\theta \geq \ell\theta/2.$$

Now consider the general case with $m \geq \ell+1$. Our induction strategy will be the following. If Y_m is at least θ, then we include Y_m's contribution towards the Top_ℓ-norm and collect the expected $\mathsf{Top}_{\ell-1}$-norm from the remaining coordinates. Otherwise, we simply collect the expected Top_ℓ-norm from the remaining coordinates. We use Y_{-m} to denote the vector $(Y_1, \ldots, Y_{m-1})$. Note that Y_{-m} follows a product distribution on $\mathbb{R}_{\geq 0}^{m-1}$, so we can apply induction to Y_{-m}.

We handle some easy cases separately. The case $\mathbf{E}[Y_m^{\geq\theta}] > \ell\theta$ causes a hindrance to applying the induction hypothesis. But this case is quite easy: we have

$$\mathbf{E}[\mathsf{Top}_\ell(Y_1,\ldots,Y_m)] \geq \mathbf{E}[Y_m] \geq \mathbf{E}[Y_m^{\geq\theta}] > \ell\theta \geq \ell\theta/2.$$

So we may assume that $\mathbf{E}[Y_m^{\geq\theta}] \leq \ell\theta$. Let $q := \mathbf{Pr}[Y_m \geq \theta]$. Another easy case that we handle separately is when $q = 0$. In this case, we have $\sum_{i\in[m-1]} \mathbf{E}[Y_i^{\geq\theta}] = \sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}]$, and hence $\mathbf{E}[\mathsf{Top}_\ell(Y)] \geq \mathbf{E}[\mathsf{Top}_\ell(Y_{-m})] > \ell\theta/2$ follows from the induction hypothesis.

We are now left with the case where $m \geq \ell+1$, $\mathbf{E}[Y_m^{\geq\theta}] \leq \ell\theta$ and $q = \mathbf{Pr}[Y_m \geq \theta] > 0$. Let $s := \mathbf{E}[Y_m | Y_m \geq \theta]$, which is well defined and is at least θ. Note that $qs = \mathbf{E}[Y_m^{\geq\theta}]$. We define two thresholds $\theta_1, \theta_2 \in [0,\theta]$ to apply the induction hypothesis to smaller cases.

$$\theta_1 := \frac{\ell\theta - \mathbf{E}[Y_m^{\geq\theta}]}{\ell} \quad \text{and} \quad \theta_2 := \min\left\{\theta, \frac{\ell\theta - \mathbf{E}[Y_m^{\geq\theta}]}{\ell-1}\right\}$$

Noting that $\mathbf{E}[Y_i^{\geq t}]$ is a nonincreasing function in t, observe that:

(A1) $\sum_{i\in[m-1]} \mathbf{E}[Y_i^{\geq\theta_2}] > (\ell-1)\theta_2$: since $\theta_2 \leq \theta$, we have

$$\sum_{i\in[m-1]} \mathbf{E}[Y_i^{\geq\theta_2}] \geq \sum_{i\in[m-1]} \mathbf{E}[Y_i^{\geq\theta}] > (\ell\theta - \mathbf{E}[Y_m^{\geq\theta}]) \geq (\ell-1)\theta_2.$$

(A2) $\sum_{i\in[m-1]} \mathbf{E}[Y_i^{\geq\theta_1}] > \ell\theta_1$: since $\theta_1 \leq \theta$, we have

$$\sum_{i\in[m-1]} \mathbf{E}[Y_i^{\geq\theta_1}] \geq \sum_{i\in[m-1]} \mathbf{E}[Y_i^{\geq\theta}] > \ell\theta - \mathbf{E}[Y_m^{\geq\theta}] = \ell\theta_1.$$

We now have the following chain of inequalities.

$$\begin{aligned}
\mathbf{E}[\mathsf{Top}_\ell(Y_1,\ldots,Y_m)] &\geq q\big(s + \mathbf{E}[\mathsf{Top}_{\ell-1}(Y_1,\ldots,Y_{m-1})]\big) + (1-q)\mathbf{E}[\mathsf{Top}_\ell(Y_1,\ldots,Y_{m-1})] \\
&> q\big(s + \frac{(\ell-1)\theta_2}{2}\big) + \frac{(1-q)\ell\theta_1}{2} \\
&\hspace{6em} \text{induction hypothesis, using (A1) and (A2)} \\
&= qs + \frac{q}{2}\cdot\min\{(\ell-1)\theta, \ell\theta - \mathbf{E}[Y_m^{\geq\theta}]\} + \frac{(1-q)}{2}\cdot\big(\ell\theta - \mathbf{E}[Y_m^{\geq\theta}]\big) \\
&= \frac{\ell\theta}{2} + \frac{qs}{2} + \frac{q}{2}\cdot\min\{qs-\theta, 0\} \qquad (\text{since } qs = \mathbf{E}[Y_m^{\geq\theta}]) \\
&\geq \frac{\ell\theta}{2} + \frac{qs}{2} - \frac{q\theta}{2} \\
&\geq \ell\theta/2. \qquad (\text{since } s \geq \theta) \quad \blacksquare
\end{aligned}$$

We now prove Theorem 3.3 by combining Lemmas 3.5 and 3.6. We use ρ_ℓ to refer to $\rho_\ell(Y)$ (see Definition 3.2).

Proof of Theorem 3.3. For the lower bound argument, consider an arbitrary scalar $\theta \in [0, \rho_\ell)$. By definition, we have $\sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] > \ell\theta$. By Lemma 3.6 we get $\mathbf{E}[\mathsf{Top}_\ell(Y)] > \ell\theta/2$. Taking a supremum over all $\theta < \rho_\ell$ gives the desired lower bound on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$.

For the upper bound argument, consider an arbitrary scalar $\theta \in (\rho_\ell, \infty)$. Again, by definition, we have $\sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] \leq \ell\theta$. By Lemma 3.5 we get $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq 2\ell\theta$. Taking an infimum over all $\theta > \rho_\ell$ gives the desired upper bound on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$. ■

3.2 Order Statistics and Proxy Function Based on $\mathbf{E}[N^{>\theta}(Y)]$

We now present an alternate proxy function for $\mathbf{E}[\mathsf{Top}_\ell(Y)]$, based on the expected number of "large" coordinates of Y. Our main result in this section is that the expression $\ell\theta + \sum_{i\in[m]} \mathbf{E}[(Y_i - \theta)^+]$ serves as a constant-factor approximation to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ when θ is such that $\mathbf{E}[N^{>\theta}(Y)] \approx \ell$. Recall that for a scalar $z \in \mathbb{R}$, $z^+ := \max(z, 0)$, and the random variable $N^{>\theta}(Y) = \sum_{i\in[m]} \mathbb{1}_{Y_i > \theta}$ counts the number of coordinates of Y that exceed θ. We now define the τ_ℓ order statistic and the γ_ℓ proxy function (see Figure 3.1 for intuition).

Definition 3.7 (τ_ℓ Order Statistic)**.**
For an m-dimensional random vector Y and a nonnegative integer ℓ, we define:

$$\tau_\ell(Y) := \inf\{\theta \in \mathbb{R}_{\geq 0} : \mathbf{E}[N^{>\theta}(Y)] < \ell\} \tag{3.2}$$

Note that $\tau_0(Y) := \infty$ and $\tau_\ell(Y) = 0$ for $\ell > m$.

We make two remarks on the $\tau_\ell(Y)$ order statistic. First, it is easy to see that τ_ℓ is nonincreasing over ℓ. Next, the infimum in the definition of τ_ℓ is actually attained, i.e., $\mathbf{E}[N^{>\tau_\ell}(Y)] < \ell$. This is because the function $\mathbf{E}[N^{>\theta}(Y)]$ is right-continuous over $\mathbb{R}$ (see Lemma 3.10 for details).

Definition 3.8 (γ_ℓ Proxy Function)**.**
For an m-dimensional random vector Y and a positive integer ℓ, we define:

$$\gamma_\ell(Y) := \ell \cdot \tau_\ell(Y) + \sum_{i\in[m]} \mathbf{E}\Big[\big(Y_i - \tau_\ell(Y)\big)^+\Big]. \tag{3.3}$$

We define $\gamma_0(Y) := 0$, and note that $\gamma_\ell(Y) = \sum_{i\in[m]} \mathbf{E}[Y_i] = \mathbf{E}[\mathsf{Top}_m(Y)]$ for $\ell > m$.

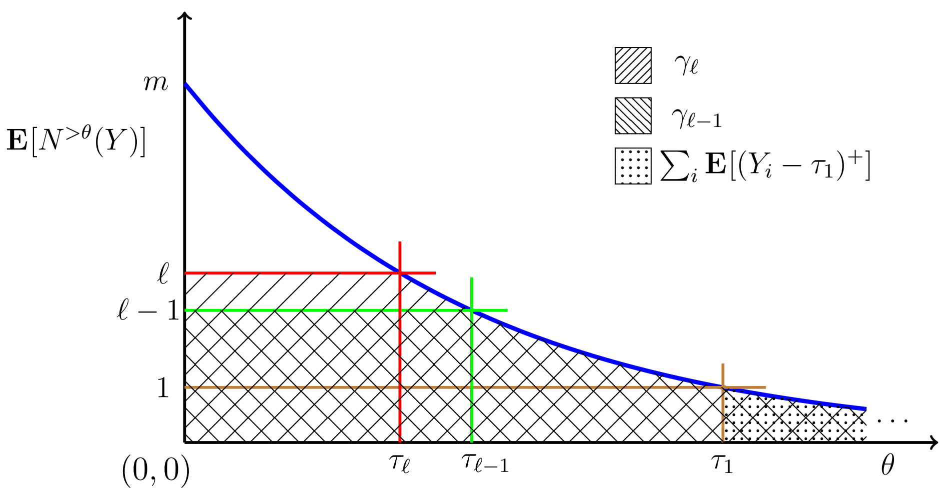

Figure 3.1: The expected histogram curve: $\mathbf{E}[N^{>\theta}(Y)]$ vs. θ.

Theorem 3.9.
Let Y be a random vector that follows a product distribution on $\mathbb{R}^m_{\geq 0}$, and $\ell \in [m]$ be a positive integer. We have:

$$\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq \gamma_\ell(Y) \leq 2\,\mathbf{E}[\mathsf{Top}_\ell(Y)].$$

Furthermore, when $\ell = 1$, we have the improved bound $\gamma_1(Y) \leq \frac{e}{e-1}\mathbf{E}[\mathsf{Top}_1(Y)]$.

Note that for a deterministic vector $y \in \mathbb{R}^m_{\geq 0}$, $\tau_\ell(y)$ is simply the ℓ^{th} largest entry in y, i.e., $\tau_\ell(y) = y^\downarrow_\ell$, and $\gamma_\ell(y) = \ell y^\downarrow_\ell + \sum_i (y_i - y^\downarrow_\ell)^+$ is the Top_ℓ-norm of y. Our key insight is that $\gamma_\ell(Y)$ continues to remain a good *proxy* for $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ in the stochastic setting.

3.2.1 Effectiveness of the τ_ℓ-Based Proxy Function: Proof of Theorem 3.9

We give some further motivation behind the definitions of τ_ℓ and γ_ℓ, which quite easily yields the upper bound on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ in Theorem 3.9. Recall from Lemma 2.4 that for any $y \in \mathbb{R}^m_{\geq 0}$, we have the following two equivalent expressions for $\mathsf{Top}_\ell(y)$:

$$\mathsf{Top}_\ell(y) \quad = \quad \ell y^{\downarrow}_\ell + \sum_{i\in[m]} (y_i - y^{\downarrow}_\ell) \cdot \mathbb{1}_{y_i > y^{\downarrow}_\ell} \quad = \quad \int_0^\infty \min\bigl(\ell, N^{>u}(y)\bigr) du. \tag{3.4}$$

Therefore, $\mathbf{E}[\mathsf{Top}_\ell(Y)] = \int_0^\infty \mathbf{E}[\min\bigl(\ell, N^{>u}(y)\bigr)]du$. (We formally justify why we can interchange the expectation and integral in the proof of Theorem 3.5.) Now, by the concavity of the min function, for any $u \in \mathbb{R}_{\geq 0}$ we have

$$\mathbf{E}_{y\sim Y}[\min\bigl(\ell, N^{>u}(y)\bigr)] \leq \min\bigl(\ell, \mathbf{E}_{y\sim Y}[N^{>u}(y)]\bigr) = \min\bigl(\ell, \mathbf{E}[N^{>u}(Y)]\bigr), \tag{3.5}$$

and using the definition of τ_ℓ, one can show that $\int_0^\infty \min(\ell, \mathbf{E}[N^{>u}(Y)])du$ is precisely $\gamma_\ell(Y)$. Thus, the difference between $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ and $\gamma_\ell(Y)$ is that we have replaced the expectation of the minimum by the minimum of an expectation, which leads to the upper bound. The following lemma will be useful in formalizing the above upper-bound argument.

Lemma 3.10. *Let $\tau_\ell = \tau_\ell(Y)$. We have: (i) $\mathbf{E}[N^{>\tau_\ell}(Y)] < \ell$; and (ii) $\mathbf{E}[N^{\geq\tau_\ell}(Y)] \geq \ell$.*

Proof. Let $\tau_\ell = \tau_\ell(Y)$. The cumulative distribution function of any random variable is right-continuous, so $\mathbf{E}[N^{>\theta}(Y)] = \sum_{i\in[m]}\bigl(1 - \mathbf{Pr}[Y_i \leq \theta]\bigr)$ is right-continuous as well. Therefore, the infimum in the definition of τ_ℓ (see Definition 3.7) is attained, and (i) holds. Furthermore, for any $\theta < \tau_\ell$ we have $\mathbf{E}[N^{>\theta}(Y)] \geq \ell$. Taking a limit supremum as $\theta \to \tau_\ell$ from below yields (ii):

$$\mathbf{E}[N^{\geq\tau_\ell}(Y)] = \limsup_{\theta\to\tau_\ell^-} \mathbf{E}[N^{>\theta}(Y)] \geq \ell. \qquad \blacksquare$$

The following lemma is analogous to Lemmas 3.4 and 3.5.

Lemma 3.11. *We have:*

(i) $\gamma_\ell(Y) = \int_0^\infty \min(\ell, \mathbf{E}[N^{>u}(Y)])\, du = \min_{\theta\geq 0}\Bigl\{\ell\theta + \sum_{i\in[m]} \mathbf{E}[(Y_i - \theta)^+]\Bigr\}$.

(ii) $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq \gamma_\ell(Y)$.

Proof. For any $\theta \in \mathbb{R}_{\geq 0}$, the inequality $\int_0^\infty \min(\ell, \mathbf{E}[N^{>u}(Y)])du \leq \ell\theta + \int_\theta^\infty \mathbf{E}[N^{>u}(Y)]du$ follows from splitting the integral at $u = \theta$. The above inequality is tight for $\theta = \tau_\ell$ because, by Lemma 3.10, $\mathbf{E}[N^{>u}(Y)] < \ell$ holds if and only if $u \geq \tau_\ell$ holds. Next, for any nonnegative random variable Z and $\theta \in \mathbb{R}_{\geq 0}$, we have

$$\int_\theta^\infty \mathbf{Pr}[Z > u]du = \int_0^\infty \mathbf{Pr}[(Z-\theta)^+ > u]du = \mathbf{E}[(Z-\theta)^+].$$

So $\int_\theta^\infty \mathbf{E}[N^{>u}(Y)]du = \sum_{i\in[m]} \int_\theta^\infty \mathbf{Pr}[Y_i > u]du = \sum_{i\in[m]} \mathbf{E}[(Y_i - \theta)^+]$. Claim (i) follows from recalling Definition 3.8: $\gamma_\ell(Y) = \ell\tau_\ell + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_\ell)^+]$.

For (ii), we start with (3.4):

$$\mathbf{E}[\mathsf{Top}_\ell(Y)] = \mathbf{E}_{y\sim Y}\Big[\int_0^\infty \min\big(\ell, N^{>u}(y)\big)\, du\Big] \tag{3.6a}$$

$$= \int_0^\infty \mathbf{E}_{y\sim Y}\big[\min\big(\ell, N^{>u}(y)\big)\big]\, du \tag{3.6b}$$

$$\leq \int_0^\infty \min\big(\ell, \mathbf{E}[N^{>u}(Y)]\big)\, du = \gamma_\ell(Y) \tag{3.6c}$$

We justify the above steps. The equality in (3.6b) follows from Fubini's theorem by interpreting the expectation operator as an integral: as $\mathbf{E}[|\mathsf{Top}_\ell(Y)|] \leq \sum_i \mathbf{E}[Y_i]$ is finite, Fubini's theorem allows the exchange of expectation and integration operators. The inequality in (3.6c) is due to concavity of the min function. ∎

We first prove a weaker version of Theorem 3.9 to give a quick proof of the effectiveness of the γ_ℓ proxy. The following notation and claim will be useful to us. For a nonnegative random variable Z and a scalar $\theta \in \mathbb{R}_{\geq 0}$, define $Z^{>\theta}$ to be the random variable $Z \cdot \mathbb{1}_{Z>\theta}$ whose support lies in $\{0\} \cup (\theta, \infty)$.

Claim 3.12. *The function* $\mathbf{E}[Z^{>\theta}]$ *is right-continuous over* θ *and satisfies* $\mathbf{E}[Z^{>\theta}] = \sup_{u>\theta} \mathbf{E}[Z^{\geq u}]$.

Proof. Since $\mathbf{Pr}[Z > \theta]$ is right-continuous over θ, $\mathbf{E}[Z^{>\theta}] = \int_0^\infty \mathbf{Pr}[Z^{>\theta} > u]\, du = \theta \cdot \mathbf{Pr}[Z > \theta] + \int_\theta^\infty \mathbf{Pr}[Z > u]\, du$ is right-continuous over θ. Next, by right-continuity of $\mathbf{E}[Z^{>\theta}]$ and the fact that $\mathbf{E}[Z^{\geq\theta}]$ is nonincreasing in θ, we get:

$$\mathbf{E}[Z^{>\theta}] = \sup_{u>\theta} \mathbf{E}[Z^{>u}] = \sup_{u>\theta} \mathbf{E}[Z^{\geq u}].$$ ∎

Lemma 3.13. *We have* $\gamma_\ell(Y) \leq 4\,\mathbf{E}[\mathsf{Top}_\ell(Y)]$.

Proof. Let $\rho_\ell = \rho_\ell(Y)$. By definition of ρ_ℓ (see (3.1)), $\sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] \leq \ell\theta$ for any $\theta > \rho_\ell$. Using Claim 3.12 and that $\sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}]$ is nonincreasing in θ, we get:

$$\sum_{i\in[m]} \mathbf{E}[Y_i^{>\rho_\ell}] = \sup_{\theta>\rho_\ell} \sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] = \limsup_{\substack{\theta\to\rho_\ell \\ \text{from right}}} \Big\{ \sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] \Big\} \leq \ell\rho_\ell.$$

Since the event $\{(Y_i - \rho_\ell)^+ \leq Y_i^{>\rho_\ell}\}$ always happens, Lemma 3.11(i) gives:

$$\gamma_\ell(Y) \leq \ell\rho_\ell + \sum_{i\in[m]} \mathbf{E}[(Y_i - \rho_\ell)^+] \leq \ell\rho_\ell + \sum_{i\in[m]} \mathbf{E}[Y_i^{>\rho_\ell}] \leq 2\ell\rho_\ell \leq 4\mathbf{E}[\mathsf{Top}_\ell(Y)],$$

where we use Theorem 3.3 in the final inequality. ■

We now delve into the proof of Theorem 3.9; note that we already showed $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq \gamma_\ell(Y)$ in Lemma 3.11(ii). We remark that the proof is long and technical, but not essential to understand the rest of the thesis. We split the proof into two parts. First, we prove the easy $\ell = 1$ case using just Lemma 2.30. We then give a proof sketch for the harder $\ell > 1$ case before giving the full proof.

Proof of Theorem 3.9 for $\ell = 1$. For any scalar $\theta \in \mathbb{R}_{\geq 0}$, the event $\{Y_1^{\downarrow} > \theta\}$ happens if and only if the event $\{N^{>\theta}(Y) \geq 1\}$ happens. A straightforward integration gives:

$$\begin{aligned}
\mathbf{E}[Y_1^{\downarrow}] &= \int_0^\infty \mathbf{Pr}[Y_1^{\downarrow} > \theta]\, d\theta = \int_0^\infty \mathbf{Pr}[N^{>\theta}(Y) \geq 1]\, d\theta && \text{(3.7a)} \\
&= \int_0^{\tau_1} \mathbf{Pr}[N^{>\theta}(Y) \geq 1]\, d\theta + \int_{\tau_1}^\infty \mathbf{Pr}[N^{>\theta}(Y) \geq 1]\, d\theta && \text{(3.7b)} \\
&\geq \int_0^{\tau_1} (1 - 1/e)\, d\theta + \int_{\tau_1}^\infty (1 - 1/e) \cdot \mathbf{E}[N^{>\theta}(Y)]\, d\theta && \text{(3.7c)} \\
&= (1 - 1/e) \cdot \Big(\tau_1 + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_1)^+]\Big) = (1 - 1/e)\, \gamma_1(Y). && \text{(3.7d)}
\end{aligned}$$

We justify the inequality in (3.7c) by using the definition of τ_1 and applying Lemma 2.30 for Bernoulli random variables $\{\mathbb{1}_{Y_i>\theta}\}_i$. For $\theta < \tau_1$, $\mathbf{E}[N^{>\theta}(Y)] \geq 1$ holds, and hence $\mathbf{Pr}[N^{>\theta}(Y) \geq 1] \geq 1 - 1/e$ by Lemma 2.30(ii). Next, for $\theta > \tau_1$, $\mathbf{E}[N^{>\theta}(Y)] < 1$ holds, and hence $\mathbf{Pr}[N^{>\theta}(Y) \geq 1] \geq (1 - 1/e)\mathbf{E}[N^{>\theta}(Y)]$ by Lemma 2.30(i). This finishes the proof for the $\ell = 1$ case. ■

For the $\ell > 1$ case we use Lemma 2.29 with a more careful choice of the g function. We want to show the following lower bound:

$$\mathbf{E}[\mathsf{Top}_\ell(Y)] \geq \frac{1}{2} \cdot \ell\tau_\ell + \sum_{i\in[m]} \frac{1}{2} \cdot \mathbf{E}[(Y_i - \tau_\ell)^+].$$

Call a coordinate Y_i of Y *large* if $Y_i > \tau_\ell$. Collecting the $\ell\tau_\ell/2$ term in the above expression is easy. For any $\theta < \tau_\ell$, we have $\mathbf{E}[N^{>\theta}(Y)] \geq \ell$, so by Lemma 2.30(ii), we get $\mathbf{Pr}[Y_\ell^\downarrow > \theta] \geq 1/2$, which implies that the expected Top_ℓ-norm has an $\ell\tau_\ell/2$ term. Arguing that we can collect at least half of every large coordinate of Y in the $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ calculation requires careful bookkeeping. Roughly speaking, the argument boils down to showing that we can pack 50% of all large coordinates of Y into ℓ slots. As the average number of large coordinates of Y is at most ℓ, a random sample drawn from the distribution of Y does not have too many large coordinates, so we should be able to pick every large coordinate at least 50% of the time. We now formalize the above argument.

Proof of Theorem 3.9 for $\ell > 1$. Fix some $\ell > 1$. Let $\tau_\ell = \tau_\ell(Y)$ and $\gamma_\ell = \gamma_\ell(Y)$. We reserve i to denote an arbitrary index in $[m]$. Let $Z_i := \mathbb{1}_{Y_i \geq \tau_\ell}$, and $S := \sum_i Z_i = N^{\geq\tau_\ell}(Y)$ denote the (random) number of coordinates of Y that are at least τ_ℓ. By Lemma 3.10, $\mathbf{E}[S] \geq \ell$. Without loss of generality, we may assume that the support of each Y_i lies in $\{0\} \cup [\tau_\ell, \infty)$. We can achieve this by replacing each Y_i by $Y_i^{\geq\tau_\ell} = Y_i \cdot Z_i$. While this operation may decrease $\mathbf{E}[\mathsf{Top}_\ell(Y)]$, it leaves τ_ℓ and hence γ_ℓ unchanged. τ_ℓ is unchanged since $\mathbf{Pr}[Y_i^{\geq\tau_\ell} > \theta] = \mathbf{Pr}[Y_i > \theta]$ for all $\theta \in [\tau_\ell, \infty)$; also, for $\theta \in [0, \tau_\ell)$, we have $\mathbf{Pr}[Y_i^{\geq\tau_\ell} > \theta] = \mathbf{Pr}[Y_i \geq \tau_\ell]$, and so $\sum_i \mathbf{Pr}[Y_i^{\geq\tau_\ell} > \theta] = \mathbf{E}[N^{\geq\tau_\ell}(Y)] \geq \ell$. Next, we may also assume that $\mathbf{E}[S] = \sum_i \mathbf{Pr}[Y_i \geq \tau_\ell] = \ell$ holds. We can achieve this by moving a suitable amount of probability mass (from some of the Y_i's) located at the point τ_ℓ to the point 0. This is possible since by Lemma 3.10 we have $\mathbf{E}[N^{\geq\tau_\ell}(Y)] \geq \ell > \mathbf{E}[N^{>\tau_\ell}(Y)]$. Again, this modification may decrease $\mathbf{E}[\mathsf{Top}_\ell(Y)]$, but τ_ℓ and γ_ℓ remain unchanged; in particular, the term $\mathbf{E}[(Y_i - \tau_\ell)^+]$ remains unchanged because the probability mass that is located strictly above τ_ℓ is not modified. So, to summarize, the above modifications yield that: the support of each Y_i lies in $\{0\} \cup [\tau_\ell, \infty)$, and $\mathbf{E}[S] = \ell$, where $S = N^{\geq\tau_\ell}(Y)$. That is, it suffices to prove the lower bound on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ when the Y_i's satisfy the above properties.

Consider the real-valued function $g_\ell : \{0, 1, \ldots, m\} \to \mathbb{R}_{\geq 0}$ defined as follows: $g(0) := 0$ and $g(r) = \min(1, \ell/r)$ for $1 \leq r \leq m$. The choice of g_ℓ has a very natural interpretation for our purposes. Suppose that we have $r > 0$ items that we want to uniformly pack in ℓ slots. If $r \leq \ell$, then we can pack each item fully. Otherwise, we can only pack each item up to ℓ/r portion. The following lemma is straightforward.

Lemma 3.14. *We have* $\mathsf{Top}_\ell(Y) \geq g_\ell(S) \cdot \sum_{i\in[m]} Y_i$.

Proof. Observe that the event $\{Z_i = 1\}$ happens if and only if the event $\{Y_i \neq 0\}$ happens (since we have that $Y_i \neq 0$ implies that $Y_i \geq \tau_\ell$), and hence S counts the random number of nonzero coordinates of Y. The claim is trivial when the event $\{S \leq \ell\}$ happens because $\mathsf{Top}_\ell(Y) = \sum_i Y_i$ and g_ℓ is bounded above by 1. On the other hand if the event $\{S > \ell\}$ happens, then $\mathsf{Top}_\ell(Y)$ is the sum of ℓ largest coordinates of Y, whereas $g_\ell(S)\sum_i Y_i$ is a weighted average of the coordinates of Y, where the weights are at most 1 and add up to ℓ. Therefore, $\mathsf{Top}_\ell(Y) \geq g_\ell(S)\sum_i Y_i$ holds. ∎

For any $i \in [m]$, let S_{-i} denote the random variable $S - Z_i$, which is the sum of Z_j's for $j \neq i$. Lemma 3.14 leads to the following lower bound on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$.

Lemma 3.15. $\mathbf{E}[\mathsf{Top}_\ell(Y)] \geq \mathbf{E}[\min(\ell, S)] \cdot \tau_\ell + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_\ell)^+] \cdot \mathbf{E}[g_\ell(1 + S_{-i})]$.

Proof. The proof is by an explicit calculation.

$$\begin{aligned}
\mathbf{E}[\mathsf{Top}_\ell(Y)] &\geq \mathbf{E}\Big[g_\ell(S) \cdot \sum_{i\in[m]} Y_i\Big] && \text{(by Lemma 3.14)}\\
&= \mathbf{E}\Big[g_\ell(S) \cdot \sum_{i\in[m]} \big(\min(Y_i, \tau_\ell) + (Y_i - \tau_\ell)^+\big)\Big] && (z = \min(z, \theta) + (z - \theta)^+)\\
&= \mathbf{E}\big[g_\ell(S) \cdot S \cdot \tau_\ell\big] + \mathbf{E}\Big[g_\ell(S) \cdot \sum_{i\in[m]} (Y_i - \tau_\ell)^+\Big] && (\min(Y_i, \tau_\ell) = Z_i \cdot \tau_\ell)\\
&= \mathbf{E}[\min(\ell, S)] \cdot \tau_\ell + \sum_{i\in[m]} \mathbf{E}\big[g_\ell(S) \cdot (Y_i - \tau_\ell)^+\big] && \text{(definition of } g_\ell)\\
&= \mathbf{E}[\min(\ell, S)] \cdot \tau_\ell + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_\ell)^+] \cdot \mathbf{E}[g_\ell(1 + S_{-i})].
\end{aligned}$$

The last equality is due to the independence of the Y_i random variables. For any $i \in [m]$, we can focus on the event $\{Y_i > \tau_\ell\}$, which implies that the event $\{Z_i = 1\}$ happens. But when the event $\{Z_i = 1\}$ happens, we have $g_\ell(S) = g_\ell(1 + S_{-i})$. Now since Y_i and S_{-i} are independent random variables, we can split the expectation of $g_\ell(S) \cdot (Y_i - \tau_\ell)^+$ into a product of expectations. Formally,

$$\begin{aligned}
\mathbf{E}[g_\ell(S) \cdot (Y_i - \tau_\ell)^+] &= \mathbf{Pr}[Y_i > \tau_\ell] \cdot \mathbf{E}[g_\ell(S) \cdot (Y_i - \tau_\ell)^+ \mid Y_i > \tau_\ell]\\
&= \mathbf{Pr}[Y_i > \tau_\ell] \cdot \mathbf{E}[g_\ell(1 + S_{-i}) \cdot (Y_i - \tau_\ell)^+ \mid Y_i > \tau_\ell]\\
&= \mathbf{Pr}[Y_i > \tau_\ell] \cdot \mathbf{E}[g_\ell(1 + S_{-i})] \cdot \mathbf{E}[(Y_i - \tau_\ell)^+ \mid Y_i > \tau_\ell]\\
&= \mathbf{E}[g_\ell(1 + S_{-i})] \cdot \mathbf{E}[(Y_i - \tau_\ell)^+]
\end{aligned}$$

∎

Recall our assumption that $\mathbf{E}[S] = \ell$. By Lemma 2.31, we have $\mathbf{E}[\min(\ell, S)] \geq \ell/2$. This gives us the $\ell\tau_\ell/2$ term in the lower bound argument. It remains to show that $\mathbf{E}[g_\ell(1 + S_{-i})] \geq 1/2$ for every $i \in [m]$. Since $S \geq S_{-i}$ and $g_\ell(1 + r) = \min(1, \ell/(1 + r))$ is nonincreasing in r, we have $\mathbf{E}[g_\ell(1 + S_{-i})] \geq \mathbf{E}[g_\ell(1 + S)]$. We now lower bound the latter term.

Lemma 3.16. *Let k and n be positive integers such that $k \leq n$. Let $B_1, \ldots, B_n$ be a collection of n independent Bernoulli trials, and $T = \sum_{j\in[n]} B_j$ denote their sum. Suppose $\mathbf{E}[T] = k$ holds. Then, $\mathbf{E}[g_k(1 + T)] \geq 1/2$.*

Proof. We prove the statement by induction on $k + n$. For the base cases with $k = n$, we have $\mathbf{E}[g_k(1+T)] = k/(1+k) \geq 1/2$. Consider k, n with $k < n$. Let $q_j := \mathbf{Pr}[B_j = 1]$ for $j \in [n]$. By Lemma 2.29, we may assume without loss of generality that there exists $q \in (0, 1)$ such that $q_j \in \{0, q, 1\}$ for all $j \in [n]$. If any q_j is 0, then we can ignore the corresponding B_j and use induction hypothesis for the smaller case. Next, let $k' \in \{0, 1, \ldots, k-1\}$ denote the number of variables with $q_j = 1$. Note that $q = (k - k')/(n - k')$ holds. If $k' > 0$, then we can use the induction hypothesis:

$$\begin{aligned}
\mathbf{E}[g_k(1 + T)] &= \mathbf{E}[g_k(1 + k' + \mathsf{Bin}(n - k', q))] && (3.8a)\\
&= \mathbf{E}\left[\min\Big(1, \frac{k}{1 + k' + \mathsf{Bin}(n - k', q)}\Big)\right] && (3.8b)\\
&\geq \mathbf{E}\left[\min\Big(1, \frac{k - k'}{1 + \mathsf{Bin}(n - k', q)}\Big)\right] && (3.8c)\\
&= \mathbf{E}[g_{k-k'}(1 + \mathsf{Bin}(n - k', q))]. && (3.8d)
\end{aligned}$$

We justify the inequality in (3.8c). Observe that if the event $\{1 + k' + \mathsf{Bin}(n - k', q) \leq k\}$ happens, then the minimums in (3.8b) and (3.8c) are both 1. Otherwise, the inequality holds because for any scalars $b \geq a > \theta$, we have $a/b \geq (a - \theta)/(b - \theta)$.

Assuming $k' = 0$, it remains to show that for any integers $1 \leq k < n$, the following inequality holds:

$$\mathbf{E}[g_k(1 + \mathsf{Bin}(n, k/n))] \geq \frac{1}{2} \tag{3.9}$$

The following notation will be useful to show that (3.9) holds. For any $r \in \{0, 1, \ldots, n\}$, define $p_r := \mathbf{Pr}[\mathsf{Bin}(n, k/n) = r]$, $p_{\geq r} := \sum_{r'=r}^{n} p_{r'}$, and $p_{\leq r} := \sum_{r'=1}^{r} p_{r'}$. We have:

$$\mathbf{E}[g_k(1 + \mathsf{Bin}(n, k/n))] = \mathbf{E}\left[\min\Big(1, \frac{k}{1 + \mathsf{Bin}(n, k/n)}\Big)\right]$$

$$
\begin{aligned}
&= \Big(\sum_{r=0}^{k-1} p_r \cdot 1\Big) + \Big(\sum_{r=k}^{n} p_r \cdot \frac{k}{1+r}\Big) \\
&= p_{\le k-1} + \sum_{r=k}^{n} \frac{n!}{r!\,(n-r)!} \cdot \left(\frac{k}{n}\right)^{r} \cdot \left(1-\frac{k}{n}\right)^{n-r} \cdot \frac{k}{1+r} \\
&= p_{\le k-1} + \frac{k}{n+1} \cdot \frac{n}{k} \cdot \sum_{r=k}^{n} \frac{(n+1)!}{(r+1)!\,(n-r)!} \cdot \left(\frac{k}{n}\right)^{r+1} \cdot \left(1-\frac{k}{n}\right)^{n-r} \\
&= p_{\le k-1} + \frac{n}{n+1} \cdot \mathbf{Pr}[\mathsf{Bin}(n+1, k/n) \ge k+1] \\
&= p_{\le k-1} + \frac{n}{n+1} \cdot \left\{\frac{k}{n} \cdot p_{\ge k} + \left(1-\frac{k}{n}\right) \cdot p_{\ge k+1}\right\} \\
&= p_{\le k-1} + \frac{k}{n+1} \cdot p_{\ge k} + \frac{n-k}{n+1} \cdot p_{\ge k+1} \\
&= p_{\le k-1} + \frac{k}{n+1} \cdot p_{k} + \frac{n}{n+1} \cdot p_{\ge k+1} \\
&= 1 - \frac{(n+1-k)p_k + p_{\ge k+1}}{n+1} \ge \frac{1}{2}.
\end{aligned}
$$

It remains to justify the final inequality. Since the median of $\mathsf{Bin}(n, k/n)$ is k (see Fact 2.3), we have $p_{\ge k+1} \le 1/2$. Next, by Lemma 2.25, we have $p_k \le 1/2$. Since $n > k \ge 1$, we get $(n+1-k)p_k + p_{\ge k+1} \le n \cdot \frac{1}{2} + \frac{1}{2} \le \frac{n+1}{2}$. ■

We finish the proof of Theorem 3.9 by using Lemmas 3.15, 2.31, and 3.16.

$$
\begin{aligned}
\mathbf{E}[\mathsf{Top}_\ell(Y)] &\ge \mathbf{E}[\min(\ell, S)] \cdot \tau_\ell + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_\ell)^+] \cdot \mathbf{E}[g_\ell(1 + S_{-i})] \\
&\ge \frac{\ell}{2} \cdot \tau_\ell + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_\ell)^+] \cdot \frac{1}{2} = \frac{\gamma_\ell(Y)}{2}.
\end{aligned}
$$
■

3.3 Some Auxiliary Results

The following result will be useful in Chapter 4.

Lemma 3.17. *For any $\ell \in \{2, \dots, m\}$, we have:*

$$\tau_\ell(Y) \le \gamma_\ell(Y) - \gamma_{\ell-1}(Y) \le \tau_{\ell-1}(Y).$$

Proof. For convenience, we drop the argument Y from $\tau_\ell(Y)$ and $\gamma_\ell(Y)$ notation. Both inequalities are easily inferred from Fig. 3.1. Algebraically, we have

$$\gamma_\ell - \gamma_{\ell-1} = \int_0^\infty \left\{\min(\ell, \mathbf{E}[N^{>\theta}(Y)]) - \min(\ell-1, \mathbf{E}[N^{>\theta}(Y)])\right\} d\theta.$$

For $\theta \geq \tau_{\ell-1}$, the integrand is 0 since $\mathbf{E}[N^{>\theta}(Y)] < \ell - 1$; for $\theta < \tau_\ell$, the integrand is 1 since $\mathbf{E}[N^{>\theta}(Y)] \geq \ell$; and for $\tau_\ell \leq \theta < \tau_{\ell-1}$, the integrand is at most 1 since $\ell - 1 \leq \mathbf{E}[N^{>\theta}(Y)] < \ell$. It follows that $\tau_{\ell-1} \geq \gamma_\ell - \gamma_{\ell-1} \geq \tau_\ell$. ∎

Theorem 3.18. *Let Y follow a product distribution on $\mathbb{R}^m_{\geq 0}$. We have:*

(i) $\mathbf{E}[Y^\downarrow_1] \geq (1-1/e)\gamma_1(Y) \geq (1-1/e)\tau_1(Y)$*; and*

(ii) *for any $\ell \in \{2, 3, \ldots, m\}$, we have $\mathbf{E}[Y^\downarrow_\ell] \geq \tau_\ell(Y)/2$.*

Proof. The first claim follows from Theorem 3.9 for the $\ell = 1$ case. For the second claim, fix some $\ell \in \{2, \ldots, m\}$. By definition of τ_ℓ, for any $\theta < \tau_\ell$ we have $\mathbf{E}[N^{>\theta}(Y)] \geq \ell$. Observe that if the event $\{N^{>\theta}(Y) \geq \ell\}$ happens, then the event $\{Y^\downarrow_\ell > \theta\}$ happens. A simple calculation gives:

$$\mathbf{E}[Y^\downarrow_\ell] \geq \int_0^{\tau_\ell} \mathbf{Pr}[Y^\downarrow_\ell > \theta] d\theta \geq \int_0^{\tau_\ell} \mathbf{Pr}[N^{>\theta}(Y) \geq \ell] d\theta \geq \tau_\ell/2.$$

In the final inequality above we use Lemma 2.30(iii) with Bernoulli variables $B_i := \mathbb{1}_{Y^\downarrow_\ell > \theta}$ and $S = N^{>\theta}(Y)$. ∎

We remark that the constants in Theorem 3.18 are essentially tight.

Remark 3.1. *Consider a random vector Y that follows a product distribution on $\mathbb{R}^m_{\geq 0}$ where for each $i \in [m]$, Y_i is a Bernoulli random variable that takes size 1 with probability $1/m$. Observe that $\tau_1(Y) = 1$ and $\mathbf{E}[Y^\downarrow_1] = 1 - (1-1/m)^m$. Thus, the bound in Theorem 3.18(i) becomes tight as $m \to \infty$.*

Remark 3.2. *Fix some $\ell \in \{2, 3, \ldots, m\}$. Consider a random vector Y that follows a product distribution on $\mathbb{R}^m_{\geq 0}$ where for each $i \in [m]$, Y_i is a Bernoulli random variable that takes size 1 with probability ℓ/m. Observe that $\tau_\ell(Y) = 1$, and*

$$\mathbf{E}[Y^\downarrow_\ell] = 1 - \mathbf{Pr}[\mathsf{Bin}(m, \ell/m) \leq \ell - 1] \approx \mathbf{Pr}[\mathsf{Pois}(\ell) \geq \ell] \approx \frac{1}{2} + \frac{\ell^\ell e^{-\ell}}{\ell!}$$

By Stirling's approximation, for a sufficiently large ℓ, the last term is $\Theta(1/\sqrt{\ell})$. Thus, the constant $1/2$ in Theorem 3.18(ii) becomes tight as $\ell \to \infty$.

Chapter 4

Expected Norm of a Random Vector

In this chapter, we prove one of the central results of this thesis: for any random vector Y that follows a product distribution and any monotone symmetric norm f, we have $\mathbf{E}[f(Y)] = \Theta(f(\mathbf{E}[Y^{\downarrow}]))$. An immediate consequence of the above result is a stochastic generalization of the majorization inequality, which we call *approximate stochastic majorization*: if $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq \mathbf{E}[\mathsf{Top}_\ell(W)]$ holds for all $\ell \in [m]$, then $\mathbf{E}[f(Y)] \leq O(\mathbf{E}[f(W)])$, where W is another random vector that follows a product distribution.

Our framework for stochastic min-norm optimization is based on two key ideas. First, we use approximate stochastic majorization to reduce stochastic min-norm optimization to simultaneous stochastic Top_ℓ-norm optimization over all $\ell \in [m]$ that are powers of 2. Then, we use techniques from Chapter 3 for controlling $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ to handle the optimization problem for individual Top_ℓ norms.

We recall some frequently-used notation. We use Y to denote an arbitrary random vector that follows a product distribution on $\mathbb{R}^m_{\geq 0}$, and f to denote an arbitrary monotone, symmetric norm. We further assume that f is normalized, i.e., $f(1, 0, \ldots, 0) = 1$. For any $\theta \in \mathbb{R}_{\geq 0}$, the integer random variable $N^{>\theta}(Y)$ is defined to be $|\{i \in [m] : Y_i > \theta\}|$. Similarly, $N^{\geq\theta}(Y) := |\{i \in [m] : Y_i \geq \theta\}|$. We use $y^{\downarrow}$ as a shorthand for the vector $y \in \mathbb{R}^m_{\geq 0}$ with its coordinates sorted in non-increasing order. For an event A, we use $\mathbb{1}_A$ to denote its indicator random variable, i.e., $\mathbb{1}_A = 1$ if and only if event A happens. For any real number z, we define $z^+ := \max(z, 0)$.

4.1 Main Theorem: $\mathbf{E}[f(Y)] = \Theta(f(\mathbf{E}[Y^{\downarrow}]))$

Our main result in this section is the following.

Theorem 4.1.
Let Y follow a product distribution on $\mathbb{R}^m_{\geq 0}$, and $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotone, symmetric norm. Then,

$$f(\mathbf{E}[Y^{\downarrow}]) \leq \mathbf{E}[f(Y)] \leq 7.634 \cdot f(\mathbf{E}[Y^{\downarrow}]).$$

The lower bound on $\mathbf{E}[f(Y)]$ in the above theorem trivially follows from symmetry and convexity of f: $\mathbf{E}[f(Y)] = \mathbf{E}[f(Y^{\downarrow})] \geq f(\mathbf{E}[Y^{\downarrow}])$. The upper bound, however, requires some work and is based on an insightful application of Chernoff upper-tail bounds.

We make a few remarks on Theorem 4.1. Observe that the non-increasing vector $\mathbf{E}[Y^{\downarrow}] \in \mathbb{R}^m_{\geq 0}$ in the lower bound expression is arguably the most natural order statistic of Y: for any $i \in [m]$, the i^{th} coordinate of $\mathbf{E}[Y^{\downarrow}]$ is the expectation of the i^{th} maximum coordinate of Y. The significance of Theorem 4.1 is that the expectation and the norm can be exchanged with at most an $O(1)$-loss in approximation as long as the Y-vector is sorted before the exchange is performed; the constant that appears on the right hand side in Theorem 4.1 must be at least 1.214 (see Remark 4.1). Next, note that $\mathbf{E}[f(Y)] = f(\mathbf{E}[Y^{\downarrow}])$ holds for all Top_ℓ norms, and more generally, all ordered norms. This is because for any non-increasing $w \in \mathbb{R}^m_{\geq 0}$ we have:

$$\mathbf{E}[w^T Y^{\downarrow}] = \mathbf{E}\Big[\sum_{i\in[m]} w_i Y_i^{\downarrow}\Big] = \sum_{i\in[m]} w_i \mathbf{E}[Y_i^{\downarrow}] = w^T \mathbf{E}[Y^{\downarrow}].$$

Theorem 4.2. *For any ordered norm f, we have $\mathbf{E}[f(Y)] = f(\mathbf{E}[Y^{\downarrow}])$.*

So, the substance of Theorem 4.1 is that the above relationship holds approximately even when f is not an ordered norm. Recall the structural result of Chakrabarty and Swamy (Theorem 2.6): any monotone symmetric norm can be expressed as a supremum of ordered norms. Using this interpretation, Theorem 4.1 reads as follows: the expectation of a supremum of ordered norms is within a constant factor of the supremum of the expectation of ordered norms.

Theorem 4.3.
Let $\mathcal{C} \subseteq \mathbb{R}^m_{\geq 0}$ denote an arbitrary (possibly uncountable) collection of ordered norms. (Every $w \in \mathcal{C}$ is non-increasing). Then,

$$\sup_{w \in \mathcal{C}} w^T \mathbf{E}[Y^{\downarrow}] \leq \mathbf{E}\left[\sup_{w \in \mathcal{C}} w^T Y^{\downarrow}\right] \leq O(1) \cdot \sup_{w \in \mathcal{C}} w^T \mathbf{E}[Y^{\downarrow}].$$

One consequence of our proof approach is that $f(\gamma_1(Y), \tau_2(Y), \ldots, \tau_m(Y))$ serves as a nice proxy for $\mathbf{E}[f(Y)]$; see (3.2) and (3.3) for definitions.

Theorem 4.4.
Let Y follow a product distribution on $\mathbb{R}^m_{\geq 0}$, and $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotone, symmetric norm. Let $\gamma_1 = \gamma_1(Y)$, and $\tau_\ell = \tau_\ell(Y)$ for all $\ell \in [m]$. Then,

$$\frac{1}{2} \cdot f(\gamma_1, \tau_2, \tau_3, \ldots, \tau_m) \leq \mathbf{E}[f(Y)] \leq 4.026 \cdot f(\gamma_1, \tau_2, \tau_3, \ldots, \tau_m).$$

Theorem 4.4 is particularly useful in settings where the Y_i's are "atomic" random variables, and we have direct access to their distributions.

We remark that in our prior work [17, 18], we prove Theorem 4.1 by working with a slightly different lower bound on $\mathbf{E}[f(Y)]$, but the high-level proof strategy is essentially the same as the one we describe in the following section. Let $\Delta\gamma(Y) := (\gamma_\ell(Y) - \gamma_{\ell-1}(Y))_{\ell \in [m]} \in \mathbb{R}^m_{\geq 0}$. Note that $\Delta\gamma$ is nonincreasing (see Figure 3.1 and Lemma 3.17) and $\mathsf{Top}_\ell(\Delta\gamma) = \gamma_\ell(Y)$. We show that $\mathbf{E}[f(Y)] = \Theta(f(\Delta\gamma))$.

Theorem 4.5.
Let Y follow a product distribution on $\mathbb{R}^m_{\geq 0}$, and $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotone, symmetric norm. For any $\ell \in \{0, 1, \ldots, m\}$, let $\gamma_\ell = \gamma_\ell(Y)$; recall $\gamma_0 = 0$. Define the nonincreasing vector $\Delta\gamma \in \mathbb{R}^m_{\geq 0}$ as follows: for $\ell \in [m]$, $(\Delta\gamma)_\ell = \gamma_\ell - \gamma_{\ell-1}$. Then,

$$\frac{1}{2} \cdot f(\Delta\gamma) \leq \mathbf{E}[f(Y)] \leq 4.026 \cdot f(\Delta\gamma)$$

Proof Sketch. Before delving into the proofs of Theorem 4.1 and (the upper bound on $\mathbf{E}[f(Y)]$) Theorem 4.4, we describe the main ideas in our proof by making some mild assumptions about the probability distribution of the Y vector. Our assumptions are stated below.

(i) The τ_ℓ's are distinct.

(ii) For each $i \in [m]$, the support of Y_i lies in $\{\tau_1, \tau_2, \ldots, \tau_m\}$.

(iii) For each $\ell \in [m]$, we have $\sum_{i\in[m]} \mathbf{Pr}[Y_i = \tau_\ell] = 1$. That is, exactly one unit of probability mass (over all coordinates) resides at each τ_ℓ.

We remark that in the actual proof we group the ℓ's into buckets based on their τ_ℓ-value. So, the first assumption is that each bucket is a singleton. The second and third assumptions are used in the actual proof, and can be achieved by a simple rounding down strategy. Note that rounding down the support of Y_i's may only lead to a decrease in $\mathbf{E}[f(Y)]$. We will argue that this decrease can be bounded by $\gamma_1(Y) = O(\mathbf{E}[Y_1^\downarrow])$, a quantity that can be charged to the lower bound $f(\mathbf{E}[Y^\downarrow])$.

Assuming the above, it remains to bound $\mathbf{E}[f(Y)]$ when the total probability mass of Y_i's at each τ_ℓ is exactly 1. We work with the more amenable expression $f(\tau)$, where $\tau = (\tau_1, \ldots, \tau_m) \in \mathbb{R}^m_{\geq 0}$, which we argue is $O(f(\mathbf{E}[Y^\downarrow]))$ (see Theorem 4.6). (Note that $\gamma_1(Y) = \tau_1(Y)$ under assumption (ii).) Consider a random sample y drawn from the distribution of Y. How likely is it for $f(y)$ to exceed $\alpha f(\tau)$ for some scalar α? This probability can be bounded by using Lemma 2.14: if $f(y) > \alpha f(\tau)$, then there exists a $k \in \{1, \ldots, \lceil m/\alpha \rceil\}$ such that $y^\downarrow_{\alpha(k-1)+1} > \tau_k$. Since we assumed that the τ_ℓ's are distinct, we get that $y^\downarrow_{\alpha(k-1)+1} \geq \tau_{k-1}$. In other words, if the event $\{f(Y) > \alpha f(\tau)\}$ happens, then for some $1 \leq k \leq \lceil m/\alpha \rceil$, the event $\{N^{\geq \tau_{k-1}}(Y) > \alpha(k-1)\}$ happens, where $N^{\geq \theta}(Y)$ denotes the (random) number of coordinates of Y that are at least θ. Observe that $N^{\geq \tau_{k-1}}(Y) = \sum_{i\in[m]} \mathbb{1}_{Y_i \geq \tau_{k-1}}$ is a sum of independent Bernoulli random variables, and by the third assumption we have $\mathbf{E}[N^{\geq \tau_{k-1}}(Y)] = k-1$. So, by Chernoff bounds, the event $\{N^{\geq \tau_{k-1}}(Y) > \alpha(k-1)\}$ happens with probability at most $\mathsf{exp}(-\Theta(\alpha k))$. A simple union bound argument over all k shows that the event $\{f(Y) > \alpha f(\tau)\}$ happens with probability at most $\mathsf{exp}(-\Theta(\alpha))$. A straightforward integration calculation gives $\mathbf{E}[f(Y)] = O(f(\tau))$. We formalize this strategy in Section 4.1.1.

We remark that [18] has a different proof of Theorem 4.1 that does not proceed via these modifications, but at the core uses the same idea that using Chernoff bounds, one can argue that $N^{>\tau_\ell}(Y)$ is concentrated around its expectation.

4.1.1 Proof of Theorems 4.1 and 4.4

We now delve into the proof details. We already showed that $\mathbf{E}[f(Y)] = \mathbf{E}[f(Y^\downarrow)] \geq f(\mathbf{E}[Y^\downarrow])$ follows from symmetry and convexity of f. The remainder of this section is devoted to proving the upper bounds on $\mathbf{E}[f(Y)]$ in Theorems 4.1 and 4.4. The following result is useful to us.

Theorem 4.6. *Let Y follow a product distribution on $\mathbb{R}^m_{\geq 0}$, and $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotone, symmetric norm. Let $\gamma_1 = \gamma_1(Y)$, and $\tau_\ell = \tau_\ell(Y)$ for $\ell \in [m]$. We have:*

$$\frac{1}{2} \cdot f(\gamma_1, \tau_2, \tau_3, \ldots, \tau_m) \leq f(\mathbf{E}[Y^{\downarrow}]).$$

Proof. By Theorem 3.18, we have $\mathbf{E}[Y^{\downarrow}_1] \geq \gamma_1/2$, and for any $\ell \in \{2, 3, \ldots, m\}$ we have $\mathbf{E}[Y^{\downarrow}_\ell] \geq \tau_\ell/2$. Since f is homogeneous and monotone, we get the desired result. ∎

As described in the proof sketch, we prove Theorems 4.1 and 4.4 in two parts. First, we modify the distribution of Y_i's so that assumptions (i), (ii) and (iii) hold in a somewhat weak form. Lemmas 4.7 and 4.9 show that the modification "costs" at most $\gamma_1(Y)$. Then, through Lemmas 4.12 and 4.13, we show that the f-norm of the modified Y-vector enjoys strong concentration properties around $f(\tau)$. Overall, we get $\mathbf{E}[f(Y)] = O(\gamma_1 + f(\tau)) = O(f(\mathbf{E}[Y^{\downarrow}]))$.

We introduce some notation to refer to distinct $\tau_\ell(Y)$'s. For notational convenience, define $\tau_0 := \infty$ and $\ell_0 := 0$. For any integer $j \geq 1$, we iteratively define ℓ_j as follows: ℓ_j is the largest index in $[m]$ satisfying $\tau_{\ell_j} = \tau_{\ell_{j-1}+1}$. We only define ℓ_j's as long as $\ell_j \leq m$, and let $m' \geq 1$ denote the largest integer satisfying $\ell_{m'} = m$. By definition, the number of distinct values in $\{\tau_1, \ldots, \tau_m\}$ is m', and for each $j \in [m']$, there are $\ell_j - \ell_{j-1}$ distinct indices $\ell \in [m]$ such that $\tau_\ell = \tau_{\ell_j}$. For instance, if $\tau = (5, 5, 4.5, 2, 2, 2, 0, 0) \in \mathbb{R}^8_{\geq 0}$, then $\ell_1 = 2$, $\ell_2 = 3$, $\ell_3 = 6$, $\ell_4 = 8$, and $m' = 4$.

At a high level, our modifications involve moving a suitable amount of probability mass (of the Y_i's) such that $\sum_{i \in [m]} \mathbf{Pr}[Y_i = \tau_{\ell_j}] = \ell_j - \ell_{j-1}$ holds for each $j \in [m']$. Since $\sum_{j \in [m']} (\ell_j - \ell_{j-1}) = \ell_{m'} - \ell_0 = m$, our modifications will ensure that the Y_i's are supported on $\{\tau_{\ell_j}\}_{j \in [m']}$. In other words, assumptions (i), (ii) and (iii) from the proof sketch that we gave earlier hold in a weak form. We also remark that our modifications do not change any of the τ_ℓ statistics, so with some abuse of notation we use τ_ℓ without explicitly stating the argument. For notational convenience, we use $\widetilde{Y}$ to denote the vector Y that is obtained after the modifications are applied.

Our modifications happen over m' iterations. In each iteration $j \in [m']$, we move at most a single unit of probability mass of the Y_i's that lies in the interval $[\tau_{\ell_j}, \tau_{\ell_{j-1}}]$ down to a single point τ_{ℓ_j}. In iteration $j = 1$, this corresponds to bounding the Y_i's from above: for $i \in [m]$, define $\widetilde{Y}_i := \min(Y_i, \tau_1)$. The following result is immediate.

Lemma 4.7. *At the end of the first modification step, we have:*

$$\mathbf{E}[f(\widetilde{Y})] \leq \mathbf{E}[f(Y)] \leq \mathbf{E}[f(\widetilde{Y})] + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_1)^+].$$

Proof. The lower bound on $\mathbf{E}[f(Y)]$ is trivial because f is monotone. For the upper bound, observe that $Y_i = \widetilde{Y}_i + (Y_i - \tau_1)^+$. Therefore, by triangle inequality and Lemma 2.2 we get:

$$\mathbf{E}[f(Y)] \leq \mathbf{E}[f(\widetilde{Y})] + \mathbf{E}[f\big((Y_1 - \tau_1)^+, \dots, (Y_m - \tau_1)^+\big)] \leq \mathbf{E}[f(\widetilde{Y})] + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_1)^+]. \quad \blacksquare$$

The following claim will be useful in describing the rounding strategy for iterations 2 through m'.

Claim 4.8. *Fix some $k \in [m'-1]$. Suppose that $\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i = \tau_{\ell_j}] = \ell_j - \ell_{j-1}$ and $\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i \in (\tau_{\ell_{j+1}}, \tau_{\ell_j})] = 0$ hold for all $j \in \{1, \dots, k-1\}$. The following are true at the beginning of iteration $k+1$:*

(i) $\sum_{i\in[m]} \mathbf{Pr}\left[\widetilde{Y}_i \in (\tau_{\ell_k+1}, \tau_{\ell_k}]\right] \leq \ell_k - \ell_{k-1} + 1$.

(ii) $\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i = \tau_{\ell_k}] \geq \ell_k - \ell_{k-1}$.

Proof. First, note that $\tau_{\ell_{k+1}} = \tau_{\ell_k+1} < \tau_{\ell_k}$. By Lemma 3.10, we get that $\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i > \tau_{\ell_k+1}] < \ell_k + 1$ and $\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i \geq \tau_{\ell_k}] \geq \ell_k$. Both conclusions of the claim easily follow for the $k=1$ case because $\ell_0 = 0$ and the $\widetilde{Y}_i$'s are capped at $\tau_{\ell_1} = \tau_1$. Now suppose that $k > 1$. By the assumptions in the claim, we have:

$$\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i > \tau_{\ell_k}] = \sum_{j=1}^{k-1} \Big(\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i = \tau_{\ell_j}] \Big) = \sum_{j=1}^{k-1} (\ell_j - \ell_{j-1}) = \ell_{k-1}.$$

Both conclusions of this claim follow from a simple calculation:

$$\sum_{i\in[m]} \mathbf{Pr}\left[\widetilde{Y}_i \in (\tau_{\ell_k+1}, \tau_{\ell_k}]\right] = \sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i > \tau_{\ell_k+1}] - \sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i > \tau_{\ell_k}] < (\ell_k + 1) - \ell_{k-1},$$

and

$$\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i = \tau_{\ell_k}] = \sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i \geq \tau_{\ell_k}] - \sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i > \tau_{\ell_k}] \geq \ell_k - \ell_{k-1}. \quad \blacksquare$$

We now formally describe the modifications that we make in iterations 2 through m'. Fix some $k \in [m'-1]$. Suppose that we are at the beginning of iteration $k+1$ and the hypothesis of Claim 4.8 holds. We arbitrarily move a probability mass of $\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i = \tau_{\ell_k}] - (\ell_k - \ell_{k-1}) \in [0,1)$ from τ_{ℓ_k} down to $\tau_{\ell_{k+1}}$. We also move every point in the support of $\widetilde{Y}_i$'s that lies in the interval $(\tau_{\ell_k+1}, \tau_{\ell_k})$ down to the point $\tau_{\ell_k+1} (= \tau_{\ell_{k+1}})$. By Claim 4.8, the above rounding procedure is well-defined and the total amount of probability mass that is moved from $(\tau_{\ell_k+1}, \tau_{\ell_k}]$ down to τ_{ℓ_k+1} is at most 1. It is easy to see that by the end of iteration $k+1$, we have ensured that the hypothesis of Claim 4.8 holds for the start of iteration $k+2$ (assuming $k+2 \leq m'$). Since the hypothesis of Claim 4.8 holds vacuously for the $k=1$ case, by the end of m' iterations we have that $\sum_{i\in[m]} \mathbf{Pr}[Y_i = \tau_{\ell_j}] = \ell_j - \ell_{j-1}$ holds for all $j \in [m']$. We again remark that our modifications do not alter the τ_ℓ statistics. This is because for any $j \in [m']$, $\sum_{i\in[m]} \mathbf{Pr}[\widetilde{Y}_i \geq \tau_{\ell_j}] = \ell_j$, so for any $\ell \in \{\ell_{j-1}+1, \ldots, \ell_j\}$, we have $\tau_\ell(\widetilde{Y}) = \tau_{\ell_j}(\widetilde{Y}) = \tau_\ell(Y)$.

For notational clarity, let $\widetilde{Y}^{(k)}$ denote the $\widetilde{Y}$-vector after k iterations of modifications have been applied.

Lemma 4.9. *At the end of iteration $k \in \{2, 3, \ldots, m'\}$, we have:*

$$\mathbf{E}[f(\widetilde{Y}^{(k)})] \leq \mathbf{E}[f(\widetilde{Y}^{(k-1)})] \leq \mathbf{E}[f(\widetilde{Y}^{(k)})] + (\tau_{\ell_{k-1}} - \tau_{\ell_k}).$$

Proof. The lower bound on $\mathbf{E}[f(\widetilde{Y}^{(k-1)})]$ is trivial because f is monotone. By the design of our modification step for iteration k, the probability mass (of $\widetilde{Y}_i$'s) that is moved from the interval $(\tau_{\ell_k}, \tau_{\ell_{k-1}}]$ to τ_{ℓ_k} is at most 1. Thus, the difference in expected f-norm of the $\widetilde{Y}$-vector before and after iteration k is bounded by $\tau_{\ell_{k-1}} - \tau_{\ell_k}$. Formally,

$$\mathbf{E}[f(\widetilde{Y}^{(k-1)})] \leq \mathbf{E}[f(\widetilde{Y}^{(k)})] + \sum_{i\in[m]} \mathbf{E}\left[\widetilde{Y}^{(k-1)} - \widetilde{Y}^{(k)}\right] \leq \mathbf{E}[f(\widetilde{Y}^{(k)})] + (\tau_{\ell_{k-1}} - \tau_{\ell_k}). \qquad \blacksquare$$

Let $\overline{Y} = \widetilde{Y}^{(m')}$. Combining Lemmas 4.7 and 4.9 gives the following.

Lemma 4.10. *The following are true:*

(i) $\mathbf{E}[f(\overline{Y})] \leq \mathbf{E}[f(Y)] \leq \mathbf{E}[f(\overline{Y})] + \gamma_1(Y)$.

(ii) *For each $j \in [m']$, we have $\mathbf{E}[N^{\geq \tau_{\ell_j}}(\overline{Y})] = \ell_j$.*

Proof. Follows from Lemmas 4.7, 4.9, and the description of our modification steps. We also use $\sum_{k=2}^{m'} (\tau_{\ell_{k-1}} - \tau_{\ell_k}) + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_1)^+] \leq (\tau_1 - \tau_m) + \sum_{i\in[m]} \mathbf{E}[(Y_i - \tau_1)^+] \leq \gamma_1(Y)$. $\blacksquare$

We now bound the expected f-norm of the well-structured $\overline{Y}$-vector.

Theorem 4.11. *We have* $\mathbf{E}[f(\overline{Y})] \leq 3.026 \cdot f(\tau)$.

The following lemmas will be useful in the proof of Theorem 4.11 and 4.1.

Lemma 4.12. *Let* $\alpha \in \mathbb{R}_{\geq 0}$ *be a scalar and* $j \in [m']$ *be a positive integer. Consider the random variable* $N^{\geq \tau_{\ell_j}}(\overline{Y}) = \sum_{i \in [m]} \mathbb{1}_{\overline{Y}_i \geq \tau_{\ell_j}}$ *that counts the number of coordinates of* $\overline{Y}$ *that are at least* τ_{ℓ_j}. *The following tail bound holds:*

$$\mathbf{Pr}\left[N^{\geq \tau_{\ell_j}}(\overline{Y}) > (1+\alpha)\ell_j\right] \leq \left(\frac{e^{\alpha}}{(1+\alpha)^{1+\alpha}}\right)^{\ell_j}.$$

Proof. By Lemma 4.10(ii), we have $\mathbf{E}[N^{\geq \tau_{\ell_j}}(\overline{Y})] = \ell_j$, so the tail bound follows from a direct application of Chernoff bounds (see Lemma 2.28). ■

Lemma 4.13. *Let* $\alpha \in \mathbb{R}_{\geq 0}$ *be a scalar. The following tail bound holds for the scalar random variable* $f(\overline{Y})$*:*

$$\mathbf{Pr}\left[f(\overline{Y}) > (1+\alpha) f(\tau)\right] \leq \frac{e^{\alpha}}{(1+\alpha)^{1+\alpha} - e^{\alpha}}.$$

Proof. Suppose that the event $\{f(\overline{Y}) > (1+\alpha)f(\tau)\}$ happens. Then, by Lemma 2.14, there exists $k \in \{1, \ldots, \lceil m/(1+\alpha) \rceil\}$ such that the event $\{\overline{Y}^{\downarrow}_{\lfloor (1+\alpha)(k-1) \rfloor + 1} > \tau_k\}$ happens. The above implication can be simplified by using distributional assumptions of $\overline{Y}$. Let $j \in [m']$ denote the largest index such that $\tau_{\ell_j} > \tau_k$ holds. Since τ is a non-increasing vector, we clearly have $\ell_j \leq k-1$. Next, recall that for any $i \in [m]$ the support of $\overline{Y}_i$ lies in $\{\tau_{\ell_j}\}_{j \in [m']}$. Thus, if the event $\{\overline{Y}^{\downarrow}_{\lfloor (1+\alpha)(k-1) \rfloor + 1} > \tau_k\}$ happens, then the event $\{\overline{Y}^{\downarrow}_{\lfloor (1+\alpha)\ell_j \rfloor + 1} \geq \tau_{\ell_j}\}$ happens. Observe that the latter event is the same as the event $\{N^{\geq \tau_{\ell_j}}(\overline{Y}) > (1+\alpha)\ell_j\}$.

The tail bound on $f(Y)$ follows from a simple union bound argument over all $j \in [m']$:

$$\begin{aligned}
\mathbf{Pr}[f(\overline{Y}) > (1+\alpha)f(\tau)] &\leq \sum_{j\in[m']} \mathbf{Pr}[N^{\geq \tau_{\ell_j}}(\overline{Y}) > (1+\alpha)\ell_j] \\
&\leq \sum_{j\in[m']} \left(\frac{e^\alpha}{(1+\alpha)^{1+\alpha}}\right)^{\ell_j} && \text{(by Lemma 4.12)} \\
&\leq \sum_{\ell=1}^{\infty} \left(\frac{e^\alpha}{(1+\alpha)^{1+\alpha}}\right)^{\ell} && (1 \leq \ell_1 < \ell_2 < \cdots < \ell_{m'} = m) \\
&= \frac{e^\alpha}{(1+\alpha)^{1+\alpha} - e^\alpha} && \blacksquare
\end{aligned}$$

We remark that the tail bound in Lemma 4.13 is nontrivial when $(1+\alpha)^{1+\alpha} > 2e^\alpha$, which is roughly $\alpha \geq 1.391$. We now prove Theorem 4.11, and subsequently Theorems 4.1 and 4.4.

Proof of Theorem 4.11. We use a simple integration calculation:

$$\begin{aligned}
\mathbf{E}[f(\overline{Y})] &= f(\tau) \cdot \int_0^\infty \mathbf{Pr}\left[f(\overline{Y}) > \alpha f(\tau)\right] d\alpha \\
&\leq f(\tau) \cdot \left(3 + \int_3^\infty \mathbf{Pr}\left[f(\overline{Y}) > \alpha f(\tau)\right] d\alpha\right) \\
&\leq f(\tau) \cdot \left(3 + \int_2^\infty \frac{e^\alpha}{(1+\alpha)^{1+\alpha} - e^\alpha} d\alpha\right) && \text{(By Lemma 4.13)} \\
&\leq f(\tau) \cdot \left(3 + \sum_{n=2}^{\infty} \frac{e^n}{(1+n)^{1+n} - e^n}\right) && \text{(integrand is a decreasing function)} \\
&\leq f(\tau) \cdot \left(3 + \sum_{n=2}^{\infty} 2^{-(n-1)}\right) = 4\, f(\tau).
\end{aligned}$$

It remains to justify that $(1+n)^{1+n} \geq (2^{n-1}+1)e^n$ holds for any integer $n \geq 2$. The inequality is true for $n = 2, 3, 4$ by an explicit calculation. For $n \geq 4$, we have:

$$(2+n)^{2+n} \geq (2+n) \cdot (1+n)^{1+n} \geq (2e) \cdot (2^{n-1}+1) \cdot e^n \geq (2^n+1) \cdot e^{n+1},$$

so the desired inequality follows by induction.

We can obtain a tighter upper bound by a computer-assisted numerical computation of the integral:

$$\mathbf{E}[f(\overline{Y})] \leq f(\tau) \cdot \left(2.391 + \int_{1.391}^\infty \frac{e^\alpha}{(1+\alpha)^{1+\alpha} - e^\alpha} d\alpha\right) \leq (2.391 + 0.635) f(\tau) = 3.026 \cdot f(\tau)$$

$\blacksquare$

Proof of Theorem 4.1. Recall that f is normalized so for any $y \in \mathbb{R}^m_{\geq 0}$ we have $\mathsf{Top}_1(y) \leq f(y)$. By Lemma 4.10(i) and Theorem 4.11:

$$\mathbf{E}[f(Y)] \leq \mathbf{E}[f(\overline{Y})] + \gamma_1(Y) \leq 3.026 \cdot f(\tau) + \gamma_1(Y). \tag{4.1}$$

Theorem 4.6 and monotonicity of f gives $f(\tau) \leq f(\gamma_1, \tau_2, \ldots, \tau_m) \leq 2\, f(\mathbf{E}[Y^{\downarrow}])$, and Theorem 3.18 gives $\gamma_1(Y) \leq \frac{e}{e-1} \cdot \mathbf{E}[Y_1^{\downarrow}] \leq \frac{e}{e-1} \cdot f(\mathbf{E}[Y^{\downarrow}])$. Combining the above, we get

$$\mathbf{E}[f(Y)] \leq 7.634 \cdot f(\mathbf{E}[Y^{\downarrow}]).$$ ∎

Proof of Theorem 4.4. We use Theorem 4.6 and monotone property of f to upper bound the right hand side in (4.1):

$$\frac{1}{2} \cdot f(\gamma_1, \tau_2, \ldots, \tau_m) \leq \mathbf{E}[f(Y)] \leq 3.026 \cdot f(\tau) + \gamma_1(Y) \leq 4.026 \cdot f(\gamma_1, \tau_2, \ldots, \tau_m).$$ ∎

Given Theorems 4.1 and 4.4, the proof of Theorem 4.5 is straightforward.

Proof of Theorem 4.5. Recall that the nonincreasing vector $\Delta\gamma \in \mathbb{R}^m_{\geq 0}$ is defined as follows: for $\ell \in [m]$, $\Delta\gamma_\ell = \gamma_\ell - \gamma_{\ell-1}$, where $\gamma_\ell = \gamma_\ell(Y)$. By definition, $\mathsf{Top}_\ell(\Delta\gamma) = \gamma_\ell$, so by Theorem 3.9 we have that $\mathsf{Top}_\ell(\Delta\gamma) \leq 2\,\mathsf{Top}_\ell(\mathbf{E}[Y^{\downarrow}])$ for all $\ell \in [m]$. Therefore, by the majorization inequality (Theorem 2.7) and Theorem 4.1, we get

$$\mathbf{E}[f(Y)] \geq f(\mathbf{E}[Y^{\downarrow}]) \geq \frac{1}{2} \cdot f(\Delta\gamma).$$

Next, we prove the upper bound on $\mathbf{E}[f(Y)]$. In Theorem 4.4, we showed $\mathbf{E}[f(Y)] \leq 4.026 \cdot f(\gamma_1, \tau_2, \ldots, \tau_m)$. Observe that $(\Delta\gamma)_1 = \gamma_1$, and for any $\ell \in \{2, 3, \ldots, m\}$ we have $(\Delta\gamma)_\ell \geq \tau_\ell$ by Lemma 3.17. Since f is monotone, we get:

$$\mathbf{E}[f(Y)] \leq 4.026 \cdot f(\gamma_1, \tau_2, \ldots, \tau_m) \leq 4.026 \cdot f(\Delta\gamma).$$ ∎

We conclude this section by showing that the worst-case multiplicative gap between $\mathbf{E}[f(Y)]$ and $f(\mathbf{E}[Y^{\downarrow}])$ is bounded away from 1.

Remark 4.1. *Consider a random vector Y that follows a product distribution on $\mathbb{R}^m_{\geq 0}$ where for each $i \in [m]$, Y_i is a Bernoulli trial that takes value 1 with probability $1/m$ (and value 0 with remaining probability). Consider the monotone symmetric norm $f : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ that* *maps* $x \in \mathbb{R}^m_{\geq 0}$ *to* $\max\left(\frac{e}{e-1} \cdot \mathsf{Top}_1(x), \mathsf{Top}_m(x)\right)$. *Observe that* $\mathbf{E}[\mathsf{Top}_m(Y)] = m \cdot \frac{1}{m} = 1$, *and* $\mathbf{E}[\mathsf{Top}_1(Y)] = 1 - (1 - \frac{1}{m})^m \approx 1 - \frac{1}{e}$. *Assuming* $m \to \infty$, *we have:*

$$f(\mathbf{E}[Y^{\downarrow}]) = \max\Big(\frac{e}{e-1} \cdot \mathbf{E}[\mathsf{Top}_1(Y)], \mathbf{E}[\mathsf{Top}_m(Y)]\Big) \to 1.$$

and

$$\begin{aligned}\mathbf{E}[f(Y)] &= \mathbf{E}[\mathsf{Top}_m(Y)] + \mathbf{E}\Big[\Big(\frac{e}{e-1} \cdot \mathsf{Top}_1(Y) - \mathsf{Top}_m(Y)\Big)^+\Big] \\ &= 1 + \frac{1}{e-1} \cdot \mathbf{Pr}[Y_1^{\downarrow} = 1 \text{ and } Y_2^{\downarrow} = 0] \\ &= 1 + \frac{1}{e-1} \cdot \binom{m}{1} \cdot \Big(\frac{1}{m}\Big) \cdot \Big(1 - \frac{1}{m}\Big)^{m-1} \to 1 + \frac{1}{e \cdot (e-1)}\end{aligned}$$

Therefore, the constant in the right hand side of Theorem 4.1 must be at least $1 + \frac{1}{e\cdot(e-1)} \approx 1.214$.

4.2 Approximate Stochastic Majorization

In this section we show that the majorization inequality extends to the stochastic setting in an approximate form. We call this result *Approximate Stochastic Majorization* and it is a direct consequence of Theorems 2.7 and 4.1.

Theorem 4.14 (Approximate Stochastic Majorization)**.**
Let Y, W *follow product distributions on* $\mathbb{R}^m_{\geq 0}$, *and* $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ *be a monotone, symmetric norm. Suppose that* $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq \mathbf{E}[\mathsf{Top}_\ell(W)]$ *holds for all* $\ell \in [m]$. *Then,*

$$\mathbf{E}[f(Y)] \leq 7.634 \cdot \mathbf{E}[f(W)].$$

Proof. Recall from Theorem 4.2 that $\mathbf{E}[\mathsf{Top}_\ell(Y)] = \mathsf{Top}_\ell(\mathbf{E}[Y^{\downarrow}])$ for any $\ell \in [m]$. We have:

$$\mathbf{E}[f(Y)] \leq 7.634 \cdot f(\mathbf{E}[Y^{\downarrow}]) \leq 7.634 \cdot f(\mathbf{E}[W^{\downarrow}]) \leq 7.634 \cdot \mathbf{E}[f(W)].$$

In the above, the first and third inequalities follow from Theorem 4.1, and the second inequality follows from Theorem 2.7 applied to the vectors $\mathbf{E}[Y^{\downarrow}]$ and $\mathbf{E}[W^{\downarrow}]$. ■

4.3 Other Stochastic Majorization Inequalities

In this section we give three versions of approximate stochastic majorization that are easier to apply in other problem settings. Recall that $\mathsf{POS}_m := \{1, 2, 4, \ldots, 2^{\lfloor \log_2 m \rfloor}\}$.

In the following theorem, we give a *relative* and *budgeted* version of approximate stochastic majorization. In the relative version, we show that bounding $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ in terms of $\mathbf{E}[\mathsf{Top}_\ell(W)]$ for all $\ell \in \mathsf{POS}$ implies a bound on $\mathbf{E}[f(Y)]$ in terms of $\mathbf{E}[f(W)]$. In the budgeted version, the bounds on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ are given by some scalars B_ℓ, and we infer a bound on $\mathbf{E}[f(Y)]$ in terms of $f(\vec{b})$, where $\vec{b}$ is such that $\mathsf{Top}_\ell(\vec{b})$ is roughly B_ℓ.

Theorem 4.15.
Let Y follow a product distribution on $\mathbb{R}^m_{\geq 0}$, and $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotone, symmetric norm. The following are true:

(i) ***(Relative Version)*** *Let W follow a product distribution on $\mathbb{R}^m_{\geq 0}$, and $\alpha \in \mathbb{R}_{>0}$ be a positive scalar. Suppose that $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq \alpha \mathbf{E}[\mathsf{Top}_\ell(W)]$ holds for all $\ell \in \mathsf{POS}_m$. Then we have:*

$$\mathbf{E}[f(Y)] \leq 2\alpha \cdot 7.634 \cdot \mathbf{E}[f(W)] = O(\alpha) \cdot \mathbf{E}[f(W)].$$

(ii) ***[Budgeted Version]*** *Let $(B_\ell)_{\ell \in \mathsf{POS}_m}$ denote a non-decreasing, nonnegative sequence, and $b : [0, m] \to \mathbb{R}_{\geq 0}$ denote the corresponding upper envelope curve (see Definition 2.16). Let $\vec{b} \in \mathbb{R}^m_{\geq 0}$ be defined as follows: for any $i \in [m]$, $\vec{b}_i := b(i) - b(i-1)$. Suppose that $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq B_\ell$ holds for all $\ell \in \mathsf{POS}_m$. Then we have:*

$$\mathbf{E}[f(Y)] \leq 2 \cdot 7.634 \cdot f(\vec{b}) = O(f(\vec{b}))$$

Proof. The first part follows from Theorem 4.14 applied to product distributions Y and $W' := 2\alpha W$, where for any $i \in [m]$, W'_i is the random variable W_i scaled by 2α. The factor 2 is to account for Top_ℓ norms with $\ell \notin \mathsf{POS}_m$.

The second part follows from Theorems 2.15 and 4.1. We have $\mathbf{E}[f(Y)] \leq 7.634 \cdot f(\mathbf{E}[Y^\downarrow])$ by Theorem 4.1. By our assumption $\mathsf{Top}_\ell(\mathbf{E}[Y^\downarrow]) = \mathbf{E}[\mathsf{Top}_\ell(Y)] \leq B_\ell$, so by Theorem 2.15 we have $f(\mathbf{E}[Y^\downarrow]) \leq 2\, f(\vec{b})$. ■

The following result is a stochastic generalization of the results that we saw in Section 2.2.1. Roughly speaking, we show that if the τ_ℓ statistics of Y are comparable to the

τ_ℓ statistics of W, then the expected f-norm of Y can be bounded in terms of the expected f-norm of W.

Theorem 4.16.
Let Y, W follow product distributions on $\mathbb{R}^m_{\geq 0}$, and $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotone, symmetric norm. Let $\lambda, B_1 \in \mathbb{R}_{\geq 0}$ denote scalars such that $\sum_{i\in[m]} \mathbf{E}[(Y_i - \lambda\tau_1(W))^+] \leq B_1$. Let $\alpha \in \mathbb{R}_{\geq 1}, \beta \in \mathbb{R}_{>0}$ denote scalars such that for every $\ell \in \mathsf{POS}_m, \alpha\ell \leq m$, the condition $\tau_{\lceil \alpha\ell \rceil}(Y) \leq \beta\tau_\ell(W)$ holds. We have:

$$\mathbf{E}[f(Y)] \leq 4.026 \cdot \{\alpha\lambda\tau_1(W) + B_1 + 2\alpha\beta f(\tau(W))\}.$$

By Theorem 4.4, we have $\mathbf{E}[f(Y)] \leq O(1) \cdot f(\gamma_1(Y), \tau_2(Y), \ldots, \tau_m(Y))$, so it remains to prove a bound on $f(\gamma_1(Y), \tau_2(Y), \ldots, \tau_m(Y))$. We do this by using triangle inequality: Lemma 4.17 handles the contribution from the first $\max(1, \lceil\alpha\rceil - 1)$ coordinates, and Lemma 4.18 takes care of the remaining coordinates.

Lemma 4.17. $f(\gamma_1(Y), \tau_2(Y), \ldots, \tau_{\lceil\alpha\rceil - 1}(Y), 0, \ldots, 0) \leq \alpha\lambda\tau_1(W) + B_1.$

Proof. We consider two cases. Suppose that $\alpha \leq 2$ holds. Taking $\ell = 1$ and $\theta = \lambda\tau_1(W)$ in Lemma 3.11(i) gives:

$$\gamma_1(Y) \leq \lambda\tau_1(W) + \sum_{i\in[m]} \mathbf{E}\big[\big(Y_i - \lambda\tau_1(W)\big)^+\big] \leq \alpha\lambda\tau_1(W) + B_1.$$

On the other hand, if $\alpha > 2$ holds, we first use Lemma 3.17 to obtain $\gamma_1(Y) + \tau_2(Y) + \cdots + \tau_{\lceil\alpha\rceil-1}(Y) \leq \gamma_{\lceil\alpha\rceil-1}(Y)$. Then, we again use Lemma 3.11 with $\theta = \lambda\tau_1(W)$ to obtain

$$\gamma_{\lceil\alpha\rceil-1}(Y) \leq (\lceil\alpha\rceil - 1) \cdot \lambda\tau_1(W) + \sum_{i\in[m]} \mathbf{E}\big[\big(Y_i - \lambda\tau_1(W)\big)^+\big] \leq \alpha\lambda\tau_1(W) + B_1. \qquad \blacksquare$$

Lemma 4.18. $f(\tau_{\lceil\alpha\rceil}(Y), \tau_{\lceil\alpha\rceil+1}(Y), \ldots, \tau_m(Y), 0, \ldots, 0) \leq 2\alpha\beta \cdot f(\tau(W))$

Proof. We use the τ_ℓ-based majorization inequality from Lemma 2.13 for this proof. Consider nonincreasing vectors $x, y \in \mathbb{R}^m_{\geq 0}$ defined as follows: for any $i \in [m]$, $x_i := \beta\tau_\ell(W)$ and $y_i := \tau_{\lceil\alpha\rceil-1+i}(Y)$. Note that $y_i = 0$ for indices $i > m - \lceil\alpha\rceil + 1$. We show that the condition in Lemma 2.13 holds for the scaling parameter 2α, thereby implying $f(y) \leq 2\alpha f(x)$.

By our assumption, for any $\ell \in \mathsf{POS}_m, \alpha\ell \leq m$ we have $y_{\lceil\alpha\ell\rceil - \lceil\alpha\rceil + 1} \leq \beta\tau_\ell(W)$. Fix one such ℓ. Observe that for any $k \in \{\lceil\alpha\ell\rceil - \lceil\alpha\rceil + 1, \ldots, \max(m, \lceil 2\alpha\ell\rceil - \lceil\alpha\rceil)\}$, we have:

$$y_k \leq y_{\lceil\alpha\ell\rceil - \lceil\alpha\rceil + 1} \leq x_\ell \leq x_{\lceil \frac{\lceil 2\alpha\ell\rceil - \lceil\alpha\rceil}{2\alpha} \rceil} \leq x_{\lceil \frac{k}{2\alpha} \rceil}. \qquad \blacksquare$$

Proof of Theorem 4.16. Follows from Lemmas 4.17, 4.18 and triangle inequality. $\blacksquare$

Chapter 5

Stochastic Minimum Norm Load Balancing

We now apply our framework to devise approximation algorithms for *stochastic minimum norm load balancing*. In this introductory chapter on stochastic min-norm load balancing, we define the problem, state our main results, and develop some basic tools that will be used in obtaining our results. In Section 5.1, we define the problem and list some important special cases. We also describe our main results here; Chapters 6-9 are devoted to deriving these results. In Section 5.2 we derive some results for handling the expectation of exceptional random variables that arise in load balancing applications; These results will be used in Chapters 6, 7 and 8.

5.1 Problem Statement

In an instance of stochastic min-norm load balancing ($\mathsf{StochNormLB}$), we are given n *stochastic* jobs that need to be processed on exactly one of m unrelated machines. Throughout, we use J and $[m]$ to denote the set of jobs and machines respectively; we use j to index jobs, and i to index machines. For each job $j \in J$ and machine $i \in [m]$, we are given a nonnegative random variable X_{ij} that denotes the processing time of job j on machine i. Jobs are independent, so X_{ij} and $X_{i'j'}$ are independent whenever $j \neq j'$; however, X_{ij} and $X_{i'j}$ could be correlated. A feasible solution is an assignment $\sigma : J \to [m]$ of jobs to machines. This induces a random load vector $\overrightarrow{\mathsf{load}}^{\sigma}$ where $\overrightarrow{\mathsf{load}}^{\sigma}(i) := \sum_{j:\sigma(j)=i} X_{ij}$ for each $i \in [m]$; note that $\overrightarrow{\mathsf{load}}^{\sigma}$ follows a product distribution on $\mathbb{R}^m_{\geq 0}$. The goal is to find an assignment σ that minimizes $\mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma})]$ for a given monotone, symmetric norm f. Here, the expectation is over the randomness in $\{X_{ij}\}_{i,j}$.

We make a few remarks on how the input is given to us. While the most natural input specification for the job-size distributions is that we are given the joint distribution of $(X_{ij})_{i\in[m]}$, all our algorithms only need the *marginal* distribution of X_{ij} for each $i \in [m]$ and job j (and in fact only require access to a certain statistic of these marginal distributions). Our algorithms only require a value oracle for the norm f.

5.1.1 Important Special Cases

There are three sources of generality in StochNormLB: the generality of monotone symmetric norms, the generality of the unrelated-machines environment, and the generality of job-size distributions. Limiting the level of generality in each of these leads to the following important special cases:

1. **Top_ℓ Norms:** The norm f is a Top_ℓ norm for some $\ell \in [m]$. So, the goal in this problem is to find an assignment of jobs to machines that minimizes the expected sum of the ℓ largest machine loads. We refer to this problem as stochastic Top_ℓ-norm load balancing and abbreviate it to $\mathsf{StochTop}_\ell\mathsf{LB}$.

2. **Identical-Machines Environment:** The machines are identical, so the processing time of a job does not depend on the machine that it is assigned to. That is, $X_{ij} = X_j$ for all machines i.

3. **StochNormLB with Special Distributions:** The following probability distributions are relevant to us:
   - **Deterministic Jobs:** Job sizes are deterministic, i.e., X_{ij} takes value p_{ij} with probability 1. Chakrabarty and Swamy [5] call this problem minimum norm load balancing. We abbreviate it to MinNormLB.
   - **Bernoulli Jobs:** The job variable X_{ij} is a weighted Bernoulli trial that takes size $s_{ij} \in \mathbb{R}_{\geq 0}$ with probability $q_{ij} \in [0, 1]$, and 0 otherwise. We call this problem as stochastic min-norm load balancing with Bernoulli jobs, and abbreviate it to BerNormLB.
   - **Poisson Jobs:** The job variable X_{ij} follows a Poisson distribution with a given mean $\lambda_{ij} \in \mathbb{R}_{\geq 0}$. That is, for any nonnegative integer k, $\mathbf{Pr}[X_{ij} = k] = e^{-\lambda_{ij}}\lambda_{ij}^k/k!$. We call this problem as stochastic min-norm load balancing with Poisson jobs, and abbreviate it to PoisNormLB. A key observation that makes

PoisNormLB easier to approximate is that the load on any machine is also a Poisson variable (see Fact 2.2).

5.1.2 Our Results

We state informal versions of our main results on stochastic min-norm load balancing. We start with our strongest results on $\mathsf{StochTop}_\ell\mathsf{LB}$ and BerNormLB, followed by our result for the general StochNormLB. Our results heavily exploit the machinery developed in Chapters 3 and 4.

Theorem 5.1 (proved in Chapter 6)**.** *There is an $O(1)$-approximation algorithm for* $\mathsf{StochTop}_\ell\mathsf{LB}$ *with arbitrary job-size distributions.*
Furthermore, when the machines are identical, a simple list-scheduling based algorithm yields an $O(1)$-approximate assignment.

Observe that the above result generalizes the $O(1)$-approximation algorithm for stochastic makespan minimization (i.e., StochNormLB with $f = \mathsf{Top}_1$-norm) given by Kleinberg et al. [23] for identical machines, and by Gupta et al. [10] for unrelated machines to all Top_ℓ norms. The above result is incomparable to that of Molinaro [34], who obtains an $O(1)$-approximation when f is an ℓ_p norm for some $p \in [1, \infty]$.

Theorem 5.2 (proved in Chapter 8)**.** *There is an $O(1)$-approximation algorithm for* BerNormLB *with an arbitrary monotone, symmetric norm.*
Furthermore, when the machines are identical, the approximation guarantee holds simultaneously for all monotone symmetric norms. That is, the algorithm returns an assignment that is simultaneously an $O(1)$-approximation for every *monotone, symmetric norm.*

As MinNormLB can be viewed as a trivial case of BerNormLB, modulo constant factors, Theorem 5.2 strictly generalizes the $O(1)$-approximation algorithm obtained by Chakrabarty and Swamy [5, 6].

We next state our approximation guarantees for general StochNormLB, i.e., where we have arbitrary job-size distributions, and an arbitrary monotone, symmetric norm.

Theorem 5.3 (proved in Chapter 7)**.** *There is an $O(\log\log m/\log\log\log m)$-approximation algorithm for* StochNormLB *on unrelated machines.*
Furthermore, when the machines are identical, the approximation guarantee holds simultaneously for all monotone symmetric norms.

In Chapter 7, we also give a very simple $O(\log\log m)$-approximation for StochNormLB on identical machines via a reduction to *vector scheduling*.

Finally, we obtain quite strong guarantees for PoisNormLB via an entirely different approach. We reduce PoisNormLB to MinNormLB with at most a $(1+\varepsilon)$-factor loss in approximation. This reduction yields the following guarantees.

Theorem 5.4 (proved in Chapter 9)**.** *For any $\varepsilon > 0$, there is a $(2+\varepsilon)$-approximation algorithm for* PoisNormLB *on unrelated machines with an arbitrary monotone, symmetric norm.*
Furthermore, when the machines are identical, there is a $(1+\varepsilon)$-approximation algorithm.

The above reduction applies in fact more generally to stochastic combinatorial optimization with Poisson distributions, and relies on a certain property, called *Schur convexity* of the objective function that arises with Poisson distributions. We remark that the f-norm in the reduced MinNormLB instance is different from the f-norm in the original PoisNormLB instance. We give full details in Chapter 9, and this chapter can be read independently of Chapters 6 – 8.

In the rest of this chapter, we build tools that will be useful in designing algorithms for StochNormLB instances with no special assumptions on the distributions of machine-load variables, unlike PoisNormLB.

5.2 Handling Sums of Independent Random Variables

In StochNormLB, an assignment $\sigma : J \to [m]$ induces a load of $\overrightarrow{\mathsf{load}}^\sigma(i) := \sum_{j\in J:\sigma(j)=i} X_{ij}$ on machine $i \in [m]$. We have two key challenges in working with these load variables. First, the load variables depend on the actual assignment, which we are trying to compute in the first place. Second, we do not have direct access to the probability distributions of load variables because the input only consists of distributions of X_{ij}'s. In the rest of this section, we focus on the problem posed by the second challenge.

Consider a composite random variable $S := \sum_{j\in[k]} Z_j$ that is a sum of k independent nonnegative random variables $\{Z_j\}_j$. We only assume that the Z_j's have finite mean, which is a reasonable assumption in load-balancing applications. We think of S as a placeholder for $\overrightarrow{\mathsf{load}}^\sigma(i)$ for some machine i. For notational convenience, let $Y := \overrightarrow{\mathsf{load}}^\sigma$ for some fixed assignment σ. When f is the Top_1 norm, the contribution of $\{Y_i\}_{i\in[m]}$ to $\mathbf{E}[f(Y)]$ can be captured by deriving probability bounds on the upper tail of Y_i for each machine i (see

[23, 10]). However, for Top_ℓ norms with $\ell > 1$ (and more generally, arbitrary monotone symmetric norms), it is unclear how to use upper-tail probability bounds to control the expected norm. Nevertheless, in Chapter 3, we saw that controlling $\sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}]$ for a suitable θ gives an approximate handle on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$, and in Chapter 4, we saw that simultaneously controlling $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ for all powers-of-2 ℓ gives an approximate handle on $\mathbf{E}[f(Y)]$.

Our goal in the rest of this section is to derive lower bounds on $\mathbf{E}[S^{\geq\theta}]$, for an arbitrary scalar $\theta \in \mathbb{R}_{\geq 0}$, by using distributional information about the Z_j's. We derive two separate lower bounds: one arising from truncated random variables ($\{Z_j^{<\theta}\}_j$ and the other arising from exceptional random variables ($\{Z_j^{\geq\theta}\}_j$). We start with the simpler bound.

5.2.1 Handling Exceptional Variables

Observe that for any $j \in [k]$, if the event $\{Z_j \geq \theta\}$ happens, then the event $\{S \geq \theta\}$ happens. The following *exceptional-items* bound is straightforward.

Lemma 5.5. *Let $S = \sum_{j\in[k]} Z_j$, where the Z_j's are independent random variables whose support lies in $\{0\} \cup [\theta, \infty)$. For any scalar $\theta \in \mathbb{R}_{\geq 0}$, we have $\mathbf{E}[S^{\geq\theta}] = \sum_{j\in[k]} \mathbf{E}[Z_j^{\geq\theta}]$.*

Proof. The support of S and Z_j's lies in $\{0\} \cup [\theta, \infty)$, so

$$\mathbf{E}[S^{\geq\theta}] = \mathbf{E}[S] = \sum_j \mathbf{E}[Z_j] = \sum_j \mathbf{E}[Z_j^{\geq\theta}]. \qquad \blacksquare$$

The separable nature of the right hand side in Lemma 5.5 makes it a convenient lower bound on $\mathbf{E}[S^{\geq\theta}]$ even without any further assumptions on the Z_j's.

5.2.2 Handling Truncated Variables

Unlike the exceptional-items bound, the bound involving truncated items requires nuanced ideas. First of all, since $S \geq \sum_j Z_j^{<\theta}$ with probability 1, we have:

$$\mathbf{E}[S^{\geq\theta}] \geq \mathbf{E}\Big[\big(\sum_j Z_j^{<\theta}\big)^{\geq\theta}\Big].$$

The above bound is ineffective because of its non-separable dependence on $Z_j^{<\theta}$. The following notion of *effective size* will be useful to us in obtaining separable bounds.

Definition 5.6 (Effective Size)**.** *For a nonnegative random variable Z and a parameter $\lambda > 1$, the λ-effective size $\beta_\lambda(Z)$ of Z is defined to be* $\log_\lambda \mathbf{E}[\lambda^Z]$. *We also define $\beta_1(Z) := \mathbf{E}[Z]$.*

Note that $\lambda^{\beta_\lambda(Z)}$ is simply the moment-generating function of Z evaluated at $\ln \lambda$.

The notion of effective size originated in queuing theory [16] and Kleinberg et al. [23] were the first to utilize this notion to obtain an $O(1)$-approximation algorithm for stochastic makespan minimization on identical machines. We give some intuition on why effective sizes are better than expected sizes in capturing the contribution of any one truncated random variable $Z_j^{<\theta}$ to $\mathbf{E}[S^{\geq\theta}]$. Observe that for any scalar $\theta' \in [0, \theta)$, the support of Z_j that lies in $[\theta', \theta)$ contributes to $\mathbf{E}[S^{\geq\theta}]$ as long as the remaining independent random variables $\{Z_{j'}^{<\theta}\}_{j' \neq j}$ in total contribute at least $\theta - \theta'$. Thus, larger values in the support of $Z_j^{<\theta}$ are represented more in $\mathbf{E}[S^{\geq\theta}]$. Note that the definition of λ-effective size inflates larger values more than smaller values. Furthermore, the inflation in itself becomes more aggressive as the parameter λ increases.

In the following we give quantitative upper and lower bounds on $\mathbf{E}[\left(\sum_j Z_j^{<\theta}\right)^{\geq\theta}]$ in terms of effective sizes of truncated random variables related to the Z_j variables. The first lemma — a straightforward application of Markov's inequality — shows that if the λ-effective size of a random variable is small, then its upper tail obeys an inverse power law in λ. The first part of Lemma 5.7 is lifted from [23].

Lemma 5.7. *Let Z be a nonnegative random variable and $\lambda \geq 1$ be a scalar. If $\beta_\lambda(Z) \leq b$ for some $b \in \mathbb{R}_{\geq 0}$, then for any $c \geq 0$, we have:*

$$\mathbf{Pr}[Z \geq b + c] \leq \lambda^{-c}.$$

Furthermore, if $\lambda \geq 2$, then

$$\mathbf{E}[Z^{\geq \beta_\lambda(Z)+1}] \leq \frac{\beta_\lambda(Z) + 3}{\lambda}.$$

Proof. The first part trivially holds for $\lambda = 1$ so assume that $\lambda > 1$. By definition, $\lambda^{\beta_\lambda(Z)} = \mathbf{E}[\lambda^Z]$, so Markov's inequality gives:

$$\mathbf{Pr}[Z \geq b + c] = \mathbf{Pr}[\lambda^Z \geq \lambda^{b+c}] \leq \frac{\mathbf{E}[\lambda^Z]}{\lambda^{b+c}} \leq \lambda^{-c}$$

For the second part, we have

$$\begin{aligned}
\mathbf{E}[Z^{\geq \beta_\lambda(Z)+1}] &= (\beta_\lambda(Z)+1)\cdot \mathbf{Pr}[Z \geq \beta_\lambda(Z)+1] + \int_{\beta_\lambda(Z)+1}^{\infty} \mathbf{Pr}[Z > u]\, du \\
&\leq \frac{\beta_\lambda(Z)+1}{\lambda} + \int_1^\infty \lambda^{-u}\, du && \text{(by the first part)} \\
&= \frac{\beta_\lambda(Z)+1}{\lambda} + \frac{1}{\ln\lambda}\cdot \int_{\ln\lambda}^\infty e^{-v}\, dv && \text{(change of variable)} \\
&= \frac{\beta_\lambda(Z)+1}{\lambda} + \frac{1}{\lambda\ln\lambda} \leq \frac{\beta_\lambda(Z)+3}{\lambda}
\end{aligned}$$

The last inequality above is because $\lambda \geq 2$. ■

The following two results will be useful to us. First, we show that the λ-effective size function is additive over sums of independent random variables. Second, we show that λ-effective size grows sublinearly in λ for $[0,1]$-bounded random variables.

Lemma 5.8. *For any two independent random variables Z and Z', we have:*

$$\beta_\lambda(Z+Z') = \beta_\lambda(Z) + \beta_\lambda(Z').$$

Proof. We have $\beta_\lambda(Z+Z') = \ln_\lambda \mathbf{E}[\lambda^{Z+Z'}] = \ln_\lambda\big(\mathbf{E}[\lambda^Z]\cdot\mathbf{E}[\lambda^{Z'}]\big) = \beta_\lambda(Z)+\beta_\lambda(Z')$, where the second equality follows from independence. ■

Lemma 5.9. *For any $[0,1]$-bounded random variable Z and a scalar $\lambda \geq 1$, we have:*

$$\beta_\lambda(Z) \leq \lambda\cdot\mathbf{E}[Z].$$

Proof. By definition, $\lambda^{\beta_\lambda(Z)} = \mathbf{E}[\lambda^Z] \leq \mathbf{E}[1+(\lambda-1)Z] = 1+(\lambda-1)\mathbf{E}[Z]$, where we use the inequality $\lambda^z \leq 1+(\lambda-1)z$ which holds for any $\lambda \geq 1$ and $z \in [0,1]$. Therefore:

$$\beta_\lambda(Z) \leq \ln_\lambda\big(1+(\lambda-1)\mathbf{E}[Z]\big) \leq \frac{(\lambda-1)\cdot\mathbf{E}[Z]}{\ln\lambda} \leq \lambda\cdot\mathbf{E}[Z].$$

In the above we use the elementary logarithmic inequality $\frac{z}{1+z} \leq \ln(1+z) \leq z$ that holds for any scalar $z \geq 0$. ■

By Lemmas 5.7 and 5.8, we can infer that if $\sum_j \beta_\lambda(Z_j^{<\theta}/\theta) = O(1)$ holds, then $\mathbf{E}[\big(\sum_j Z_j^{<\theta}/\theta\big)^{\geq\Omega(1)}] \leq O(1)/\lambda$, or equivalently $\mathbf{E}[\big(\sum_j Z_j^{<\theta}\big)^{\geq\Omega(\theta)}] \leq O(\theta)/\lambda$. A key contribution of Kleinberg et al. [23] is encapsulated by Lemma 5.10 below, which yields a lower bound on $\mathbf{E}[S^{\geq\theta}]$ in terms of the $\beta_\lambda(Z_j)$-effective sizes that complements the above upper bound. This lemma is obtained by combining various results from [23].

Lemma 5.10.
Let $S = \sum_{j\in[k]} Z_j$, where the Z_j's are independent $[0,\theta]$-bounded random variables. The following are true:

(i) *For any scalar $\lambda \geq 1$, we have*

$$\mathbf{E}[S^{\geq\theta}] \geq \theta \cdot \frac{\sum_{j\in[k]} \beta_\lambda(Z_j/4\theta) - 6}{4\lambda}.$$

(ii) *For $\lambda \in [1,2]$, we have $\mathbf{E}[S^{\geq\theta}] \geq \theta \cdot \left(\frac{1}{2} \cdot \sum_j \beta_\lambda(Z_j/\theta) - 1\right)$.*

(iii) *For $\lambda = 1$, we trivially have $\mathbf{E}[S^{\geq\theta}] \geq \theta \cdot (\sum_{j\in[k]} \mathbf{E}[Z_j/\theta] - 1)$.*

We devote the rest of this Chapter to the proof of Lemma 5.10. The proof is long and technical, so we split it into multiple parts.

5.2.3 A Lower Bound Based on Effective Sizes: Proof of Lemma 5.10

Without loss of generality, we may assume that $\theta = 1$. This is because if we set $Z'_j = Z_j/\theta$ for all $j \in [k]$ and define $S' := S/\theta$, then the Z'_j variables are supported on $[0,1]$, and $\mathbf{E}[S'^{\geq 1}] = \mathbf{E}[S^{\geq\theta}]/\theta$; so applying the result for $\theta = 1$ yields the desired inequalities.

To keep notation simple, we always reserve j to index over the set $[k]$. We first prove the second and third claims in the lemma. When $\lambda = 1$, we have $\mathbf{E}[S^{\geq 1}] \geq \mathbf{E}[S] - 1 = \sum_j \mathbf{E}[Z_j] - 1$, so we are done. Next, if $\lambda \leq 2$, then Lemma 5.9 gives:

$$\mathbf{E}[S^{\geq 1}] \geq \sum_j \mathbf{E}[Z_j] - 1 \geq \frac{1}{2} \cdot \sum_j \beta_\lambda(Z_j) - 1.$$

The lower bound in part (ii) is stronger than the lower bound in part (i), so for the rest of the proof we assume that $\lambda \geq 2$. We prove the lemma by showing that the following "volume" inequality holds:

$$\sum_j \beta_\lambda(Z_j/4) \leq (3\lambda + 2)\mathbf{E}[S^{\geq 1}] + 6 \,, \tag{5.1}$$

holds for any scalar $\lambda \geq 2$. As $3\lambda + 2 \leq 4\lambda$ for $\lambda \geq 2$, (5.1) implies the lemma.

At a high level, we combine and adapt the proofs of Lemmas 3.2 and 3.4 from [23] to obtain our result. We say that a Bernoulli trial is of type (q, s) if it takes size s with probability q, and size 0 with probability $1-q$. For a Bernoulli trial B of type (q, s), Kleinberg et al. [23] define a modified notion of effective size: $\beta'_\lambda(B) := \min(s, sq\lambda^s)$. The following claim will be useful.

Claim 5.11 (Proposition 2.5 from [23])**.** $\beta_\lambda(B) \leq \beta'_\lambda(B)$.

Roughly speaking, inequality (5.1) states that each unit of λ-effective size contributes "λ^{-1}-units" towards $\mathbf{E}[S^{\geq 1}]$. This is indeed what we show, but we first reduce to the setting of Bernoulli trials.

The proof will involve various transformations of the Z_j random variables, and the notion of stochastic dominance will be convenient to compare the various random variables so obtained. A random variable B *stochastically dominates* another random variable R, denoted $R \preceq_{\text{sd}} B$, if $\mathbf{Pr}[B \geq t] \geq \mathbf{Pr}[R \geq t]$ for all $t \in \mathbb{R}$. We will use the following well-known facts about stochastic dominance.

(F1) If $B \preceq_{\text{sd}} R$, then for any non-decreasing function $u : \mathbb{R} \mapsto \mathbb{R}$, we have $\mathbf{E}[u(B)] \leq \mathbf{E}[u(R)]$.

(F2) If $B_i \preceq_{\text{sd}} R_i$ for all $i = 1, \ldots, k$, then $\left(\sum_{i=1}^k B_i\right) \preceq_{\text{sd}} \left(\sum_{i=1}^k R_i\right)$.

All the random variables encountered in the proof will be nonnegative, and we will often omit stating this explicitly.

Bernoulli Decomposition

We utilize a result of [23] that shows how to replace an arbitrary random variable R with a sum of Bernoulli trials that is "close" to R in terms of stochastic dominance. This allows us to reduce to the case where all random variables are Bernoulli trials (Lemma 5.14).

Following [23], we say that a random variable is *geometric* if its support is contained in $\{2^r : r \in \mathbb{Z}\}$. We use $\text{supp}(R)$ to denote the support of a random variable R.

Lemma 5.12 (Lemma 3.10 from [23])**.** *Let R be a geometric random variable. Then there exists a set of independent Bernoulli trials $B_1, \ldots, B_p$ such that $B = B_1 + \cdots + B_p$ satisfies $\mathbf{Pr}[R = t] = \mathbf{Pr}[t \leq B < 2t]$ for all $t \in \text{supp}(R)$. Furthermore, the support of each B_i is contained in the support of R.*

Corollary 5.13. *Let R be a geometric random variable, and B be the corresponding sum of Bernoulli trials given by Lemma 5.12. Then, we have $B/2 \preceq_{\mathrm{sd}} R \preceq_{\mathrm{sd}} B$.*

Proof. To see that $R \preceq_{\mathrm{sd}} B$, consider any $t \geq 0$. We have

$$\mathbf{Pr}[R \geq t] = \sum_{t' \in \mathrm{supp}(R): t' \geq t} \mathbf{Pr}[R = t'] = \sum_{t' \in \mathrm{supp}(R): t' \geq t} \mathbf{Pr}[t' \leq B < 2t'] \leq \mathbf{Pr}[B \geq t]$$

where the second equality is due to Lemma 5.12, and the last inequality follows since the intervals $[t', 2t')$ are disjoint for $t' \in \mathrm{supp}(R)$, as R is a geometric random variable.

To show that $B/2 \preceq_{\mathrm{sd}} R$, we argue that $\mathbf{Pr}[R < t] \leq \mathbf{Pr}[B < 2t]$ for all $t \in \mathbb{R}$. This follows from a very similar argument as above. We have

$$\mathbf{Pr}[R < t] = \sum_{t' \in \mathrm{supp}(R): t' < t} \mathbf{Pr}[R = t'] = \sum_{t' \in \mathrm{supp}(R): t' < t} \mathbf{Pr}[t' \leq B < 2t'] \leq \mathbf{Pr}[B < 2t].$$

The last inequality again follows because the intervals $[t', 2t')$ are disjoint for $t' \in \mathrm{supp}(R)$, as R is a geometric random variable. ∎

We intend to apply Lemma 5.12 to the Z_j's to obtain a collection of Bernoulli trials, but first we need to convert them to geometric random variables. For technical reasons that will be clear soon, we first scale our random variables by a factor of 4; then, we convert each scaled random variable to a geometric random variable by rounding up the values in its support to the closest power of 2. Formally, for each j, let Z_j^{rd} denote the geometric random variable obtained by *rounding up* each value in the support of $Z_j/4$ to the closest power of 2. (So, for instance, 0.22 would get rounded up to 0.25, whereas 1/8 would stay the same.) Note that Z_j^{rd} is a geometric $[0, 1/4]$-bounded random variable.

We now apply Lemma 5.12 to each Z_j^{rd} to obtain a collection $\{Z_\ell^{\mathsf{ber}}\}_{\ell \in F_j}$ of independent Bernoulli trials. Let F be the disjoint union of the F_j's, so F is the collection of all the Bernoulli variables so obtained. Define $S^{\mathsf{ber}} := \sum_{\ell \in F} Z_\ell^{\mathsf{ber}}$. For $\ell \in F$, let Z_ℓ^{ber} be a Bernoulli trial of type (q_ℓ, s_ℓ); from Lemma 5.12, we have that $s_\ell \in [0, 1/4]$ and is a (inverse) power of 2. We now argue that it suffices to prove (5.1) for the Bernoulli trials $\{Z_\ell^{\mathsf{ber}}\}_{\ell \in F}$.

Lemma 5.14. *Inequality (5.1) follows from the inequality:*

$$\sum_{\ell \in F} \beta_\lambda(Z_\ell^{\mathsf{ber}}) \leq (3\lambda + 2)\mathbf{E}[(S^{\mathsf{ber}})^{\geq 1}] + 6$$

By Claim 5.11, this in turn is implied by the inequality

$$\sum_{\ell \in F} \beta'_\lambda \left(Z_\ell^{\mathsf{ber}} \right) \leq (3\lambda + 2)\mathbf{E}[(S^{\mathsf{ber}})^{\geq 1}] + 6. \tag{5.2}$$

Proof. Fix any $j \in [k]$. By Corollary 5.13, we have $Z_j^{\mathsf{rd}} \preceq_{\text{sd}} \sum_{\ell \in F_j} Z_\ell^{\mathsf{ber}}$. We also have that $Z_j/4 \leq Z_j^{\mathsf{rd}}$, and so $Z_j/4 \preceq_{\text{sd}} \sum_{\ell \in F_j} Z_\ell^{\mathsf{ber}}$. Therefore, since λ^x is an increasing function of x, we have $\beta_\lambda(Z_j/4) \leq \beta_\lambda(\sum_{\ell \in F_j} Z_\ell^{\mathsf{ber}}) = \sum_{\ell \in F_j} \beta_\lambda(Z_\ell^{\mathsf{ber}})$. Summing over all j, we obtain that $\sum_{j \in [k]} \beta_\lambda(Z_j/4) \leq \sum_{\ell \in F} \beta_\lambda(Z_\ell^{\mathsf{ber}})$.

Corollary 5.13 also yields that $\sum_{\ell \in F_j} Z_\ell^{\mathsf{ber}} \preceq_{\text{sd}} 2 \cdot Z_j^{\mathsf{rd}}$ for all j. We also have $Z_j^{\mathsf{rd}} \leq 2(Z_j/4)$, and therefore, we have $\sum_{\ell \in F_j} Z_\ell^{\mathsf{ber}} \preceq_{\text{sd}} Z_j$ for all j. Using Fact (F2), this implies that $S^{\mathsf{ber}} \preceq_{\text{sd}} S$, and hence, by Fact (F1), we have $\mathbf{E}[(S^{\mathsf{ber}})^{\geq 1}] \leq \mathbf{E}[S^{\geq 1}]$.

To summarize, we have shown that $\sum_{j \in [k]} \beta_\lambda(Z_j/4)$ is at most $\sum_{\ell \in F} \beta_\lambda(Z_\ell^{\mathsf{ber}})$, which is at most the LHS of (5.2) (by Claim 5.11), and the RHS of (5.1) is at least the RHS of (5.2). ∎

Proof of Inequality (5.2)

We now focus on proving inequality (5.2). Let $F^{\mathsf{sml}} := \{\ell \in F : \lambda^{s_\ell} \leq 2\}$ index the set of *small* Bernoulli trials, and let $F^{\mathsf{lg}} := F \setminus F^{\mathsf{sml}}$ index the remaining *large* Bernoulli trials.

It is easy to show that the total modified effective size of small Bernoulli trials is at most $2\,\mathbf{E}[S^{\mathsf{ber}}] \leq 2\,\mathbf{E}[(S^{\mathsf{ber}})^{\geq 1}]+2$ (see Claim 5.15 and inequality (vol-small)). Bounding the modified effective size of the large Bernoulli trials is more involved. Roughly speaking, we first consolidate these random variables by replacing them with "simpler" Bernoulli trials, and then show that each unit of total modified effective size of these simpler Bernoulli trials makes a contribution of λ^{-1} towards $\mathbf{E}[(S^{\mathsf{ber}})^{\geq 1}]$. The constant 6 in (5.2) arises due to two reasons: (i) because we bound the modified effective size of the small Bernoulli trials by $O\big(\mathbf{E}[S^{\mathsf{ber}}]\big)$ (as opposed to $O\big(\mathbf{E}[(S^{\mathsf{ber}})^{\geq 1}]\big)$); and (ii) because we lose some modified effective size in the consolidation of large Bernoulli trials. [1]

Claim 5.15 (Shown in Lemma 3.4 in [23])**.** *Let B be a Bernoulli trial of type (q, s) with $\lambda^s \leq 2$. Then, $\beta'_\lambda(B) \leq 2\,\mathbf{E}[B]$.*

[1] Note that some additive constant must *unavoidably* appear on the RHS of inequalities (5.1) and (5.2); that is, we cannot bound the total effective size (or modified effective size) by a purely multiplicative factor of $\mathbf{E}[S^{\geq 1}]$, even for Bernoulli trials. This is because if $\lambda = 1$, and say we have only one (Bernoulli) random variable Z that is strictly less than 1, then its effective size (as also its modified effective size) is simply $\mathbf{E}[Z]$, whereas $\mathbf{E}[S^{\geq 1}] = \mathbf{E}[Z^{\geq 1}] = 0$.

Proof. By definition, $\beta'_\lambda(B) = \min(s, sq\lambda^s) \le 2qs = 2\,\mathbf{E}[B]$. ■

By the above claim we get the following volume inequality for the small Bernoulli trials.

$$\sum_{\ell \in F^{\mathsf{sml}}} \beta'_\lambda\left(Z_\ell^{\mathsf{ber}}\right) \le 2 \sum_{\ell \in F^{\mathsf{sml}}} \mathbf{E}[Z_\ell^{\mathsf{ber}}] \le 2\,\mathbf{E}[S^{\mathsf{ber}}] \le 2\,\mathbf{E}[(S^{\mathsf{ber}})^{\ge 1}] + 2. \qquad \text{(vol-small)}$$

We now handle the large Bernoulli trials. For each $\ell \in F^{\mathsf{lg}}$, we set $q'_\ell = \min(q_\ell, \lambda^{-s_\ell})$. Observe that this operation does not change the modified effective size, and we have that $\beta'_\lambda((q_\ell, s_\ell)) = \beta'_\lambda((q'_\ell, s_\ell)) = s_\ell q'_\ell \lambda^{s_\ell}$. The following claim from [23] is useful in consolidating Bernoulli trials of the same size.

Claim 5.16 (Claim 3.1 from [23])**.** *Let $\mathcal{E}_1, \ldots, \mathcal{E}_p$ be independent events, with $\mathbf{Pr}[\mathcal{E}_i] = p_i$. Let $\mathcal{E}'$ be the event that at least one of these events occurs. Then $\mathbf{Pr}[\mathcal{E}'] \ge \frac{1}{2}\min(1, \sum_i p_i)$.*

Consolidation for a fixed size. For each s that is an inverse power of 2, we define $F^s := \{\ell \in F^{\mathsf{lg}} : s_\ell = s\}$, so that $\bigcup_s F^s$ is a partition of F^{lg}. Next, we further partition F^s into sets $P_1^s, \ldots, P_{n_s}^s$ such that for all $i = 1, \ldots, n_s - 1$, we have $2\lambda^{-s} \le \sum_{\ell \in P_i^s} q'_\ell < 3\lambda^{-s}$ and $\sum_{\ell \in P_{n_s}^s} q'_\ell < 2\lambda^{-s}$. Such a partitioning always exists since for each $\ell \in F^s$ we have $q'_\ell \le \lambda^{-s}$ by definition. We now apply Claim 5.16 to "consolidate" each P_i^s: by this, we mean that for $i = 1, \ldots, n_s - 1$, we think of representing P_i^s by the "simpler" Bernoulli trial B_i^s of type (λ^{-s}, s) and using this to replace the individual random variables in P_i^s.

By Claim 5.16, for any $i = 1, \ldots, n_s - 1$, we have

$$\mathbf{Pr}\Big[\sum_{\ell \in P_i^s} Z_\ell^{\mathsf{ber}} \ge s\Big] \ge \lambda^{-s} = \mathbf{Pr}[B_i^s = s]$$

(note that this only works for large Bernoulli trials since $2\lambda^{-s} \le 1$); hence, it follows that $B_i^s \preceq_{\mathrm{sd}} \sum_{\ell \in P_i^s} Z_\ell^{\mathsf{ber}}$.

Note that

$$\sum_{\ell \in P_i^s} \beta'_\lambda\left(Z_\ell^{\mathsf{ber}}\right) = s\lambda^s \sum_{\ell \in P_i^s} q'_\ell < \begin{cases} 3s; & \text{if } i \in \{1, \ldots, n_s - 1\} \\ 2s; & \text{if } i = n_s. \end{cases}$$

Also, $\beta'_\lambda(B_i^s) = \min(s, s\lambda^{-s}\lambda^s) = s$. Putting everything together,

$$(n_s - 1)s = \sum_{i=1}^{n_s - 1} \beta'_\lambda(B_i^s) \ge \frac{\sum_{\ell \in F^s} \beta'_\lambda\left(Z_\ell^{\mathsf{ber}}\right) - 2s}{3}.$$

Consolidation across different sizes. Summing the above inequality for all s (note that each s is an inverse power of 2 and at most $1/4$), we obtain that

$$\sum_s \sum_{i=1}^{n_s-1} \beta'_\lambda(B_i^s) \geq \frac{\sum_{\ell \in F^{\mathsf{lg}}} \beta'_\lambda\left(Z_\ell^{\mathsf{ber}}\right) - 1}{3}. \tag{5.3}$$

Let vol denote the RHS of (5.3), and let $m := \lfloor \mathsf{vol} \rfloor$. (Note that m could be 0; but if $m = 0$, then $\sum_{\ell \in F^{\mathsf{lg}}} \beta'_\lambda\left(Z_\ell^{\mathsf{ber}}\right) < 4$, which combined with (vol-small) yields (5.2).) Since each B_i^s is a Bernoulli trial of type (λ^{-s}, s), where s is an inverse power of 2, we can obtain m disjoint subsets $A_1, \ldots, A_m$ of (s, i) pairs from the entire collection $\{B_i^s\}_{s,i}$ of Bernoulli trials, such that $\sum_{(s,i)\in A_u} \beta'_\lambda(B_i^s) = \sum_{(s,i)\in A_u} s = 1$ for each $u \in [m]$. [2] For each subset A_u,

$$\mathbf{Pr}\Big[\sum_{(s,i)\in A_u} B_i^s = 1\Big] = \prod_{(s,i)\in A_u} \mathbf{Pr}[B_i^s = s] = \prod_{(s,i)\in A_u} \lambda^{-s} = \lambda^{-1}\ .$$

Finishing up the proof of inequality (5.2). For any nonnegative random variables R_1, R_2, we have $\mathbf{E}[(R_1 + R_2)^{\geq 1}] \geq \mathbf{E}[R_1^{\geq 1}] + \mathbf{E}[R_2^{\geq 1}]$. So,

$$\mathbf{E}\left[\Big(\sum_s \sum_{i=1}^{n_s-1} B_i^s\Big)^{\geq 1}\right] \geq \sum_{u=1}^{m} \mathbf{E}\left[\Big(\sum_{(s,i)\in A_u} B_i^s\Big)^{\geq 1}\right] = \frac{m}{\lambda} \geq \frac{\sum_{\ell \in F^{\mathsf{lg}}} \beta'_\lambda\left(Z_\ell^{\mathsf{ber}}\right) - 4}{3\lambda}.$$

As noted earlier, we have that $B_i^s \preceq_{\mathrm{sd}} \sum_{\ell \in P_i^s} Z_\ell^{\mathsf{ber}}$ for all s, and all $i = 1, \ldots, n_s - 1$. By Fact (F2), it follows that $\left(\sum_s \sum_{i=1}^{n_s-1} B_i^s\right) \preceq_{\mathrm{sd}} \left(\sum_s \sum_{i=1}^{n_s-1} \sum_{\ell \in P_i^s} Z_\ell^{\mathsf{ber}}\right)$. Also,

$$\sum_s \sum_{i=1}^{n_s-1} \sum_{\ell \in P_i^s} Z_\ell^{\mathsf{ber}} \leq \sum_{\ell \in F^{\mathsf{lg}}} Z_\ell^{\mathsf{ber}} \leq \sum_{\ell \in F} Z_\ell^{\mathsf{ber}} = S^{\mathsf{ber}},$$

and combining the above with Fact (F1), we obtain that

$$\mathbf{E}\Big[\Big(\sum_s \sum_{i=1}^{n_s-1} B_i^s\Big)^{\geq 1}\Big] \leq \mathbf{E}[(S^{\mathsf{ber}})^{\geq 1}].$$

[2] To justify this statement, it suffices to show the following. Suppose we have created some r sets $A_1, \ldots, A_r$, where $r < m$, and let I be the set of (s, i) pairs indexing the Bernoulli trials that are not in $A_1, \ldots, A_r$; then, we can find a subset $I' \subseteq I$ such that $\sum_{(s,i)\in I'} s = 1$. To see this, first since $r < m$, we have $\sum_{(s,i)\in I} s \geq 1$. We sort the (s, i) pairs in I in non-increasing order of s; to avoid excessive notation, let I denote this sorted list. Now since each s is an inverse power of 2, it is easy to see by induction that if J is a prefix of I such that $\sum_{(s,i)\in J} s < 1$, then $1 - \sum_{(s,i)\in J} s$ is at least as large as the s-value of the pair in I appearing immediately after J. Coupled with the fact that $\sum_{(s,i)\in I} s \geq 1$, this implies that there is a prefix I' such that $\sum_{(s,i)\in I'} s = 1$.

Thus, we have shown that

$$\sum_{\ell \in F^{\mathsf{lg}}} \beta'_\lambda \left(Z_\ell^{\mathsf{ber}}\right) \leq 3\lambda \mathbf{E}[(S^{\mathsf{ber}})^{\geq 1}] + 4. \tag{vol-large}$$

We finish the proof of inequality (5.2), and hence the lemma by adding (vol-small) and (vol-large):

$$\sum_{\ell \in F} \beta'_\lambda \left(Z_\ell^{\mathsf{ber}}\right) \leq (3\lambda + 2)\mathbf{E}[(S^{\mathsf{ber}})^{\geq 1}] + 6. \qquad \blacksquare$$

Chapter 6

Stochastic Load Balancing with Top_ℓ Norms

In this chapter, we design an $O(1)$-approximation algorithm for $\mathsf{StochTop}_\ell\mathsf{LB}$. Recall that the objective in this problem is to find an assignment $\sigma : J \to [m]$ of jobs to machines that minimizes $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)]$ for a specified $\ell \in [m]$. For the rest of this chapter we reserve ℓ to refer to the ℓ in the Top_ℓ objective. Let σ^* denote an optimal solution and $\mathsf{OPT}_\ell := \mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})]$ denote the optimal solution value. We split this chapter into two sections: first, we restrict ourselves to the simpler identical-machines setting and give a combinatorial $O(1)$-approximation algorithm; and we then give a more involved LP-based approximation algorithm for the unrelated-machines setting.

6.1 Identical Machines

In the identical-machines setting, the processing time of a job j (on any machine) is X_j. We assume that a succinct representation of the distribution of X_j is given as part of the input. The main result in this section is the following.

Theorem 6.1.
There is an $O(1)$-approximation algorithm for stochastic Top_ℓ-norm load balancing on identical machines with arbitrary job-size distributions.

Our approximation strategy for $\mathsf{StochTop}_\ell\mathsf{LB}$ is similar to the one used in [23] for minimizing expected makespan, i.e., the Top_1-norm case. Suppose that we guess t_ℓ such that

ℓt_ℓ is roughly OPT_ℓ. Our algorithm in Theorem 6.1 is based on simultaneously approximating two sub-instances, *exceptional* and *truncated*, of the given $\mathsf{StochTop}_\ell\mathsf{LB}$ instance. As indicated by their names, the size of a job j is $X_j^{\geq t_\ell}$ in the exceptional sub-instance, and $X_j^{<t_\ell}$ in the truncated sub-instance. The exceptional sub-instance is easy to approximate because the expected Top_ℓ norm of *every* assignment is $\Theta(\mathsf{OPT}_\ell)$. Up to an $O(1)$-loss in approximation, we reduce the truncated sub-instance to (deterministic) makespan minimization problem with job sizes given by $p_j := \beta_{m/\ell}(X_j^{<t_\ell})$. A simple greedy algorithm applied to the latter problem yields the desired $O(1)$-approximate assignment for the original $\mathsf{StochTop}_\ell\mathsf{LB}$ instance. We now formalize the above strategy.

6.1.1 The Exceptional Sub-Instance

Since we do not have explicit access to t_ℓ (that is roughly OPT_ℓ/ℓ), we work with an arbitrary guess θ of t_ℓ.

Lemma 6.2. *Let θ be a positive scalar. Consider an instance of* $\mathsf{StochTop}_\ell\mathsf{LB}$ *where for each job j, the support of X_j lies in $\{0\} \cup [\theta, \infty)$ (equivalently, $\mathbf{Pr}[0 < X_j < \theta] = 0$). Let $\sigma : J \to [m]$ denote an arbitrary assignment. The following are true:*

(i) *If $\sum_{j\in J} \mathbf{E}[X_j] \leq \ell\theta$, then $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] \leq 2\ell\theta$.*

(ii) *If $\sum_{j\in J} \mathbf{E}[X_j] > \ell\theta$, then $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] > \ell\theta/2$.*

Proof. Let $Y := \overrightarrow{\mathsf{load}}^\sigma$ denote the load vector induced by σ. Observe that for any machine i, the support of its load variable Y_i lies in $\{0\} \cup [\theta, \infty)$. Thus, $\mathbf{E}[Y_i^{\geq\theta}] = \mathbf{E}[Y_i] = \sum_{j\in J:\sigma(j)=i} \mathbf{E}[X_j]$. Summing this equation over all machines gives a term that is independent of σ:

$$\sum_{i\in[m]} \mathbf{E}[Y_i^{\geq\theta}] = \sum_{j\in J} \mathbf{E}[X_j].$$

The first and second parts follow from Lemmas 3.5 and 3.6, respectively. ■

The above lemma implies that if we take $\theta \geq 2\,\mathsf{OPT}_\ell/\ell$, then any assignment induces expected Top_ℓ load at most $2\ell\theta$ for the exceptional sub-instance. In particular, for $\theta = 2\mathsf{OPT}_\ell/\ell$, this shows that every assignment is a 4-approximation for the exceptional sub-instance.

6.1.2 The Truncated Sub-Instance

We use techniques from Chapter 5 to approximate the truncated sub-instance. Recall that in the truncated sub-instance, the support of each X_j lies in $[0, \theta)$. Consider an arbitrary assignment $\sigma : J \to [m]$, and let $Y := \overrightarrow{\mathsf{load}}^\sigma$ denote the induced load vector. By Lemma 5.10, for any machine $i \in [m]$, effective-size parameter $\lambda_i \geq 1$, and a scalar $\theta \in \mathbb{R}_{\geq 0}$, we have:

$$\mathbf{E}[Y_i^{\geq\theta}] \geq \theta \cdot \frac{\sum_{j \in J : \sigma(j)=i} \beta_{\lambda_i}(X_j/4\theta) - 6}{4\lambda_i} \qquad \text{(eff-size-lb-m/c)}$$

Since the machines are identical, it is natural to set all λ_i to a common value $\lambda \geq 1$ and sum the above inequality over all machines $i \in [m]$. This yields a lower bound on $\sum_i \mathbf{E}[Y_i^{\geq\theta}]$ that is independent of the assignment σ.

$$\sum_{i \in [m]} \mathbf{E}[Y_i^{\geq\theta}] \geq \theta \cdot \frac{\sum_{j \in J} \beta_\lambda(X_j/4\theta) - 6m}{4\lambda} \qquad \text{(eff-size-lb-all)}$$

Recall from Lemma 3.6 that the lower bound on $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ is strongest when θ is such that $\sum_i \mathbf{E}[Y_i^{\geq\theta}]$ is just above $\ell\theta$. Now observe that if we fix $\lambda = m/\ell$ and choose the largest θ such that $\sum_j \beta_\lambda(X_j/4\theta) \approx 10\,m$ holds, then (eff-size-lb-all) yields $\sum_i \mathbf{E}[Y_i^{\geq\theta}] \approx \ell\theta$. This motivates our choice of $\lambda = m/\ell$ for the identical-machines setting.

The following technical lemma is the counterpart of Lemma 6.2.

Lemma 6.3. *Fix $\lambda = m/\ell$. Let θ be a positive scalar. Consider an instance of* $\mathsf{StochTop}_\ell\mathsf{LB}$ *where for each job j, the support of X_j lies in $[0, \theta)$ (equivalently, $\mathbf{Pr}[X_j \geq \theta] = 0$). Let $\sigma : J \to [m]$ denote an arbitrary assignment, and let $\alpha := \max_{i \in [m]} \sum_{j:\sigma(j)=i} \beta_\lambda(X_j/4\theta)$ be the "β_λ-makespan" of this assignment. The following are true:*

(i) *Suppose that $\sum_{j \in J} \beta_\lambda(X_j/4\theta) \leq 10\,m$ holds. For $\ell \leq m/2$ we have $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] \leq (8\alpha + 16)\ell\theta$, and for $\ell > m/2$ we have $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] \leq 80\,\ell\theta$.*

(ii) *Suppose that $\sum_{j \in J} \beta_\lambda(X_j/4\theta) > 10\,m$ holds. Then we have $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] > \ell\theta/2$.*

Proof. For notational convenience, let $Y := \overrightarrow{\mathsf{load}}^\sigma/\theta$ denote the load vector induced by σ scaled down by a factor θ. Note $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] = \theta\,\mathbf{E}[\mathsf{Top}_\ell(Y)]$. We split the proof of the first part into two cases. If $\lambda < 2$ (i.e., $\ell > m/2$), we have:

$$\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq \mathbf{E}[\mathsf{Top}_m(Y)] = \sum_{j \in J} \mathbf{E}[X_j/\theta] \leq 4 \cdot \sum_{j \in J} \beta_\lambda(X_j/4\theta) \leq 40\,m \leq 80\,\ell,$$

where the second inequality is because the effective size function is non-decreasing over the parameter λ. Therefore, $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] \leq 80\,\ell\theta$.

Next, suppose that $\lambda \geq 2$. Since X_j's are independent random variables, for any machine i we have $\beta_\lambda(Y_i/4) = \sum_{j:\sigma(j)=i} \beta_\lambda(X_j/4\theta) \leq \alpha$, by the definition of α. By Lemma 5.7:

$$\mathbf{E}\left[\left(\frac{Y_i}{4}\right)^{\geq \alpha+1}\right] \leq \mathbf{E}\left[\left(\frac{Y_i}{4}\right)^{\geq \beta_\lambda(Y_i/4)+1}\right] \leq \frac{\beta_\lambda(Y_i/4)+3}{\lambda} \leq \frac{\alpha+3}{\lambda}.$$

Summing over all machines, we get

$$\sum_{i\in[m]} \mathbf{E}\left[\left(\frac{Y_i}{4}\right)^{\geq \alpha+1}\right] \leq (\alpha+3)\cdot\frac{m}{\lambda} = (\alpha+3)\ell.$$

Equivalently, $\sum_i \mathbf{E}[Y_i^{\geq 4\alpha+4}] \leq 4(\alpha+3)\ell$. Next, we use Lemma 3.4 with $\theta = 4\alpha+4$ to get $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq (4\alpha+4)\ell + 4(\alpha+3)\ell = (8\alpha+16)\ell$. Therefore, $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] \leq (8\alpha+16)\ell\theta$.

For the second part, suppose that $\sum_{j\in J}\beta_\lambda(X_j/4\theta) > 10m$ holds. Observe that for any $i \in [m]$, $Y_i = \sum_{j:\sigma(j)=i}(X_j/\theta)$ is a sum of independent $[0,1]$-bounded random variables, so Lemma 5.10 gives:

$$\sum_{i\in[m]} \mathbf{E}[Y_i^{\geq 1}] \geq \frac{1}{4\lambda}\cdot\sum_{i\in[m]}\Big(\sum_{j:\sigma(j)=i}\beta_\lambda(X_j/4\theta) - 6\Big) = \frac{\ell}{4m}\cdot\Big(\sum_{j\in J}\beta_\lambda(X_j/4\theta) - 6m\Big) > \ell.$$

Now Lemma 3.6 implies that $\mathbf{E}[\mathsf{Top}_\ell(Y)] > \ell/2$, which is equivalent to $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] > \ell\theta/2$. ■

6.1.3 Our Algorithm

Lemmas 6.2 and 6.3 suggest the following natural approach to obtain a constant-factor approximation for $\mathsf{StochTop}_\ell\mathsf{LB}$ on identical machines. For a given scalar $\theta > 0$, "split" each job j as $X_j = X_j^{<\theta} + X_j^{\geq\theta}$. This yields the *exceptional sub-instance* with the exceptional job-variables $\{X_j^{\geq\theta}\}_j$, and the *truncated sub-instance* with the truncated job-variables $\{X_j^{<\theta}\}_j$. Clearly, for any assignment σ, if $\overline{Y}$ and $\widetilde{Y}$ denote the respective load vectors in these two sub-instances, we have $\overrightarrow{\mathsf{load}}^\sigma = \overline{Y} + \widetilde{Y}$. Now, we can use binary search to find the *right* threshold t_ℓ, such that for $\theta = t_\ell$, the conditions in part (i) of Lemmas 6.2 and 6.3 hold, and for some $\theta \geq t_\ell/(1+\epsilon)$, the opposite is true. Note that the latter condition above implies that $\mathsf{OPT}_\ell = \Omega(\ell t_\ell)$. A simple list-scheduling algorithm for deterministic scheduling with job sizes $p_j := \beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell)$ gives us an assignment with objective value $O(\ell t_\ell)$ in the original instance. We give full details now.

We first establish suitable upper and lower bounds for the right t_ℓ so that it can be computed efficiently. Define $\kappa := \max_{j\in J} \mathbf{E}[X_j]$ (which can be easily computed from the input data). By the contrapositive of part (ii) of Lemmas 6.2 and 6.3, for any $\theta \geq 2\mathsf{OPT}_\ell/\ell$ the following inequalities hold: $\sum_{j\in J} \mathbf{E}[X_j^{\geq\theta}] \leq \ell\theta$ and $\sum_{j\in J} \beta_{m/\ell}(X_j^{<\theta}/4\theta) \leq 8m$. In particular, since $\mathsf{OPT}_\ell \leq \mathbf{E}[\mathsf{Top}_m(\overrightarrow{\mathsf{load}}^{\sigma^*})] = \sum_{j\in J} \mathbf{E}[X_j] \leq n\kappa$, the above two inequalities hold for $\theta = \mathsf{hi} := 2n\kappa \geq 2\mathsf{OPT}_\ell/\ell$. Next, observe that for any job k and $\theta = \mathbf{E}[X_k]/(m+2)$, we have

$$\sum_j \mathbf{E}[X_j^{\geq\theta}] \geq \mathbf{E}[X_k^{\geq\theta}] \geq \mathbf{E}[X_k] - \theta = (m+1)\theta > \ell\theta.$$

Thus, for any $\theta \leq \mathsf{low} := \kappa/(m+2)$, we have $\sum_j \mathbf{E}[X_j^{\geq\theta}] > \ell\theta$, so the condition in Lemma 6.2(ii) holds.

Proof of Theorem 6.1. Let $\epsilon > 0$ be a small constant (say, 1/1000). By enumerating over scalars of the form $(1+\epsilon)^r$ with $r \in \mathbb{Z}$, we can efficiently find a scalar $t_\ell \in \mathbb{R}_{\geq 0}$ satisfying:

(i) $\mathsf{low}/(1+\epsilon) < t_\ell < \mathsf{hi}(1+\epsilon)$.

(ii) $\sum_{j\in J} \mathbf{E}[X_j^{\geq t_\ell}] \leq \ell t_\ell$ and $\sum_{j\in J} \beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell) \leq 10\,m$.

(iii) $\sum_{j\in J} \mathbf{E}[X_j^{\geq t'_\ell}] > \ell t'_\ell$ or $\sum_{j\in J} \beta_{m/\ell}(X_j^{<t'_\ell}/4t'_\ell) > 10\,m$, where $t'_\ell := t_\ell/(1+\epsilon)$.

We first establish a lower bound on OPT_ℓ. Depending on which condition in (iii) holds, the optimum value of the exceptional or the truncated sub-instance (w.r.t. the scalar $\theta = t'_\ell$) is larger than $\ell t'_\ell/2$. So, $\mathsf{OPT}_\ell \geq \ell t_\ell/2(1+\epsilon)$.

We now describe the algorithm for obtaining an assignment σ with $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})] = O(\ell t_\ell)$. Consider an instance of *deterministic* (makespan-minimization) scheduling with job sizes $p_j := \beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell) \in [0,1]$. By Property (ii) above, we have $\sum_{j\in J} p_j \leq 10\,m$. Thus, a simple greedy algorithm gives an assignment $\sigma : J \to [m]$ that assigns total p_j-load at most $\sum_{j\in J} p_j/m + 1 \leq 11 =: \alpha$ on any machine. Let $\overline{Y}$ and $\widetilde{Y}$ denote the load vector induced by σ in the exceptional and truncated sub-instances w.r.t. the threshold $\theta = t_\ell$. Lemma 6.2 (i) gives $\mathbf{E}[\mathsf{Top}_\ell(\overline{Y})] \leq 2\ell t_\ell$, and Lemma 6.3 (i) gives

$$\mathbf{E}[\mathsf{Top}_\ell(\widetilde{Y})] \leq \max\{(8\alpha+16), 80\} \cdot \ell t_\ell \leq 104\,\ell t_\ell.$$

As $\overrightarrow{\mathsf{load}}^{\sigma} = \overline{Y} + \widetilde{Y}$, triangle inequality of norms gives:

$$\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})] \leq \mathbf{E}[\mathsf{Top}_\ell(\overline{Y})] + \mathbf{E}[\mathsf{Top}_\ell(\widetilde{Y})] \leq 106\, \ell t_\ell.$$

Therefore, σ is an $O(1)$-approximate solution to the given instance of $\mathsf{StochTop}_\ell\mathsf{LB}$. ■

6.2 Unrelated Machines

In the unrelated-machines setting, the stochastic processing time of a job j on machine i is denoted X_{ij}.[1] Our main result here is the following.

Theorem 6.4.
There is an $O(1)$-approximation algorithm for stochastic Top_ℓ-norm load balancing on unrelated machines with arbitrary job-size distributions.

Our approximation algorithm for the unrelated-machines setting is much more technical: we solve a linear program (LP) to obtain a fractional assignment, which is subsequently rounded to an integral assignment that is within a constant-factor of the "cost" of the fractional solution. In the following three sections, we state the LP-relaxation for $\mathsf{StochTop}_\ell\mathsf{LB}$, describe an LP-rounding strategy, and analyze the approximation quality of the rounded solution.

6.2.1 LP-Relaxation

Recall that $\mathsf{OPT}_\ell := \mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})]$ denotes the objective value of an optimal assignment σ^*. For notational convenience, we use the shorthand $j \mapsto i$ to denote that job j is assigned to machine i under some arbitrary but fixed assignment.

Similar to our approach for the identical-machines setting, we work with a guess $t_\ell \in \mathbb{R}_{\geq 0}$ for OPT_ℓ/ℓ. Our feasibility LP, which we denote ($\mathsf{LP}(\ell, t_\ell)$), seeks a fractional assignment with objective value $O(\ell t_\ell)$. The guess t_ℓ gives rise to exceptional and truncated sub-instances consisting of job-size variables of the form $X_{ij}^{\geq t_\ell}$ and $X_{ij}^{<t_\ell}$, respectively.

[1] Recall our assumption that the input consists of joint distributions of $\{X_{ij}\}_{i\in[m]}$ for each job j. From the description of our algorithms it will be clear that correlations between X_{ij} and $X_{i'j}$ can be ignored: it suffices to assume that the input consists of marginal distributions $\{X_{ij}\}_{i,j}$ that is often easier to represent succinctly than the joint distributions $\{(X_{ij})_i\}_j$.

By triangle inequality and Lemma 3.5, it suffices for our LP to model the constraints $\sum_i \mathbf{E}[\overline{Y}_i^{\geq t_\ell}] = O(\ell t_\ell)$ and $\sum_i \mathbf{E}[\widetilde{Y}_i^{\geq t_\ell}] = O(\ell t_\ell)$, where $\overline{Y}$ and $\widetilde{Y}$ are the "load vectors" induced by the fractional assignment in the exceptional and truncated sub-instances. Before delving into the specifics of our LP, we remark that $(\mathsf{LP}(\ell, t_\ell))$ is feasible whenever $t_\ell \geq 2 \cdot \mathsf{OPT}_\ell/\ell$ (see Claim 6.5). So, using binary search we can find a t_ℓ such that $(\mathsf{LP}(\ell, t_\ell))$ is feasible and $t_\ell = O(\mathsf{OPT}_\ell/\ell)$. In Section 6.2.2, we describe a rounding strategy that converts any fractional solution for $(\mathsf{LP}(\ell, t_\ell))$ to an assignment σ satisfying $\sum_i \mathbf{E}[\overrightarrow{\mathsf{load}}^\sigma(i)^{\geq t_\ell}] = O(\ell t_\ell)$, which implies that $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] = O(\ell t_\ell)$. Overall, this leads to an $O(1)$-approximation algorithm.

As is standard in load-balancing LPs, we introduce variables $\{z_{ij}\}_{i,j}$ indicating if job j is assigned to machine i, so z belongs to the assignment polytope

$$\mathcal{Q}^{\mathsf{asgn}} := \Big\{ z \in \mathbb{R}_{\geq 0}^{m \times J} : \sum_{i \in [m]} z_{ij} = 1 \quad \forall i \in [m] \Big\}$$

The constraint for the exceptional sub-instance is fairly straightforward: we simply linearize the constraint $\sum_{j \in J} \mathbf{E}[X_{\sigma(j),j}^{\geq t_\ell}] = O(\ell t_\ell)$ to get (6.1). Handling the truncated sub-instance is more complicated. For each machine i, let $L_i := \sum_{j: j \mapsto i} X_{ij}^{<t_\ell}/t_\ell$ denote the scaled load on machine i due to truncated jobs assigned to it. We use the LP variable ξ_i to model $\mathbf{E}[L_i^{\geq 1}]$, so that $t_\ell \xi_i$ models $\mathbf{E}[\big(\sum_{j:j\mapsto i} X_{ij}^{<t_\ell}\big)^{\geq t_\ell}]$. Since L_i is a sum of independent $[0, 1]$-random variables, Lemma 5.10 yields various lower bounds on $\mathbf{E}[L_i^{\geq 1}]$ in terms of effective sizes of truncated jobs assigned to i. For any parameter $\lambda_i \geq 1$ we have:

$$\mathbf{E}[L_i^{\geq 1}] \geq \frac{\sum_{j:j\mapsto i} \beta_{\lambda_i}(X_{ij}^{<t_\ell}/4t_\ell) - 6}{4\lambda_i}$$

In Section 6.1.2, when the machines were identical, it made sense to set all λ_i's to a uniform value ($= m/\ell$). However, when machines are unrelated, it is apriori unclear what the *right* choice of λ_i is. To circumvent this issue, we essentially include constraints of the above form for all λ_i's; once we solve the resulting LP, we will utilize the lower bounds arising from a suitable λ_i value for each machine i. More precisely, we include constraints (see (6.2) and (6.3)) of the above form for a sufficiently large collection of λ_i's so that one of them is close enough to the right choice. We remark that the idea of using different λ's is also present in Gupta et al. [10], albeit it is exploited in a much-more limited fashion therein. Finally, we include the constraint (6.4) to model $\sum_i \mathbf{E}[L_i^{\geq 1}] \leq \ell$. In all, for any positive parameter t_ℓ, our LP-relaxation $(\mathsf{LP}(\ell, t_\ell))$ is given by (6.1)–(6.5).

$$
(\mathsf{LP}(\ell, t_\ell)) \qquad
\begin{aligned}
\sum_{i\in[m], j\in J} \mathbf{E}[X_{ij}^{\geq t_\ell}] z_{ij} &\leq \ell t_\ell && && (6.1)\\
\sum_{j\in J} \mathbf{E}[X_{ij}^{<t_\ell}/t_\ell] z_{ij} - 1 &\leq \xi_i && \forall i \in [m] && (6.2)\\
\frac{\sum_{j\in J} \beta_\lambda(X_{ij}^{<t_\ell}/4t_\ell) z_{ij} - 6}{4\lambda} &\leq \xi_i && \forall i \in [m], \forall \lambda \in \{2, \ldots, 100m\} && (6.3)\\
\sum_{i\in[m]} \xi_i &\leq \ell && && (6.4)\\
\xi \geq 0, z \in & \mathcal{Q}^{\mathsf{asgn}}. && && (6.5)
\end{aligned}
$$

Claim 6.5. $(\mathsf{LP}(\ell, t_\ell))$ *is feasible for any* $t_\ell \geq 2 \cdot \mathsf{OPT}_\ell/\ell$.

Proof. We introduce some notation for convenience. Let $Y := \overrightarrow{\mathsf{load}}^{\sigma^*}$. Let $\overline{Y}$ and $\widetilde{Y}$ denote the exceptional and truncated load vectors induced by σ^*. That is, $\overline{Y}_i = \sum_{j:\sigma^*(j)=i} X_{ij}^{\geq t_\ell}$ and $\widetilde{Y}_i = \sum_{j:\sigma^*(j)=i} X_{ij}^{<t_\ell}$ for any i. Also, let $L_i := \widetilde{Y}_i/t_\ell$ denote the truncated load on machine i scaled down by t_ℓ. Consider the solution (z^*, ξ^*) induced by σ^*, where z^* is the indicator vector of σ^*, and $\xi_i^* := \mathbf{E}[L_i^{\geq 1}]$ for any i. By our assumption on t_ℓ, the contrapositive of Lemma 3.6 implies that $\sum_i \mathbf{E}[Y_i^{\geq t_\ell}] \leq \ell t_\ell$. Thus, we trivially have $\sum_i \mathbf{E}[\overline{Y}_i^{\geq t_\ell}] \leq \ell t_\ell$ and $\sum_i \mathbf{E}[\widetilde{Y}_i^{\geq t_\ell}] \leq \ell t_\ell$.

The exceptional-jobs constraint (6.1) holds because the support of any $\overline{Y}_i$ lies in $\{0\} \cup [t_\ell, \infty)$, and hence

$$
\sum_{i,j} \mathbf{E}[X_{ij}^{\geq t_\ell}] z_{ij}^* = \sum_{j\in J} \mathbf{E}[X_{\sigma^*(j),j}^{\geq t_\ell}] = \sum_{i\in[m]} \mathbf{E}[\overline{Y}_i] = \sum_{i\in[m]} \mathbf{E}[\overline{Y}_i^{\geq t_\ell}] \leq \ell t_\ell.
$$

Next, Lemma 5.10 applied to the composite random variable L_i, which is a sum of independent $[0,1]$-bounded random variables, shows that constraints (6.2) and (6.3) hold. Finally, $t_\ell \xi_i^* = \mathbf{E}[\widetilde{Y}^{\geq t_\ell}]$ for each i, and the upper bound of ℓt_ℓ on $\sum_i \mathbf{E}[\widetilde{Y}^{\geq t_\ell}]$ implies that $\sum_i \xi_i^* \leq \ell$. ■

6.2.2 LP-Rounding Strategy

In this section we describe our rounding strategy that takes an arbitrary fractional solution to $(\mathsf{LP}(\ell, t_\ell))$, for some $t_\ell > 0$, and rounds it to an integral assignment with objective value $O(\ell t_\ell)$. We fix some feasible fractional solution $(\overline{z}, \overline{\xi})$ to $(\mathsf{LP}(\ell, t_\ell))$.

We start with some intuition on what properties are desirable in a rounded solution. Ideally, we would like to round $\overline{z}$ while only violating budget constraints (6.1), (6.2) and (6.3) up to $O(1)$ factors; this would immediately yield the $O(1)$-approximation guarantee. Unfortunately, this is prohibitive: Usually, LP-rounding algorithms are harder to design and analyze as the number of constraints grow. However, as in Gupta et al. [10], it suffices to work with only $O(m)$ many budget constraints. For each machine i, we carefully choose a single budget constraint from among its truncated jobs constraints (6.2) and (6.3), and round the fractional assignment $\overline{z}$ to obtain an assignment σ such that: (i) these budget constraints for the machines are satisfied approximately; and (ii) the total contribution from the exceptional jobs (across all machines) remains at most ℓt_ℓ. The rounding step amounts to rounding a fractional solution to an instance of the *generalized assignment problem* (GAP), for which we can utilize the algorithm of [40], or use the iterative-rounding result from Theorem 2.32.

The budget constraint that we include for a machine is tailored to ensure that the total $\beta_\lambda(X_{ij}^{<t_\ell}/4t_\ell)$-effective load on a machine under the assignment σ is not too large; due to Lemma 5.7, this will imply a suitable bound on $\mathbf{E}[L_i^{\geq \Omega(1)}]$, where $L_i = \sum_{j:j\mapsto i} X_{ij}^{<t_\ell}/4t_\ell$. Ideally, for each machine i we would like to choose constraint (6.3) for $\lambda_i = 1/\overline{\xi}_i$. This yields $\sum_j \beta_{\lambda_i}(X_{ij}^{<t_\ell}/4t_\ell)\overline{z}_{ij} \leq 4\lambda_i\overline{\xi}_i + 6 = 10$. So if this budget constraint is approximately satisfied in the rounded solution, say with RHS equal to some *constant* b, then Lemma 5.7 roughly gives us $\mathbf{E}[L_i^{\geq b+1}] \leq (b+3)/\lambda_i = (b+3)\overline{\xi}_i$. This in turn implies that

$$\sum_i \mathbf{E}\left[\Big(\sum_{j:j\mapsto i} X_{ij}^{<t_\ell}\Big)^{\geq 4(b+1)t_\ell}\right] = 4t_\ell \cdot \sum_i \mathbf{E}[L_i^{\geq b+1}] \leq 4t_\ell(b+3)\cdot \sum_i \overline{\xi}_i \leq 4(b+3)\ell t_\ell,$$

where the last inequality follows due to (6.4). The upshot is that

$$\sum_i \mathbf{E}\Big[\big(\sum_{j:j\mapsto i} X_{ij}^{<t_\ell}\big)^{\geq \Omega(t_\ell)}\Big] = O(\ell t_\ell);$$

coupled with the fact that $\sum_j \mathbf{E}[X_{\sigma(j),j}^{\geq t_\ell}] \leq \ell t_\ell$, we obtain $\sum_i \mathbf{E}\big[\overrightarrow{\mathsf{load}}^\sigma(i)^{\geq \Omega(t_\ell)}\big] = O(\ell t_\ell)$, and hence $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] = O(\ell t_\ell)$.

A slight complication is that $1/\overline{\xi}_i$ need not be an integer in $[100m]$, so we modify the choice of λ_i's appropriately to deal with this.

We remark that whereas we work with a more general norm than ℓ_∞, our entire approach—polynomial-size LP-relaxation, rounding algorithm, and analysis—is in fact simpler and cleaner than the one used in [10] for the special case of Top_1 norm. Our savings can be traced to the fact that we leverage the notion of effective size in a more powerful way, by utilizing it at multiple scales to obtain lower bounds on $\mathbf{E}[(\sum_{j\mapsto i} X_{ij}^{<t_\ell}/t_\ell)^{\geq 1}]$ (Lemma 5.10). Our rounding algorithm is summarized below.

T1. Define $\mathcal{M}_{\mathsf{low}} := \{i \in [m] : \overline{\xi}_i < 1\}$, and let $\mathcal{M}_{\mathsf{hi}} := [m] \setminus \mathcal{M}_{\mathsf{low}}$. For every $i \in [m]$, define $\lambda_i := \min(\lceil 1/\overline{\xi}_i \rceil, 100m) \in \{2, \ldots, 100m\}$ if $i \in \mathcal{M}_{\mathsf{low}}$, and $\lambda_i := 1$ otherwise.

T2. Consider the following LP that has one budget constraint for each machine i corresponding to the truncated jobs budget constraint picked from either (6.2) or (6.3) depending on the parameter λ_i. We rearrange the constraints for clarity.

$$\min \quad \sum_{i\in[m],j\in J} \mathbf{E}[X_{ij}^{\geq t_\ell}]\eta_{ij} \tag{Aux-Top_ℓ-LP}$$

$$\text{s.t.} \quad \sum_{j\in J} \beta_{\lambda_i}(X_{ij}^{<t_\ell}/4t_\ell)\eta_{ij} \leq 14 \qquad \forall\, i \in \mathcal{M}_{\mathsf{low}} \tag{6.6}$$

$$\sum_{j\in J} \mathbf{E}[X_{ij}^{<t_\ell}/t_\ell]\eta_{ij} \leq \overline{\xi}_i + 1 \qquad \forall\, i \in \mathcal{M}_{\mathsf{hi}} \tag{6.7}$$

$$\eta \in \mathcal{Q}^{\mathsf{asgn}} .$$

Clearly, $\overline{z}$ is a feasible solution to (Aux-Top_ℓ-LP) with objective value at most ℓt_ℓ. Observe that $\mathcal{Q}^{\mathsf{asgn}}$ is the base polytope of the partition matroid encoding that each job is assigned to at most one machine. We round $\overline{z}$ to obtain an integral assignment σ, either by using GAP rounding, or by invoking Theorem 2.32.

6.2.3 Analysis

We now show that $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] = O(\ell t_\ell)$. We first note that Theorem 2.32 directly shows that σ satisfies constraints (6.6) and (6.7) with an additive violation of at most 1, and the total contribution from exceptional jobs is at most ℓt_ℓ.

Claim 6.6. *The assignment σ satisfies:*

(a) $\sum_{j\in J} \mathbf{E}[X_{\sigma(j),j}^{\geq t_\ell}] \leq \ell t_\ell$.

(b) $\sum_{j\in J:\sigma(j)=i} \beta_{\lambda_i}(X_{ij}^{<t_\ell}/4t_\ell) \leq 15$ *for any machine* $i \in \mathcal{M}_{\mathsf{low}}$.

(c) $\sum_{j\in J:\sigma(j)=i} \mathbf{E}[X_{ij}^{<t_\ell}/t_\ell] \leq \overline{\xi}_i + 2$ *for any machine* $i \in \mathcal{M}_{\mathsf{hi}}$.

Proof. This follows directly from Theorem 2.32 by noting that the parameter ν, denoting an upper bound on the column sum of a variable, is at most 1, and since $\bar{z}$ is a feasible solution to (Aux-Top_ℓ-LP) of objective value at most ℓt_ℓ. ■

Next, we bound $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)]$ by bounding the expected Top_ℓ-norm of the load induced by three different sources: exceptional jobs, truncated jobs in $\mathcal{M}_{\mathsf{low}}$, and truncated jobs in $\mathcal{M}_{\mathsf{hi}}$. Let $j : j \mapsto i$ be a shorthand for $\{j \in J : \sigma(j) = i\}$. Observe that

$$\overrightarrow{\mathsf{load}}^\sigma = Y^{\mathsf{excep}} + Y^{\mathsf{low}} + Y^{\mathsf{hi}},$$

where $Y_i^{\mathsf{excep}} := \sum_{j:j\mapsto i} X_{ij}^{\geq t_\ell}$,

$$Y_i^{\mathsf{low}} := \begin{cases} \sum_{j:j\mapsto i} X_{ij}^{<t_\ell}; & \text{if } i \in \mathcal{M}_{\mathsf{low}} \\ 0; & \text{otherwise} \end{cases} \qquad \text{and} \qquad Y_i^{\mathsf{hi}} := \begin{cases} \sum_{j:j\mapsto i} X_{ij}^{<t_\ell}; & \text{if } i \in \mathcal{M}_{\mathsf{hi}} \\ 0; & \text{otherwise} \end{cases}$$

All three random vectors follow a product distribution on $\mathbb{R}^m_{\geq 0}$. By the triangle inequality, it suffices to bound the expected Top_ℓ norm of each vector by $O(\ell t_\ell)$. It is easy to bound even the expected Top_m norms of Y^{hi} and Y^{excep} (Lemma 6.7); to bound $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{low}})]$ (Lemma 6.8), we utilize properties of effective sizes.

Lemma 6.7. *We have (i)* $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{excep}})] \leq \ell t_\ell$, *and (ii)* $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{hi}})] \leq 3\ell t_\ell$.

Proof. Part (i) follows immediately from Claim 6.6(a) since:

$$\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{excep}})] \leq \mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{excep}})] = \sum_{j\in J} \mathbf{E}[X_{\sigma(j),j}^{\geq t_\ell}] \leq \ell t_\ell.$$

For part (ii), we utilize Claim 6.6(c) which gives $\mathbf{E}[Y_i^{\mathsf{hi}}] = \sum_{j:j\mapsto i} \mathbf{E}[X_{ij}^{<t_\ell}] \leq t_\ell(\bar{\xi}_i + 2)$ for any machine $i \in \mathcal{M}_{\mathsf{hi}}$. Note that $|\mathcal{M}_{\mathsf{hi}}| \leq \ell$ because $\bar{\xi}_i \geq 1$ for any $i \in \mathcal{M}_{\mathsf{hi}}$ and $\sum_{i\in[m]} \bar{\xi}_i \leq \ell$ by the LP constraint (6.4). Therefore,

$$\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{hi}})] \leq \mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{hi}})] \leq t_\ell \cdot \sum_{i\in\mathcal{M}_{\mathsf{hi}}} (\bar{\xi}_i + 2) \leq 3\ell t_\ell.$$ ■

Lemma 6.8. *We have* $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{low}})] \leq 140\, \ell t_\ell$.

Proof. Let $W := Y^{\mathsf{low}}/4t_\ell$. Observe that it suffices to show that

$$\sum_{i\in[m]} \mathbf{E}[W_i^{\geq 16}] = \sum_{i\in\mathcal{M}_{\mathsf{low}}} \mathbf{E}[W_i^{\geq 16}] \leq 19\, \ell$$

holds. Then by Lemma 3.4 we immediately get $\mathbf{E}[\mathsf{Top}_\ell(W)] \leq 35\,\ell$, or equivalently, $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{low}})] \leq 140\,\ell t_\ell$. To this end, Claim 6.6(b) gives $\beta_{\lambda_i}(W_i) \leq 15$ for any machine $i \in \mathcal{M}_{\mathsf{low}}$, where $\lambda_i = \min(100m, \lceil 1/\bar{\xi}_i \rceil) \geq 2$. Using Lemma 5.7 we get:

$$\mathbf{E}\big[W_i^{\geq 16}\big] \leq 18/\lambda_i = 18\,\max\Big(\frac{1}{100m}, \frac{1}{\lceil 1/\bar{\xi}_i \rceil}\Big) \leq 18\,\bar{\xi}_i + \frac{18}{100m}.$$

Summing over all machines in $\mathcal{M}_{\mathsf{low}}$ gives $\sum_{i\in\mathcal{M}_{\mathsf{low}}} \mathbf{E}[W_i^{\geq 16}] \leq 19\ell$. ∎

Combining the two lemmas above yields the following result.

Theorem 6.9. *The assignment σ satisfies $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] \leq 144\,\ell t_\ell$.*

6.2.4 Our Algorithm

We now give a full description of our LP-based approximation algorithm for $\mathsf{StochTop}_\ell\mathsf{LB}$ on unrelated machines.

Proof of Theorem 6.4. Given Theorem 6.9 and Claim 6.5, it is clear that if we work with $t_\ell = O(\mathsf{OPT}_\ell/\ell)$ such that $(\mathsf{LP}(\ell, t_\ell))$ is feasible and run our LP-rounding algorithm, then we obtain an $O(1)$-approximate assignment. As is standard, we can find such a t_ℓ, within a $(1+\varepsilon)$-factor, via binary search. To perform this binary search, we show that we can come up with an upper bound UB such that $\mathsf{UB}/m \leq \mathsf{OPT} \leq \mathsf{UB}$. We show this even in the general setting where we have an arbitrary monotone, symmetric norm f.

Lemma 6.10. *Let f be a normalized, monotone, symmetric norm, and OPT_f denote the optimal value for a given instance of the stochastic f-norm load balancing problem. Define $\mathsf{UB} := \sum_{j\in J}\big(\min_{i\in[m]} \mathbf{E}[X_{ij}]\big)$, which can be easily computed from the input data. We have $\mathsf{UB}/m \leq \mathsf{OPT}_f \leq \mathsf{UB}$.*

Proof. Notice that UB is the optimal value of stochastic Top_m-norm load-balancing, i.e., it is the objective value of the assignment that minimizes the sum of the expected machine loads. By our assumption that f is normalized, we have $\mathsf{Top}_1(\cdot) \leq f(\cdot)$ (see Lemma 2.2). Therefore, for any assignment $\sigma' : J \to [m]$ we have:

$$\frac{1}{m}\cdot \mathbf{E}[\mathsf{Top}_m(\overrightarrow{\mathsf{load}}^{\sigma'})] \leq \mathbf{E}[\mathsf{Top}_1(\overrightarrow{\mathsf{load}}^{\sigma'})] \leq \mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma'})] \leq \mathbf{E}[\mathsf{Top}_m(\overrightarrow{\mathsf{load}}^{\sigma'})].$$

Taking the minimum over all assignments, and plugging in $\mathsf{UB} = \min_{\sigma'} \mathbf{E}[\mathsf{Top}_m(\overrightarrow{\mathsf{load}}^{\sigma'})]$, as noted above, it follows that $\mathsf{UB}/m \leq \min_{\sigma':J\to[m]} \mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma'})] = \mathsf{OPT}_f \leq \mathsf{UB}$. ∎

Thus, if we binary search in the interval $[0, 2 \cdot \mathsf{UB}/\ell]$, for any $\varepsilon > 0$, we can find in $\mathrm{poly}(\log(m/\varepsilon))$ iterations a scalar $t_\ell \le 2 \cdot \mathsf{OPT}_\ell/\ell + \varepsilon \cdot \mathsf{UB}/m^2 \le (2+\varepsilon)\mathsf{OPT}_\ell/\ell$ such that $(\mathsf{LP}(\ell, t_\ell))$ is feasible. By Theorem 6.9, we obtain an assignment whose expected Top_ℓ norm is at most $144\,\ell t_\ell \le (288 + O(\varepsilon))\,\mathsf{OPT}_\ell$. ■

Chapter 7

Stochastic Load Balancing with General Monotone Symmetric Norms

In this chapter we design approximation algorithms for stochastic min-norm load balancing for general monotone symmetric norms. Recall that the goal in $\mathsf{StochNormLB}$ is to obtain an assignment $\sigma : J \to [m]$ of jobs to machines that minimizes $\mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma})]$, for a given monotone, symmetric norm $f : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$. We reserve σ^* to denote an optimal solution and $\mathsf{OPT}_f := \mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma^*})]$ to denote the optimal solution value. We drop the subscript f in OPT_f whenever the norm is clear from the context.

The starting point for our algorithms in this chapter is approximate stochastic majorization: to obtain an $O(\alpha)$-approximate assignment σ, it suffices to ensure that for all $\ell \in \mathsf{POS} := \{1, 2, 4, \dots, 2^{\lfloor \log_2 m \rfloor}\}$, $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})] \leq \alpha \cdot \mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})]$ holds. We refer the reader to Chapter 4 for a quick refresher on Theorems 4.1, 4.14 and 4.15. In Chapter 6 we designed approximation algorithms for $\mathsf{StochTop}_\ell\mathsf{LB}$ that were based on a reduction to deterministic makespan-minimization load balancing. Thus, a natural approximation strategy for $\mathsf{StochNormLB}$ involves a reduction to makespan-minimization load balancing when jobs have multidimensional sizes, one size for every $\ell \in \mathsf{POS}$. This latter problem is known as *vector scheduling* in Scheduling Theory literature, but we refer to it as *vector load balancing* to stay consistent with our terminology.

We divide this chapter into two sections. In Section 7.1, we restrict ourselves to the identical-machines setting and give a simple $O(\log \log m)$-approximation algorithm. This algorithm is extremely easy to describe and the analysis only uses Chernoff upper tail bounds. In fact, we show that the assignment produced by our algorithm is *simultaneously* an $O(\log\log m)$-approximate solution for *every* monotone symmetric norm. This simultaneous every-monotone-symmetric-norm approximation guarantee is a common feature of our algorithms in Chapters 7 and 8 for the identical-machines setting.

In Section 7.2, we give an improved $O(\log\log m/\log\log\log m)$-approximation algorithm which holds even in the unrelated-machines setting. At a high level, our algorithm is based on applying a sophisticated randomized-rounding procedure to an LP solution for $\mathsf{StochNormLB}$. The LP that we use is a straightforward generalization of the LP for $\mathsf{StochTop}_\ell\mathsf{LB}$ from Section 6.2: we essentially include constraints from ($\mathsf{LP}(\ell, t_\ell)$) for all $\ell \in \mathsf{POS}$. A useful property of our LP is that the column sums of the constraint matrix are bounded by $D := O(\log m)$. The powerful randomized-rounding result of Harris and Srinivasan [14] for column-sparse assignment problems with packing constraints gives the desired $O(\log D/\log\log D)$-approximate assignment. While the $O(\log\log m/\log\log\log m)$-approximation guarantee is both stronger, and holds in a more general machine environment than the $O(\log\log m)$-approximation guarantee, the overall algorithm is rather opaque and involved, since we use the result of [14] (which is a bit complicated) in a black-box fashion.

7.1 Identical Machines

Recall that in the identical machines setting, the job-size distribution for job j (on any machine $i \in [m]$) is denoted X_j. Our main result in this section is a randomized $O(\log\log m)$-approximation for $\mathsf{StochNormLB}$ when machines are identical.

Theorem 7.1.
We can compute (in polynomial time) an assignment $\sigma : J \to [m]$ such that, with probability at least $2/3$, σ is an $O(\log\log m)$-approximate assignment to the given instance of $\mathsf{StochNormLB}$.

In Section 6.1, we saw an $O(1)$-approximation algorithm for $\mathsf{StochTop}_\ell\mathsf{LB}$ (for some fixed $\ell \in [m]$) on identical machines via a reduction to deterministic makespan-minimization load balancing. We recall the high-level approximation strategy. We used binary search to compute a scalar t_ℓ such that: (i) $t_\ell = O(\mathsf{OPT}_\ell/\ell)$; Let OPT_ℓ be the optimal value when the objective function is a Top_ℓ norm; (ii) $\sum_{j\in J}\mathbf{E}[X_j^{\geq t_\ell}] \leq \ell t_\ell$; and (iii) $\sum_{j\in J}\beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell) \leq 10\,m$. Since the load vector, induced by any assignment, in the exceptional sub-instance has expected Top_ℓ-norm bounded by $O(\ell t_\ell)$ (see Lemma 6.2 (a)), we simply ignored the

contribution from exceptional jobs. For the truncated sub-instance, obtaining an $O(1)$-approximate assignment reduces to evenly distributing (up to $O(1)$ additive factors) the total $\beta_{m/\ell}$-effective-size over all machines (see Lemma 6.3).

The above approximation strategy can be naturally extended to handle $\mathsf{StochNormLB}$ via a reduction to vector load balancing. For each $\ell \in \mathsf{POS}$, we first obtain t_ℓ such that $t_\ell = O(\mathsf{OPT}_\ell/\ell)$. Next, each $\ell \in \mathsf{POS}$ gives rise to exceptional and truncated sub-instances w.r.t. the threshold t_ℓ. Like before, for any assignment σ the objective value arising from exceptional jobs w.r.t. threshold t_ℓ can be charged to $O(\ell t_\ell) = O(\mathsf{OPT}_\ell) = O(\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})])$. So it suffices to simultaneously handle the contribution from truncated jobs w.r.t. thresholds t_ℓ for each $\ell \in \mathsf{POS}$. By our choice of t_ℓ's, we have $\sum_{j\in J} \beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell) \le 10\,m$ for every $\ell \in \mathsf{POS}$. If we can obtain an assignment σ such that for all $\ell \in \mathsf{POS}$ and machine $i \in [m]$, $\sum_{j:\sigma(j)=i} \beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell) = O(\alpha)$ holds for some α, then by Lemma 6.3 we get $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})] = O(\alpha \cdot \mathsf{OPT}_\ell) = O(\alpha \cdot \mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})]$. Thus, by Theorem 4.15 (i), σ is an $O(\alpha)$-approximate solution to the given instance of $\mathsf{StochNormLB}$.

Note that the quality of approximation, α, yielded by the above approach depends on the approximation ratio that we can get for $O(\log m)$-dimensional vector load balancing instances with respect to a lower bound that, loosely speaking, measures the maximum average load in any dimension. We will shortly define the d-dimensional vector load balancing problem and state this lower bound precisely, and give a simple $O(\log d)$-approximation algorithm for it relative to this lower bound. Thus, this approach gives an $O(\log\log m)$-approximation for $\mathsf{StochNormLB}$.

In the preceding discussion, observe that the choice of threshold t_ℓ is based only on the optimal assignment for the Top_ℓ objective. Furthermore, the crude upper bound argument on $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})]$ that we described above is also w.r.t. OPT_ℓ. Thus, our approximation guarantee holds simultaneously for all monotone symmetric norms. We formalize this in Section 7.1.3.

Theorem 7.2.
Consider n stochastic jobs $\{X_j\}_{j\in J}$ and m identical machines. We can compute (in polynomial time) an assignment $\sigma : J \to [m]$ such that, with probability at least $2/3$, σ is simultaneously an $O(\log\log m)$-approximation to all Stoch-h-LB instances where $h : \mathbb{R}^m_{\ge 0} \to \mathbb{R}_{\ge 0}$ is a monotone, symmetric norm.

The randomization above stems from our randomized guarantee for vector load balancing. We remark that the probability of success in Theorem 7.2 is of a detectable (polytime-verifiable) event (see Remark 7.1). Therefore, we can lower the failure probability to δ, for any $\delta > 0$, by repeating the algorithm in Theorem 7.2 $O(\log(1/\delta))$ times.

7.1.1 The Vector Load Balancing Problem

In the d-dimensional vector load balancing (a.k.a. vector scheduling) problem, we have a set J of n jobs that are to be processed on exactly one of m identical machines. Each job $j \in J$ has a processing-time (or size) *vector* $p_j = (p_{j,1}, \ldots, p_{j,d}) \in \mathbb{R}^d_{\geq 0}$, where $p_{j,r}$ denotes the size of job j in dimension $r \in [d]$. For notational convenience, we overload some of the notation used in stochastic min-norm load balancing. An assignment $\sigma : J \to [m]$ of jobs to machines induces an $(m \times d)$-dimensional load vector $\overrightarrow{\mathsf{load}}^\sigma$ (one entry per machine-dimension pair): for machine $i \in [m]$ and dimension $r \in [d]$, the load in the r^{th} dimension of machine i is $\overrightarrow{\mathsf{load}}^\sigma(i,r) := \sum_{j \in J: \sigma(j)=i} p_{j,r}$. The *makespan objective* of an assignment σ is defined as $\max_{i\in[m], r\in[d]} \overrightarrow{\mathsf{load}}^\sigma(i,r)$, i.e., the maximum load across all machines and all dimensions. The goal in vector load balancing is to find an assignment σ that minimizes the makespan.

Chekuri and Khanna [7] were the first to consider the vector scheduling problem, and they gave an $O(\log^2 d)$-approximation algorithm. Meyerson, Roytman and Tagiku [32] improved the approximation guarantee to a factor $O(\log d)$; in fact, they gave an $O(\log d)$-competitive algorithm for the online version of the problem where jobs arrive one at a time and have to be assigned irrevocably to a machine on arrival. The current best approximation for the problem is an $O(\log d/\log\log d)$-competitive algorithm by Im, Kell, Kulkarni and Panigrahi [21], and an offline $O(\log d/\log\log d)$-approximation algorithm by Harris and Srinivasan [14] that also works for the more-general setting with unrelated machines. In terms of hardness, very recently, Sai Sandeep [38] (see [39] for a full version) showed that, for any $\epsilon > 0$, (offline) vector scheduling is hard to approximate to a factor $O((\log d)^{1-\epsilon})$ under some complexity theoretic assumptions.

A natural lower bound on the optimal makespan for vector load balancing is given by:

$$lb := \max\left\{ \max_{\substack{\text{job } j\in J \\ \text{dimension } r\in[d]}} p_{j,r}, \quad \frac{1}{m}\cdot \max_{\text{dimension } r\in[d]} \sum_{\text{job } j\in J} p_{j,r} \right\}. \tag{7.1}$$

The first term above arises because each job has to assigned to some machine, and the second term arises because in each dimension r, a total load of $\sum_j p_{j,r}$ is distributed among m machines. Note that the expression for lb is simply a maximum over all dimensions of the classical lower bound on the optimal makespan for the 1-dimensional case.

Motivated by applications to stochastic min-norm load balancing, as alluded to earlier, we want to design approximation algorithms for vector load balancing with respect to the natural lower bound lb. Our main result on vector load balancing is a simple lb-relative $O(\log d)$-approximation algorithm.

Theorem 7.3.
There is a randomized algorithm for the d-dimensional vector scheduling problem that computes an assignment σ with makespan at most $O(\log d) \cdot lb$. The algorithm runs in time polynomial in the input size, and succeeds in finding an approximate assignment with probability at least $2/3$.

Before proving Theorem 7.3, we make a few remarks comparing our result to prior work. The $O(\log d)$ approximation guarantee of Meyerson et al. [32] is also with respect to the lower bound lb, although it is not explicitly stated in this form. While their guarantee is deterministic and holds also in the online setting, our chief notable feature is the simplicity of our algorithm and analysis.[1] From a pure approximation-standpoint, the $O(\log d/\log\log d)$-approximation of [21] is of course better, but from the description of their algorithm it is unclear if their guarantee holds with respect to the lower bound lb; also, their algorithm and analysis are significantly more involved. In Section 7.2, we state an improved $O(\log d/\log\log d)$-approximation algorithm for vector load balancing that works for the more-general setting of unrelated machines (the size of a job j on machine i in dimension r is $p_{i,j,r} \in \mathbb{R}_{\geq 0}$). This result is due to Harris and Srinivasan [14], and its specialization to identical machines works with the lower bound lb. However, the overall algorithm and its analysis are quite sophisticated. Finally, we note that the hardness result in [38] essentially shows that our approximation guarantee is tight up to $\text{poly}(\log\log d)$ factors.

7.1.2 Approximation Algorithm for Vector Load Balancing

In this section, we describe our lb-relative approximation algorithm for vector load balancing. To keep the notation simple, we reserve $i \in [m]$ to index the machine-set, $j \in J$ to index the job-set, and $r \in [d]$ to index the dimensions. By scaling, we may assume without loss of generality that $lb = 1$; this implies $p_{j,r} \in [0,1]$ for all j, r, and $\sum_j p_{j,r} \leq m$ for all r. We prove Theorem 7.3 by proving a slightly stronger result.

Theorem 7.4. *Consider an instance of vector load balancing with $p_{j,r} \in [0,1]$ for all $j \in J, r \in [d]$, and $\sum_{j \in J} p_{j,r} \leq m \log d$ for all $r \in [d]$. There is a randomized algorithm*

[1] The algorithm of Meyerson et al. is also fairly simple—it assigns the recently-arrived job to the machine that leads to the least increase in an exponential potential function—but the analysis is slightly involved.

that produces an assignment σ whose makespan is $O(\log d)$. The algorithm succeeds with probability at least $2/3$.

Our randomized algorithm in Theorem 7.4 is based on finding a subset of jobs $S \subseteq J$ that can all be processed on a single machine by incurring a load of at most $O(\log d)$ in each dimension $r \in [d]$, while ensuring that the residual problem is also a valid sub-instance of the problem with $m-1$ machines. We use Chernoff tail bounds to obtain this subset S.

Lemma 7.5. *Suppose $m \geq 9$ and $d \geq 2$. Consider a random job-set $S \subseteq J$ where job $j \in J$ is independently included in S with probability $q := 9/m$. With probability at least $1/2$, for all $r \in [d]$ we have $\sum_{j \in S} p_{j,r} \leq 18 \log d$ and $\sum_{j \in J \setminus S} p_{j,r} \leq (m-1) \log d$.*

Proof. The proof is based on a straightforward application of Chernoff bounds. By our choice of q, for any dimension r we have $\mathbf{E}[\sum_{j \in S} p_{j,r}] \leq 9 \log d$. Using Lemma 2.28 with $\delta = 1$, for any dimension r we have:

$$\mathbf{Pr}\left[\sum_{j \in S} p_{j,r} > 18 \log d\right] \leq \exp\left(\frac{-9 \log d}{3}\right) \leq \frac{1}{d^3}.$$

Next, we use Chernoff bounds for the lower tail to prove the remaining size-bound on jobs in $J \setminus S$. Let $B := \{r \in [d] : \sum_{j \in J} p_{j,r} > (m-1) \log d\}$. Since $m \geq 9$, for any dimension $r \in B$ we have $\mathbf{E}[\sum_{j \in S} p_{j,r}] \geq (1 - 1/m) \cdot 9 \log d \geq 8 \log d$. Using Lemma 2.28 with $\delta = 7/8$ we get:

$$\mathbf{Pr}\left[\sum_{j \in S} p_{j,r} \leq \log d\right] \leq \exp\left(-\frac{49}{64} \cdot \frac{8 \log d}{2}\right) \leq \frac{1}{d^3}.$$

By union-bound, with probability at least $1/2$, we have $\sum_{j \in S} p_{j,r} \leq 18 \log d$ for all $r \in [d]$, and $\sum_{j \in S} p_{j,r} \geq \log d$ for all $r \in B$. The desired bounds follow. ∎

Proof of Theorem 7.4. We call an instance of vector scheduling as a *valid instance with m machines* if $p_{j,r} \in [0,1]$ for all j, r, and $\sum_j p_{j,r} \leq m \log d$ for all r. Note that in Theorem 7.4 we are given a valid instance with m machines. We will assume that $d \geq 2$, since otherwise a greedy list-scheduling algorithm gives an easy 2-approximate assignment.

Let $N := \lceil \log_2(3m) \rceil = O(\log m)$. We now describe the randomized algorithm for Theorem 7.4. The overall procedure has at most m iterations, and if the algorithm is successful, it produces an assignment with makespan $O(\log d)$. Consider the start of an iteration t for some $t \geq 1$. We have a valid instance with $m - t + 1$ identical machines. If $m - t + 1 = 8$, then we simply assign all remaining jobs to a single machine; note that $\sum_j p_{j,r} \leq 8 \log d$ for all dimensions r, so we are not assigning too much load to this machine in any dimension. Otherwise, $m - t + 1 \geq 9$. Consider random subsets $S_1, \ldots, S_N$ where each $S_\ell \subseteq J, \ell \in [N]$ is obtained by independently including job j in S_ℓ with probability $q := 9/m$. By Lemma 7.5, with probability at most $2^{-N} \leq \frac{1}{3m}$, none of the S_ℓ sets satisfy the conclusion of the lemma; in this case we terminate the algorithm with a failure. Otherwise, for some $\ell \in [N]$, we have for all r, $\sum_{j \in S_\ell} p_{j,r} \leq 18 \log d$ and $\sum_{j \in J \setminus S_\ell} p_{j,r} \leq (m - t) \log d$. In this case, we assign all jobs in S_ℓ to one machine and end the iteration with a valid instance with $m - t$ machines and residual job-set $J \setminus S_\ell$. The probability that the algorithm terminates with a failure in any given iteration is at most $1/3m$, so the overall failure probability of the algorithm is at most $1/3$. In other words, with probability at least $2/3$, we obtain an assignment whose makespan is at most $18 \log d$. ∎

7.1.3 Reduction from StochNormLB to Vector Load Balancing

We prove Theorem 7.2 in this section, thereby also proving Theorem 7.1. We fix some job-set J, m identical machines, and stochastic job-sizes $\{X_j\}_{j \in J}$. Since we will be designing a simultaneous approximation algorithm for all monotone symmetric norms, we reserve h to denote an arbitrary norm from this family. Let σ_h denote an optimal assignment for the Stoch-h-LB instance.

Let $\mathcal{A}$ be a $\gamma(d)$-approximation algorithm for the d-dimensional vector load balancing problem with respect to the natural lower bound lb. As we show below, we use $\mathcal{A}$ in a black-box fashion, so if $\mathcal{A}$ is deterministic (respectively randomized), then our simultaneous-approximation for StochNormLB is also deterministic (respectively randomized). Theorem 7.3 shows that we can take $\gamma(d) = O(\log d)$. We remark that our algorithm crucially leverages the fact that the approximation guarantee for vector load balancing is with respect to the natural lower bound lb.

Proof of Theorem 7.2. We first obtain the right thresholds t_ℓ for each $\ell \in \mathsf{POS}$ satisfying $t_\ell = O(\mathsf{OPT}_\ell/\ell)$, where OPT_ℓ is the optimal objective for the $\mathsf{StochTop}_\ell\mathsf{LB}$ instance. Let $\epsilon > 0$ be a small constant (say, $1/1000$). For each $\ell \in \mathsf{POS}$, we repeat the binary search procedure from Section 6.1.3 to obtain a scalar t_ℓ satisfying the following three conditions:

(i) $\mathsf{OPT}_\ell > \ell t_\ell / 2(1 + \epsilon)$, so that $t_\ell = O(\mathsf{OPT}_\ell/\ell)$.

(ii) $\sum_{j\in J}\mathbf{E}[X_j^{\geq t_\ell}] \leq \ell t_\ell$.

(iii) $\sum_{j\in J}\beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell) \leq 10\,m$.

Consider an instance of $|\mathsf{POS}|$-dimensional vector load balancing on m identical machines where the dimensions are indexed by $\mathsf{POS} = \{1,2,4,\ldots,2^{\lfloor \log_2 m\rfloor}\}$. In this instance, the processing-time vector for job $j\in J$ is $p_j = (p_{j,\ell})_{\ell\in\mathsf{POS}}$, where we define $p_{j,\ell} := \beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell)$. Note that $|\mathsf{POS}| = O(\log m)$, $p_{j,\ell}\in[0,1]$, and by our choice of t_ℓ we have $lb \leq 10$ for the vector load balancing instance.

Using algorithm $\mathcal{A}$, we can obtain an assignment $\sigma : J\to[m]$ with makespan at most $O(\gamma(|\mathsf{POS}|)\cdot lb) = O(\log\log m)$ for the above vector-load-balancing instance. Fix some $\ell\in\mathsf{POS}$ and consider the exceptional and truncated sub-instances of $\mathsf{StochTop}_\ell\mathsf{LB}$ w.r.t. the threshold t_ℓ. Let $\overline{Y}$ and $\widetilde{Y}$ denote the exceptional and truncated load vectors induced by σ in these sub-instances. That is, $\overline{Y}(i) = \sum_{j:\sigma(j)=i} X_j^{\geq t_\ell}$ and $\widetilde{Y}(i) = \sum_{j:\sigma(j)=i} X_j^{<t_\ell}$. By Lemma 6.2 (i) and condition (ii) above, we get $\mathbf{E}[\mathsf{Top}_\ell(\overline{Y})] = O(\ell t_\ell) = O(\mathsf{OPT}_\ell)$. By Lemma 6.3 (i), condition (iii) above, and approximation guarantee of σ for vector load balancing, we get $\mathbf{E}[\mathsf{Top}_\ell(\widetilde{Y})] = O(\log\log m)\cdot\mathsf{OPT}_\ell$. Let $\overrightarrow{\mathsf{load}}^\sigma$ denote the load vector induced by σ in any $\mathsf{Stoch}\text{-}h\text{-}\mathsf{LB}$ instance, i.e., $\overrightarrow{\mathsf{load}}^\sigma(i) = \sum_{j:\sigma(j)=i} X_j$ for any machine i. By definition $\overrightarrow{\mathsf{load}}^\sigma = \overline{Y}+\widetilde{Y}$, so triangle inequality gives:

$$\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] \leq O(\log\log m)\cdot\mathsf{OPT}_\ell \leq O(\log\log m)\cdot\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma_h})],$$

where σ_h is an optimal assignment for the $\mathsf{Stoch}\text{-}h\text{-}\mathsf{LB}$ instance. By Theorem 4.15 (i), we get that $\mathbf{E}[h(\overrightarrow{\mathsf{load}}^\sigma)] = O(\log\log m)\cdot\mathbf{E}[h(\overrightarrow{\mathsf{load}}^{\sigma_h})]$, thereby showing that σ is a simultaneous $O(\log\log m)$-approximate assignment for every monotone, symmetric norm. ■

Remark 7.1. *If the algorithm $\mathcal{A}$ is randomized (as in Theorem 7.4), then we detect if the assignment σ returned by $\mathcal{A}$ satisfies $\sum_{j:\sigma(j)=i} p_{j,\ell} \leq O(\log|\mathsf{POS}|)$ for all dimensions $\ell\in\mathsf{POS}$ and all machines $i\in[m]$, i.e., is the approximation guaranteed by $\mathcal{A}$, and if not, return failure. Then, the success probability in the statement of Theorem 7.2 is the probability of success of $\mathcal{A}$ (which lower bounds the probability of obtaining a simultaneous $O(\log\log m)$-approximation).*

7.2 Unrelated Machines

In this section we consider $\mathsf{StochNormLB}$ on unrelated machines. In this setting, the processing time of a job j on machine i is a nonnegative random variable denoted by

X_{ij}, and we want to find an assignment $\sigma : J \to [m]$ that minimizes $\mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma})]$ for a given monotone, symmetric norm f. Our main result here is an $O(\log\log m/\log\log\log m)$-approximation algorithm.

Theorem 7.6.
There is an $O(\log\log m/\log\log\log m)$-approximation algorithm for stochastic min-norm load balancing on unrelated machines with arbitrary job-size distributions. The approximation guarantee is deterministic, and the expected running time of the algorithm is polynomial in the size of the input.

As usual, our algorithm is guided by approximate stochastic majorization: we want to find an assignment σ such that $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})] \le \alpha \cdot \mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})]$ holds for all $\ell \in \mathsf{POS} = \{1, 2, 4, \dots, 2^{\lfloor \log_2 m \rfloor}\}$, where α is a small factor and σ^* is an optimal solution to the given instance of $\mathsf{StochNormLB}$. It is therefore natural to leverage the insights gained in Section 6.2 from the study of the Top_ℓ-norm problem. Since we need to simultaneously work with all Top_ℓ norms, we now work with a guess t_ℓ of the quantity $O(\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})]/\ell)$ for every $\ell \in \mathsf{POS}$. For each $\vec{t} = (t_\ell)_{\ell \in \mathsf{POS}}$ vector, we write an LP-relaxation ($\mathsf{LP}(\vec{t})$) that generalizes ($\mathsf{LP}(\ell, t_\ell)$), and if it is feasible, we round its feasible solution to obtain an assignment of jobs to machines. We argue that one can limit the number of $\vec{t}$ vectors to consider to a polynomial-size set, so this yields a polynomial number of candidate solutions. We remark that, interestingly, ($\mathsf{LP}(\vec{t})$), the rounding algorithm, and the resulting set of solutions generated, are *independent* of the norm f: they only depend only on the underlying $\vec{t}$ vector. The norm f is used only in the final step to select one of the candidate solutions as the desired near-optimal solution, utilizing the budgeted version of approximate stochastic majorization (see Theorem 4.15 (ii)).

7.2.1 LP-Relaxation

The LP-relaxation we work with is an easy generalization of ($\mathsf{LP}(\ell, t_\ell)$). We have the usual z_{ij} variables encoding a fractional assignment. For each $\ell \in \mathsf{POS}$, there is a different definition of truncated variables $X_{ij}^{<t_\ell}$ and exceptional variables $X_{ij}^{\ge t_\ell}$. Correspondingly, for each index $\ell \in \mathsf{POS}$, we have a separate set of constraints (7.2)–(7.5) involving the z_{ij}'s, a variable $\xi_{i,\ell}$ (that represents ξ_i for the index ℓ, i.e., $\mathbf{E}[(\sum_{j:j\mapsto i} X_{ij}^{<t_\ell}/t_\ell)^{\ge 1}]$), and the guess t_ℓ. For technical reasons that will become clear when we discuss the rounding algorithm (see Claim 7.9), we include additional constraints (7.6), which enforce that a job j cannot be assigned to a machine i if $\mathbf{E}[X_{ij}^{\ge t_1}] > t_1$; observe that this is valid for the optimal integral solution whenever $t_1 \ge 2\,\mathbf{E}[\mathsf{Top}_1(\overrightarrow{\mathsf{load}}^{\sigma^*})]$. This yields the following LP-relaxation.

$$\sum_{i\in[m],j\in J}\mathbf{E}[X_{ij}^{\geq t_\ell}]z_{ij} \leq \ell t_\ell \qquad \forall \ell \in \mathsf{POS} \tag{7.2}$$

$$\sum_{j\in J}\mathbf{E}[X_{ij}^{<t_\ell}/t_\ell]z_{ij} - 1 \leq \xi_{i,\ell} \qquad \forall i \in [m], \ell \in \mathsf{POS} \tag{7.3}$$

$$(\mathsf{LP}(\vec{t})) \qquad \frac{\sum_{j\in J}\beta_\lambda(X_{ij}^{<t_\ell}/4t_\ell)z_{ij} - 6}{4\lambda} \leq \xi_{i,\ell} \qquad \forall i \in [m], \ell \in \mathsf{POS}, \lambda \in \{2,\ldots,100m\} \tag{7.4}$$

$$\sum_{i\in[m]}\xi_{i,\ell} \leq \ell \qquad \forall \ell \in \mathsf{POS} \tag{7.5}$$

$$z_{ij} = 0 \qquad \forall i \in [m], j \in J \text{ with } \mathbf{E}[X_{ij}^{\geq t_1}] > t_1 \tag{7.6}$$

$$\xi \geq 0, z \in \mathcal{Q}^{\mathsf{asgn}}. \tag{7.7}$$

Claim 6.5 easily generalizes to the following.

Claim 7.7. *Let $\vec{t}$ be such that $t_\ell \geq 2\mathbf{E}[\mathsf{Top}_\ell(\mathsf{load}_{\sigma^*})]/\ell$ for all $\ell \in \mathsf{POS}$. Then,* $(\mathsf{LP}(\vec{t}))$ *is feasible.*

7.2.2 Rounding Column-Sparse LPs for Assignment Problems with Packing Constraints

Designing an LP-rounding algorithm is substantially more challenging now. In the Top_ℓ-norm case, we set up an auxiliary LP (Aux-Top_ℓ-LP) by extracting a single budget constraint for each machine that served to bound the contribution from the truncated jobs on that machine. This LP was quite easy to round (e.g., using Theorem 2.32) because each z_{ij} variable participated in exactly one constraint (thereby trivially yielding an $O(1)$ bound on the column sum for each variable). Since we now have to simultaneously control multiple Top_ℓ-norms, for each machine i, we will now need to include a budget constraint *for every index* $\ell \in \mathsf{POS}$ so as to bound the contribution from the truncated jobs for index ℓ (i.e., $\mathbf{E}[\big(\sum_{j:j\mapsto i}X_{ij}^{<t_\ell}\big)^{\geq\Omega(t_\ell)}]$). Additionally, unlike (Aux-Top_ℓ-LP), wherein the contribution from the exceptional jobs was bounded by incorporating it in the objective function, we will now need, for each $\ell \in \mathsf{POS}$, a separate constraint to bound the total contribution from the exceptional jobs for index ℓ.

Thus, while we can set up an auxiliary LP similar to (Aux-Top_ℓ-LP) containing these various budget constraints (see (6.6) and (6.7)), rounding a fractional solution to this

LP to obtain an assignment that approximately satisfies these various budget constraints presents a significant technical hurdle. The column-sums of our auxiliary LP are bounded by $D := O(\log m)$, so simply utilizing Theorem 2.32 leads to an $O(D)$-factor violation in the budget constraints. Overall, this would lead to an $O(\log m)$-approximation algorithm, which is much worse than the $O(\log\log m/\log\log\log m)$ guarantee claimed in Theorem 7.6. To obtain the improved approximation guarantee, we use a very powerful randomized-rounding result of Harris and Srinivasan for column-sparse assignment problems with packing constraints. We paraphrase a sub-case of Theorem 4.7 from [14] in a form that is suitable for us, and consistent with our notation.

Theorem 7.8. *Consider a feasibility LP for an assignment problem with finitely many packing constraints:*

(HS-LP)

$$\sum_{i\in[m],j\in J} a_{i,j,k} z_{ij} \le c_k \qquad \forall k \in [N] \tag{7.8}$$

$$\sum_{i\in[m]} z_{ij} = 1 \qquad \forall j \in J \tag{7.9}$$

$$z_{ij} \ge 0 \qquad \forall i \in [m], j \in J, \tag{7.10}$$

where $c_k \in [1,\infty)$ *for all* $k \in [N]$. *Let* $\overline{z}$ *denote a feasible fractional solution to this LP. Further assume that for each* $i \in [m], j \in J$ *with* $\overline{z}_{ij} > 0$, *we have* $a_{i,j,k} \in [0,1]$ *for all* $k \in [N]$ *and* $\sum_{k\in[N]} a_{i,j,k} \le D$ *for some parameter* $D \ge 2$. *In expected time polynomial in* m, $|J|$ *and* N, *we can find an integral assignment* $\hat{z} \in \{0,1\}^{[m]\times J}$ *satisfying:*

(i) *if* $\overline{z}_{ij} = 0$ *for some* $i \in [m], j \in J$, *then* $\hat{z}_{ij} = 0$ *holds.*

(ii) $\sum_{i\in[m],j\in J} a_{i,j,k}\hat{z}_{ij} \le O(\log D/\log\log D)\cdot c_k$ *for all* $k \in [N]$.

We make a few remarks on the above result before applying it to our setting. In [14], the bounds on $a_{i,j,k}$ and $\sum_{k\in[N]} a_{i,j,j}$ are stated for all $i \in [m], j \in J$ (irrespective of whether $\overline{z}_{ij} = 0$ holds or not). This is simply a matter of style since we can always drop the $a_{i,j,k}z_{ij}$ terms whenever $z_{ij} = 0$ holds. Next, the above rounding algorithm is based on a single application of the *Partial Resampling Algorithm* (PRA), a powerful generalization of the Moser-Tardos framework for designing constructive versions of the Lovász Local Lemma. At a high level, PRA uses the fractional assignment $\overline{z}$ to integrally assign jobs to machines, i.e., job j is assigned to machine i with probability $\overline{z}_{ij}$). Whenever a packing constraint $\sum_{i,j} a_{i,j,k}\hat{z}_{ij} \le O(1)$ is violated by an $\Omega(\log D/\log\log D)$ factor, a subset of jobs belonging to the violated constraint are carefully selected and reassigned using $\{\overline{z}_{ij}\}$ probabilities. As mentioned before, the actual algorithm and its analysis are quite complicated, and we will only be using it in a black-box fashion. Lastly, the expected running time of the algorithm in Theorem 4.7 from [14] is stated to be independent of N, but this requires that one has an efficient separation oracle for (HS-LP).

7.2.3 LP-Rounding Strategy

We now give full details about our procedure for rounding the stochastic load-balancing LP ($\mathsf{LP}(\vec{t})$). Consider a nonincreasing sequence $\vec{t} = (t_\ell)_{\ell \in \mathsf{POS}}$ such that ℓt_ℓ is nondecreasing over $\ell \in \mathsf{POS}$; since ℓt_ℓ is supposed to model $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})]$, this is a valid assumption. Suppose that ($\mathsf{LP}(\vec{t})$) is feasible, and let $(\{\bar{z}_{ij}\}_{i,j}, \{\bar{\xi}_{i,\ell}\}_{i,\ell})$ be a feasible fractional solution to this LP. We will show that $\bar{z}$ can be rounded to an integral assignment σ whose expected Top_ℓ norms are bounded by $O(\log\log m/\log\log\log m) \cdot \ell t_\ell$.

G1. For each $\ell \in \mathsf{POS}$, define the following quantities. Define $\mathcal{M}^\ell_{\mathsf{low}} := \{i \in [m] : \bar{\xi}_{i,\ell} < 1\}$, and let $\mathcal{M}^\ell_{\mathsf{hi}} := [m] \setminus \mathcal{M}^\ell_{\mathsf{low}}$. For every $i \in [m]$, set $\lambda_{i,\ell} := \min(\lceil 1/\bar{\xi}_{i,\ell}\rceil, 100m) \in \{2, \dots, 100m\}$ if $i \in \mathcal{M}^\ell_{\mathsf{low}}$, and $\lambda_{i,\ell} := 1$ otherwise.

G2. The auxiliary LP will enforce constraints (7.2), and constraint (7.3) or (7.4) for every machine $i \in [m]$ and index $\ell \in \mathsf{POS}$, depending on the parameter $\lambda_{i,\ell}$. The auxiliary LP is therefore as follows.

(Aux-Gen-LP)

$$\sum_{i\in[m],j\in J} \Big(\mathbf{E}[X_{ij}^{\geq t_\ell}]/(2\ell t_\ell)\Big)\eta_{ij} \leq 1 \qquad \forall \ell \in \mathsf{POS} \tag{7.11}$$

$$\sum_{j\in J} \beta_{\lambda_{i,\ell}}(X_{ij}^{<t_\ell}/4t_\ell)\eta_{ij} \leq 14 \qquad \forall\, \ell \in \mathsf{POS}, i \in \mathcal{M}^\ell_{\mathsf{low}} \tag{7.12}$$

$$\sum_{j\in J} \mathbf{E}[X_{ij}^{<t_\ell}/t_\ell]\eta_{ij} \leq \bar{\xi}_{i,\ell} + 1 \quad \forall\, \ell \in \mathsf{POS}, i \in \mathcal{M}^\ell_{\mathsf{hi}} \tag{7.13}$$

$$\eta \in \mathcal{Q}^{\mathsf{asgn}}. \tag{7.14}$$

We have scaled constraints (7.2) for the exceptional jobs to reflect the fact that we can afford to incur an $O(\ell t_\ell)$ violation in the constraint for index ℓ. Also, note that the RHS of this constraint was increased from $1/2$ to 1. This is done to ensure that the hypothesis of Theorem 7.8 is met.

G3. The fractional assignment $\bar{z}$ is a feasible solution to (Aux-Gen-LP). We apply Theorem 7.8 with the above system to round $\bar{z}$ and obtain an integral assignment σ.

7.2.4 Analysis

We first note that the fractional assignment $\overline{z}$ (to the assignment-LP (Aux-Gen-LP)) satisfies the hypothesis of Theorem 7.8 with $D = O(\log m)$. Thus, the theorem directly shows that σ satisfies constraints (7.11), (7.12), and (7.13) with an additive violation of at most $\log D/\log\log D$. A straightforward calculation readily yields $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] = O(\log D/\log\log D)\cdot \ell t_\ell$ for all $\ell \in \mathsf{POS}$.

Claim 7.9. *Let $D := 2\cdot|\mathsf{POS}| = O(\log m)$. For any $\ell \in \mathsf{POS}$, the assignment σ satisfies:*

(a) $\sum_{j\in J} \mathbf{E}[X^{\geq t_\ell}_{\sigma(j),j}] \leq O(\log D/\log\log D)\cdot \ell t_\ell$.

(b) $\sum_{j\in J:\sigma(j)=i} \beta_{\lambda_{i,\ell}}(X^{<t_\ell}_{ij}/4t_\ell) \leq O(\log D/\log\log D)$ *for any machine* $i \in \mathcal{M}^\ell_{\mathsf{low}}$.

(c) $\sum_{j\in J:\sigma(j)=i} \mathbf{E}[X^{<t_\ell}_{ij}/t_\ell] \leq O(\log D/\log\log D)\cdot(\overline{\xi}_{i,\ell}+1)$ *for any machine* $i \in \mathcal{M}^\ell_{\mathsf{hi}}$.

Proof. Fix some $i \in [m]$ and $j \in J$ with $\overline{z}_{ij} > 0$. Since $\overline{z}$ is feasible for $(\mathsf{LP}(\vec{t}))$, constraint (7.6) gives $\mathbf{E}[X^{\geq t_1}_{ij}] \leq t_1$, which further implies $\mathbf{E}[X^{\geq t_\ell}_{ij}] \leq t_1 + \mathbf{E}[X^{\geq t_1}_{ij}] \leq 2t_1$. By our assumption that ℓt_ℓ is non-decreasing over $\ell \in \mathsf{POS}$, we get $\mathbf{E}[X^{\geq t_\ell}_{ij}]/(2\ell t_\ell) \in [0,1]$. Next, we trivially have $\beta_{\lambda_{i,\ell}}(X^{<t_\ell}_{ij}/4t_\ell), \mathbf{E}[X^{<t_\ell}_{ij}/t_\ell] \in [0,1]$. For each $\ell \in \mathsf{POS}$, the variable $\overline{z}_{ij}$ participates in constraint (7.6) and exactly one of constraint (7.12) or (7.13). Therefore, $\overline{z}$ satisfies the hypothesis of Theorem 7.8 with $D = 2\cdot|\mathsf{POS}|$, and all three conclusions directly follow from this theorem. ∎

The following result is similar to Lemmas 6.7 and 6.8. Let $j : j \mapsto i$ be a shorthand for $\{j \in J : \sigma(j) = i\}$.

Lemma 7.10. *Fix some $\ell \in \mathsf{POS}$. Define random vectors $Y^{\mathsf{excep},\ell}, Y^{\mathsf{low},\ell}, Y^{\mathsf{hi},\ell}$ following product distributions on $\mathbb{R}^m_{\geq 0}$ as follows: for any machine $i \in [m]$, $Y^{\mathsf{excep},\ell}_i := \sum_{j:j\mapsto i} X^{\geq t_\ell}_{ij}$,*

$$Y^{\mathsf{low},\ell}_i := \begin{cases} \sum_{j:j\mapsto i} X^{<t_\ell}_{ij}; & \text{if } i \in \mathcal{M}^\ell_{\mathsf{low}} \\ 0; & \text{otherwise} \end{cases} \qquad \text{and } Y^{\mathsf{hi},\ell}_i := \begin{cases} \sum_{j:j\mapsto i} X^{<t_\ell}_{ij}; & \text{if } i \in \mathcal{M}^\ell_{\mathsf{hi}} \\ 0; & \text{otherwise.} \end{cases}$$

For each $Y \in \{Y^{\mathsf{excep},\ell}, Y^{\mathsf{hi},\ell}, Y^{\mathsf{low},\ell}\}$, we have $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq O(\log\log m/\log\log\log m)\cdot \ell t_\ell$.

Proof. Modulo $O(\log\log m/\log\log\log m)$ multiplicative factors, the proof is essentially the same as that of Lemmas 6.7 and 6.8. ∎

By definition, for any $\ell \in \mathsf{POS}$ we have $\overrightarrow{\mathsf{load}}^{\sigma} = Y^{\mathsf{excep},\ell} + Y^{\mathsf{low},\ell} + Y^{\mathsf{hi},\ell}$. By triangle inequality of norms, we get the following.

Theorem 7.11. *For any index $\ell \in \mathsf{POS}$, the assignment σ satisfies:*

$$\mathbf{E}[\mathsf{Top}_{\ell}(\overrightarrow{\mathsf{load}}^{\sigma})] \leq O\Big(\frac{\log\log m}{\log\log\log m}\Big) \cdot \ell t_{\ell}.$$

7.2.5 Our Algorithm

We now describe our entire algorithm for $\mathsf{StochNormLB}$. We begin with a high-level description. Consider a nonincreasing sequence $\vec{t^*} = (t^*_\ell)_{\ell \in \mathsf{POS}}$ such that $t^*_\ell = \Theta(\mathbf{E}[\mathsf{Top}_{\ell}(\overrightarrow{\mathsf{load}}^{\sigma^*})]/\ell)$ for all indices $\ell \in \mathsf{POS}$. Given Theorem 7.11 and Claim 7.7, it is clear that the assignment σ obtained from rounding a fractional assignment to $(\mathsf{LP}(\vec{t^*}))$ is an $O(\log\log m/\log\log\log m)$-approximation to the given instance of $\mathsf{StochNormLB}$. Since we do not have direct access to $\vec{t^*}$, we instead consider a poly(m)-size set $\mathcal{T} \subseteq \mathbb{R}^{\mathsf{POS}}_{\geq 0}$ that contains $\vec{t^*}$. While we cannot necessarily identify $\vec{t^*}$ in $\mathcal{T}$, we use Theorem 4.15(ii) to infer that if $(\mathsf{LP}(\vec{t}))$ is feasible, then we can use $\vec{t}$ to come up with a good estimate of the expected f-norm of the solution computed from rounding a solution for $(\mathsf{LP}(\vec{t}))$. This will imply that the "best" vector $\vec{t} \in \mathcal{T}$ yields the desired approximate solution. We formally state the algorithm below.

Proof of Theorem 7.6. For $\ell \in \mathsf{POS}$, let t^*_ℓ be the smallest number of the form 2^r, where $r \in \mathbb{Z}$ (and could be negative) such that $t^*_\ell \geq 2\,\mathbf{E}[\mathsf{Top}_{\ell}(\overrightarrow{\mathsf{load}}^{\sigma^*})]/\ell$. By Lemma 6.10, we have a bound on OPT and t^*_1:

$$\frac{\mathsf{UB}}{m} \leq \mathbf{E}[\mathsf{Top}_1(\overrightarrow{\mathsf{load}}^{\sigma^*})] \leq \mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma^*})] \leq \mathsf{UB},$$

where $\mathsf{UB} := \sum_{j \in J}\big(\min_{i \in [m]} \mathbf{E}[X_{ij}]\big)$ can be computed explicitly from input data. Observe that $\mathbf{E}[\mathsf{Top}_{\ell}(\overrightarrow{\mathsf{load}}^{\sigma^*})]/\ell$ does not increase as ℓ increases. It follows that

$$\frac{2 \cdot \mathsf{UB}}{m^2} \leq \frac{2 \cdot \mathbf{E}[\mathsf{Top}_{\ell}(\overrightarrow{\mathsf{load}}^{\sigma^*})]}{\ell} \leq 2 \cdot \mathsf{UB},$$

holds for all $\ell \in \mathsf{POS}$. Define

$$\mathcal{T} := \Big\{\vec{t} \in \mathbb{R}^{\mathsf{POS}}_{\geq 0} : \forall \ell \in \mathsf{POS}, \quad \frac{2 \cdot \mathsf{UB}}{m^2} \leq t_\ell < 4 \cdot \mathsf{UB}, \quad t_\ell \text{ is a power of 2, and } \frac{t_\ell}{t_{2\ell}} \in \{1, 2\} \text{ whenever } 2\ell \in \mathsf{POS}\Big\} \tag{7.15}$$

Claim 7.12. *The vector $\vec{t^*} = (t^*_\ell)_{\ell\in\mathsf{POS}}$ belongs to $\mathcal{T}$, and $|\mathcal{T}| = O(m\log m)$.*

Proof. Note that each vector $\vec{t} \in \mathcal{T}$ is completely defined by specifying t_1, and the set of "breakpoint" indices $\ell \in \mathsf{POS}$ for which $t_\ell/t_{2\ell} = 2$. There are $O(\log m)$ choices for t_1, and at most $2^{|\mathsf{POS}|} \le m$ choices for the set of breakpoint indices; hence, $|\mathcal{T}| = O(m\log m)$. To see that $\vec{t^*} \in \mathcal{T}$, by definition, each t^*_ℓ is a power of 2. The bounds shown above on $2\,\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})]/\ell$ show that t^*_ℓ lies within the stated bounds. The only nontrivial condition to check is $t^*_\ell/t^*_{2\ell} \in \{1,2\}$ whenever $2\ell \in \mathsf{POS}$. We have

$$\frac{(2\ell)\,t^*_{2\ell}}{2} \ge \mathbf{E}[\mathsf{Top}_{2\ell}(\overrightarrow{\mathsf{load}}^{\sigma^*})] \ge \mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})] > \frac{\ell t^*_\ell}{4},$$

which implies $t^*_\ell < 4\,t^*_{2\ell}$. Next, since $\mathsf{Top}_{2\ell}(\cdot) \le 2\mathsf{Top}_\ell(\cdot)$ holds, we have

$$\frac{\ell t^*_\ell}{2} \ge \mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})] \ge \frac{1}{2}\mathbf{E}[\mathsf{Top}_{2\ell}(\overrightarrow{\mathsf{load}}^{\sigma^*})] > \frac{(2\ell)t^*_{2\ell}}{8},$$

which implies $t^*_\ell > \frac{1}{2}\,t^*_{2\ell}$. As the t^*_ℓs are powers of 2, we get that $t^*_\ell/t^*_{2\ell} \in \{1,2\}$. ∎

We enumerate over all guess vectors $\vec{t} \in \mathcal{T}$ and check if (LP($\vec{t}$)) is feasible. For each $\vec{t} \in \mathcal{T}$ and $\ell \in \mathsf{POS}$, define $B_\ell(\vec{t}) := \ell t_\ell$. For ease of notation, we drop the argument $\vec{t}$ when it is clear from the context. Let $b : [0,m] \to \mathbb{R}_{\ge 0}$ denote the upper envelope curve (see Definition 2.16) for the sequence $\big(B_\ell(\vec{t})\big)_{\ell\in\mathsf{POS}}$. Define $\vec{b}(\vec{t}) \in \mathbb{R}^m_{\ge 0}$ as follows: for $i \in [m]$, $\vec{b}_i := b(i) - b(i-1)$. Among all feasible $\vec{t}$ vectors, let $\vec{t'}$ be a vector that minimizes $f\big(\vec{b}(\vec{t'}\,)\big)$.

Claim 7.13. *We have $f\big(\vec{b}(\vec{t'}\,)\big) \le 4\,\mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma^*})]$.*

Proof. By definition, $f\big(\vec{b}(\vec{t'}\,)\big) \le f\big(\vec{b}(\vec{t^*}\,)\big)$. Let $y := \mathbf{E}[(\overrightarrow{\mathsf{load}}^{\sigma^*})^{\downarrow}]$. Since $B_\ell(\vec{t^*}\,) = \ell t^*_\ell \le 4\,\mathsf{Top}_\ell(y)$ for all $\ell \in \mathsf{POS}$, Theorem 2.15 gives $f\big(\vec{b}(\vec{t^*}\,)\big) \le 4f(y) \le 4\,\mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma^*})]$, where the last inequality follows from convexity of norms. ∎

Now let σ be the assignment obtained by rounding a feasible solution to (LP($\vec{t'}$)). By Theorem 7.11 we have: $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})] \le O(\log\log m/\log\log\log m)\cdot B_\ell(\vec{t'}\,)$ for all $\ell \in \mathsf{POS}$. Therefore, by Theorem 4.15(ii), we get:

$$\mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma})] \le O\Big(\frac{\log\log m}{\log\log\log m}\Big)\cdot f(\vec{b}(\vec{t'}\,)) \le O\Big(\frac{\log\log m}{\log\log\log m}\Big)\cdot \mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma^*})].$$

∎

Remark 7.2. *When the machines are identical, the proof of Theorem 7.6 gives a simultaneous* $O(\log\log m/\log\log\log m)$*-approximation guarantee for all monotone symmetric norms. To see this, observe that after obtaining the right thresholds* $(t_\ell)_{\ell\in\mathsf{POS}}$ *(see the discussion in Section 7.1.3), the auxiliary LP only consists of assignment constraints (7.14) and budget constraints of the form (7.12) with* $\lambda_{i,\ell} = m/\ell$ *for all machines* $i \in [m]$ *and indices* $\ell \in \mathsf{POS}$*. Note that the uniform assignment* $\overline{z}_{ij} := 1/m$ *for all* $i \in [m], j \in J$ *is feasible to the auxiliary LP. Let* σ *denote the assignment obtained from rounding* $\overline{z}$ *using Theorem 7.8. Using Claim 7.9(b) and (a suitable modification of) Lemma 6.3 we can show that* $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] = O(\log\log m/\log\log\log m)\cdot \ell t_\ell$ *for all* $\ell \in \mathsf{POS}$*, which implies that* σ *is simultaneously an* $O(\log\log m/\log\log\log m)$*-approximate assignment for every monotone, symmetric norm.*

Remark 7.3. *The simultaneous every-monotone-symmetric-norm approximation guarantee does not extend to the unrelated-machines setting. This is because there is no way of guaranteeing feasibility of (Aux-Gen-LP) unless the* t_ℓ*'s come from a single assignment.*

Chapter 8

Stochastic Load Balancing with Weighted Bernoulli Jobs

In this chapter we consider stochastic min-norm load balancing with weighted Bernoulli jobs, which we abbreviate to $\mathsf{BerNormLB}$. In this setting, the processing time of a job j on machine i is distributed as a weighted Bernoulli variable of type (q_{ij}, s_{ij}) for some $q_{ij} \in [0,1]$ and $s_{ij} \in \mathbb{R}_{\geq 0}$. Recall that a weighted Bernoulli trial of type (q,s) takes size s with probability q, and 0 otherwise. Note that the job-size distributions can be fully described in the input by explicitly specifying $\{(q_{ij}, s_{ij})\}_{i\in[m],j\in J}$; our algorithms in Chapters 6 and 7 did not use any information about the correlations among $\{X_{ij}\}_{i\in[m]}$ for any job $j \in J$, and the same will be true for the algorithms in this chapter.

Our main result in this chapter is an $O(1)$-approximation algorithm for $\mathsf{BerNormLB}$ on unrelated machines, and a simultaneous $O(1)$-approximation when the machines are identical.

Theorem 8.1.
There is an $O(1)$-approximation algorithm for stochastic min-norm load balancing on unrelated machines when jobs are distributed as weighted Bernoulli trials.
Furthermore, when the machines are identical, the approximation guarantee holds simultaneously for all monotone symmetric norms.

Let σ^* denote an optimal assignment for the given instance of $\mathsf{BerNormLB}$, and $\mathsf{OPT}_f := \mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma^*})]$ denote the optimal objective value. In Chapter 7, we saw an LP-based $O(\log\log m/\log\log\log m)$-approximation algorithm for $\mathsf{StochNormLB}$ (see Theorem 7.6).

The main idea there was that the constraint matrix of the auxiliary LP (see (Aux-Gen-LP)) has column sums that are bounded by $D := O(\log m)$, so using the sophisticated rounding algorithm in Theorem 7.8 we obtained an $O(\log D/\log\log D)$-approximate assignment. Our approximation strategy for the Bernoulli case has an extra filtering step where we remove redundant constraints from the auxiliary LP. This leads to $O(1)$-bounded column sums in the auxiliary LP for the Bernoulli case, so the simpler iterative-rounding result in Theorem 2.32 is sufficient to obtain an $O(1)$-approximate assignment.

We recall some notation and definitions from prior chapters. The quantity $\mathsf{UB} = \sum_{j\in J}\big(\min_{i\in[m]} \mathbf{E}[X_{ij}]\big) = \sum_{j\in J}\big(\min_{i\in[m]} q_{ij}s_{ij}\big)$ serves as a useful bound on the Top_ℓ norms of the optimal assignment (see Lemma 6.10). For $\ell \in \mathsf{POS}$, t^*_ℓ denotes the smallest power of 2 (possibly with a negative exponent) satisfying $t^*_\ell \geq 2\,\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma^*})]/\ell$. The poly$(m)$-size collection $\mathcal{T}$ containing the correct guess-vector $\vec{t^*} = (t^*_\ell)_{\ell\in\mathsf{POS}}$ is given by:

$$\mathcal{T} = \Big\{\vec{t} \in \mathbb{R}^{\mathsf{POS}}_{\geq 0} : \forall \ell \in \mathsf{POS},\quad \frac{2\cdot\mathsf{UB}}{m^2} \leq t_\ell < 4\cdot\mathsf{UB},$$
$$t_\ell \text{ is a power of 2, and } \frac{t_\ell}{t_{2\ell}} \in \{1,2\} \text{ whenever } 2\ell \in \mathsf{POS}\Big\}.$$

8.1 Filtering the Auxiliary LP

Fix some $\vec{t} = (t_\ell)_{\ell\in\mathsf{POS}} \in \mathcal{T}$ such that $(\mathsf{LP}(\vec{t}))$ is feasible[1]. Let $(\overline{z}, \overline{\xi})$ be a feasible fractional solution to this LP. In this section, we identify redundancies in LP constraints arising from exceptional and truncated jobs w.r.t. thresholds $\{t_\ell\}_{\ell\in\mathsf{POS}}$.

8.1.1 Redundant Constraints Arising from Truncated Jobs

Recall that for a machine $i \in [m]$ and index $\ell \in \mathsf{POS}$, the LP-variable $\xi_{i,\ell}$ models $\mathbf{E}[(\sum_{j:j\mapsto i} X^{<t_\ell}_{ij}/t_\ell)^{\geq 1}]$. Constraints (7.3) and (7.4) capture the contribution of truncated jobs, w.r.t. threshold t_ℓ, that are assigned to machine i, and yield that:

$$\xi_{i,\ell} \geq \max\left\{\sum_{j\in J}\mathbf{E}[X^{<t_\ell}_{ij}/t_\ell]z_{ij} - 1,\ \max_{\lambda\in\{2,3,\ldots,100m\}}\Big(\frac{\sum_{j\in J}\beta_\lambda(X^{<t_\ell}_{ij}/4t_\ell)z_{ij} - 6}{4\lambda}\Big)\right\}.$$

[1] In Section 7.2.3, we only assumed that t_ℓ is non-increasing over $\ell \in \mathsf{POS}$ and ℓt_ℓ is non-decreasing over $\ell \in \mathsf{POS}$. The assumption that t_ℓ's are powers of 2 and $t_\ell/t_{2\ell} \in \{1,2\}$ whenever $\ell, 2\ell \in \mathsf{POS}$ will be useful to us in this chapter; the bounds on t_ℓ in terms of UB are not needed for the rounding step.

Constraint (7.5), which reads as $\sum_{i\in[m]} \xi_{i,\ell} \leq \ell$, imposes that the expected Top_ℓ norm of the "t_ℓ-truncated" load vector is $O(\ell t_\ell)$.

Consider some $\ell \in \mathsf{POS}$ such that $2\ell \in \mathsf{POS}$ and $t_\ell = t_{2\ell}$. Observe that the constraints arising from truncated jobs for the index 2ℓ are implied by the corresponding constraints for the index ℓ: indeed, if the $\xi_{i,\ell}$'s satisfy the constraints for index ℓ, then setting $\xi_{i,2\ell} = \xi_{i,\ell}$ for all machines $i \in [m]$ satisfies all the constraints for index 2ℓ. This motivates the following definition:

$$\mathsf{POS}^{<} := \{\ell \in \mathsf{POS} : \ell = 1 \text{ or } t_{\ell/2} = 2t_\ell\}. \tag{8.1}$$

By the above discussion, it suffices to only consider truncated-jobs constraints for indices $\ell \in \mathsf{POS}^{<}$.

8.1.2 Redundant Constraints Arising from Exceptional Jobs

For any index $\ell \in \mathsf{POS}$, constraint (7.2), which reads as $\sum_{i\in[m],j\in J} \mathbf{E}[X_{ij}^{\geq t_\ell}]z_{ij} \leq \ell t_\ell$, imposes that the expected Top_ℓ norm of the "t_ℓ-exceptional" load vector is $O(\ell t_\ell)$. Again, observe that the constraint for index $\ell \in \mathsf{POS} \setminus \mathsf{POS}^{<}$ is implied by the constraint for index $\ell/2$, so it suffices to only focus on indices in $\mathsf{POS}^{<}$.

Now consider an index $\ell \in \mathsf{POS}^{<}$ such that $2\ell \in \mathsf{POS}^{<}$, i.e., $2\ell \in \mathsf{POS}$ and $t_\ell = 2t_{2\ell}$ holds. Since $(2\ell)t_{2\ell} = \ell t_\ell$, constraint (7.2) for the index 2ℓ implies the corresponding constraint for the index ℓ. This motivates the following definition:

$$\mathsf{POS}^{=} := \{\ell \in \mathsf{POS} : 2\ell \notin \mathsf{POS} \text{ or } t_\ell = t_{2\ell}\}. \tag{8.2}$$

By the above discussion, it suffices to only consider exceptional-jobs constraints for indices $\ell \in \mathsf{POS}^{<} \cap \mathsf{POS}^{=}$.

8.1.3 The Filtered Auxiliary LP

By the discussion in Section 8.1.1, we may assume that $\xi_{i,\ell} = \xi_{i,\ell/2}$ for any machine $i \in [m]$ and any index $\ell \in \mathsf{POS} \setminus \mathsf{POS}^{<}$. We recall some definitions from Chapter 7 that were used to describe the auxiliary LP. For each $\ell \in \mathsf{POS}$ we defined: $\mathcal{M}^\ell_{\mathsf{low}} := \{i \in [m] : \overline{\xi}_{i,\ell} < 1\}$ and $\mathcal{M}^\ell_{\mathsf{hi}} := [m] \setminus \mathcal{M}^\ell_{\mathsf{low}}$. For each machine $i \in [m]$ and index $\ell \in \mathsf{POS}$, we set $\lambda_{i,\ell} := \min(\lceil 1/\overline{\xi}_{i,\ell}\rceil, 100m) \in \{2, \ldots, 100m\}$ if $i \in \mathcal{M}^\ell_{\mathsf{low}}$, and $\lambda_{i,\ell} := 1$ otherwise. We use the same definitions here, but we write down the auxiliary LP after filtering out all redundant constraints due to indices not in $\mathsf{POS}^{<}$ or $\mathsf{POS}^{=}$. In the following we state the auxiliary LP that remains after filtering out all redundant constraints. Like before, we scale constraints (7.2) for the exceptional jobs to reflect the fact that we can afford to incur an additive $O(\ell t_\ell)$ violation in the constraint for index ℓ.

$$\text{(Aux-Ber-LP)} \qquad \begin{aligned} \sum_{i\in[m],j\in J} \big(\mathbf{E}[X_{ij}^{\geq t_\ell}]/(\ell t_\ell)\big)\eta_{ij} &\leq 1 && \forall \ell \in \mathsf{POS}^< \cap \mathsf{POS}^= && (8.3)\\ \sum_{j\in J} \beta_{\lambda_{i,\ell}}(X_{ij}^{<t_\ell}/4t_\ell)\eta_{ij} &\leq 14 && \forall\, \ell \in \mathsf{POS}^<, i \in \mathcal{M}^\ell_{\mathsf{low}} && (8.4)\\ \sum_{j\in J} \mathbf{E}[X_{ij}^{<t_\ell}/t_\ell]\eta_{ij} &\leq \overline{\xi}_{i,\ell}+1 && \forall\, \ell \in \mathsf{POS}^<, i \in \mathcal{M}^\ell_{\mathsf{hi}} && (8.5)\\ \eta &\in \mathcal{Q}^{\mathsf{asgn}}. && && (8.6)\end{aligned}$$

Since $\overline{z}$ is feasible to the above auxiliary LP, to obtain an $O(1)$-approximate assignment it suffices to show that the column sums in the above LP are bounded by $O(1)$ for elements in the support of $\overline{z}$. We start with the constraints arising from truncated jobs where we use the fact that the job-size variables are weighted Bernoulli trials.

Claim 8.2. *Fix some machine $i \in [m]$ and a job $j \in J$. Let a^ℓ_{ij} denote the coefficient of η_{ij} variable in constraints (8.4) and (8.5). We have $\sum_{\ell\in\mathsf{POS}^<} a^\ell_{ij} \leq 2$.*

Proof. For any $[0,\theta]$-bounded random variable Z and a parameter $\lambda \in \mathbb{R}_{\geq 1}$, we have $\beta_\lambda(Z) \leq \theta$ (see Definition 5.6). Let $q := q_{ij}$ and $s := s_{ij}$. Since X_{ij} takes value s with probability q, and value 0 with probability $1-q$, for any $\ell \in \mathsf{POS}^<$, we have $a^\ell_{ij} \leq (s/t_\ell)\cdot \mathbb{1}_{s<t_\ell}$. [2] Here $\mathbb{1}_{\mathcal{E}} \in \{0,1\}$ is 1 if and only if the event $\mathcal{E}$ happens. By definition of $\mathsf{POS}^<$, the t_ℓ's decrease geometrically over $\ell \in \mathsf{POS}^<$, so

$$\sum_{\ell\in\mathsf{POS}^<} a^\ell_{ij} \leq \sum_{\ell\in\mathsf{POS}^<} \frac{s}{t_\ell}\cdot \mathbb{1}_{\left(\frac{s}{t_\ell}<1\right)} \leq 1+\frac{1}{2}+\frac{1}{4}+\cdots = 2. \qquad \blacksquare$$

Next, we prove a similar result for constraints arising from exceptional jobs. We remark that this bound holds for any distribution.

Claim 8.3. *Fix some machine $i \in [m]$ and job $j \in J$ such that $\overline{z}_{ij} > 0$. We have $\sum_{\ell\in\mathsf{POS}^<\cap\mathsf{POS}^=} \big(\mathbf{E}[X_{ij}^{\geq t_\ell}]/\ell t_\ell\big) \leq 4$.*

[2] If $i \in \mathcal{M}^\ell_{\mathsf{low}}$, then a^ℓ_{ij} is in fact at most $s/(4t_\ell)$, but otherwise, we can only say that $a^\ell_{ij} \leq s/t_\ell$.

Proof. We repeat the argument from Claim 7.9 where we showed that $\mathbf{E}[X_{ij}^{\geq t_\ell}] \leq 2t_1$ holds for any index $\ell \in \mathsf{POS}$. Due to (7.6), we have $\mathbf{E}[X_{ij}^{\geq t_1}] \leq t_1$, so $\mathbf{E}[X_{ij}^{\geq t_\ell}] \leq \mathbf{E}[X_{ij}^{\geq t_1}] + \mathbf{E}[X_{ij}^{< t_1}] \leq 2t_1$ follows.

Observe that for any index $\ell \in \mathsf{POS}^< \cap \mathsf{POS}^=$, the index 2ℓ is not in $\mathsf{POS}^<$. Also, t_ℓ drops by a factor of exactly 2 as ℓ steps over the indices in $\mathsf{POS}^<$. So, ℓt_ℓ increases by a factor of at least 2 as ℓ steps over the indices in $\mathsf{POS}^< \cap \mathsf{POS}^=$. Therefore,

$$\sum_{\ell \in \mathsf{POS}^< \cap \mathsf{POS}^=} \frac{\mathbf{E}[X_{ij}^{\geq t_\ell}]}{\ell t_\ell} \leq 2t_1 \left(\frac{1}{t_1} + \frac{1}{2t_1} + \frac{1}{4t_1} + \dots \right) \leq 4. \qquad \blacksquare$$

8.2 LP-Rounding Strategy

The fractional assignment $\overline{z}$ is a feasible solution to the filtered auxiliary LP (Aux-Ber-LP). We apply Theorem 2.32 for this linear system to round $\overline{z}$ and obtain an integral assignment $\sigma : J \to [m]$.

Claim 8.4. *The assignment σ satisfies:*

(a) $\sum_{j \in J} \mathbf{E}[X_{\sigma(j),j}^{\geq t_\ell}] \leq 7\ell t_\ell$ *for any index* $\ell \in \mathsf{POS}^< \cap \mathsf{POS}^=$.

(b) $\sum_{j \in J : \sigma(j) = i} \beta_{\lambda_{i,\ell}}(X_{ij}^{< t_\ell}/4t_\ell) \leq 20$ *for any index* $\ell \in \mathsf{POS}^<$ *and machine* $i \in \mathcal{M}_{\mathsf{low}}^\ell$.

(c) $\sum_{j \in J : \sigma(j) = i} \mathbf{E}[X_{ij}^{< t_\ell}/t_\ell] \leq \overline{\xi}_{i,\ell} + 7$ *for any index* $\ell \in \mathsf{POS}^<$ *and machine* $i \in \mathcal{M}_{\mathsf{hi}}^\ell$.

Proof. By Claims 8.2 and 8.3, the hypothesis of Theorem 2.32 holds for the violation parameter $\nu = 6$. All claims follow directly from the theorem. $\blacksquare$

8.2.1 Analysis

The analysis follows the template from Section 6.2.3 for the Top_ℓ case, but we need to also account for the redundant constraints that were dropped. For any index $\ell \in \mathsf{POS}$, we show that $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] = O(\ell t_\ell)$ by separately bounding the expected Top_ℓ-norm of the load vector induced by exceptional jobs, truncated jobs in $\mathcal{M}_{\mathsf{low}}^\ell$, and truncated jobs in $\mathcal{M}_{\mathsf{hi}}^\ell$. As usual, let $j : j \mapsto i$ be a shorthand for $\{j \in J : \sigma(j) = i\}$. We have

$$\overrightarrow{\mathsf{load}}^\sigma = Y^{\mathsf{excep},\ell} + Y^{\mathsf{low},\ell} + Y^{\mathsf{hi},\ell},$$

where for any machine $i \in [m]$, $Y_i^{\mathsf{excep},\ell} := \sum_{j:j\mapsto i} X_{ij}^{\geq t_\ell}$,

$$Y_i^{\mathsf{low},\ell} := \begin{cases} \sum_{j:j\mapsto i} X_{ij}^{<t_\ell}; & \text{if } i \in \mathcal{M}_{\mathsf{low}}^\ell \\ 0; & \text{otherwise} \end{cases} \quad \text{and} \quad Y_i^{\mathsf{hi},\ell} := \begin{cases} \sum_{j:j\mapsto i} X_{ij}^{<t_\ell}; & \text{if } i \in \mathcal{M}_{\mathsf{hi}}^\ell \\ 0; & \text{otherwise.} \end{cases}$$

We start with the easier bounds on $Y^{\mathsf{hi},\ell}$ and $Y^{\mathsf{excep},\ell}$.

Lemma 8.5. *We have* $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{hi},\ell})] \leq 8\ell t_\ell$ *for any index* $\ell \in \mathsf{POS}$.

Proof. We first recall (8.1): $\mathsf{POS}^< = \{\ell \in \mathsf{POS} : \ell = 1 \text{ or } t_{\ell/2} = 2t_\ell\}$. We prove the lemma by bounding the expected Top_m norm of $Y^{\mathsf{hi},\ell}$. Note that we only need to show this for $\ell \in \mathsf{POS}^<$: if $\ell \in \mathsf{POS} \setminus \mathsf{POS}^<$, then if we consider the largest index $\ell' \in \mathsf{POS}^<$ that is at most ℓ, we have $t_{\ell'} = t_\ell$, and so $Y^{\mathsf{hi},\ell'} = Y^{\mathsf{hi},\ell}$; thus, the bound on $\mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{hi},\ell})]$ follows from that on $\mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{hi},\ell'})]$. So suppose $\ell \in \mathsf{POS}^<$. By Claim 8.4(c), for any machine $i \in \mathcal{M}_{\mathsf{hi}}^\ell$ we have:

$$\mathbf{E}[Y_i^{\mathsf{hi},\ell}] = \sum_{j:j\mapsto i} \mathbf{E}[X_{ij}^{<t_\ell}] \leq t_\ell(\overline{\xi}_{i,\ell} + 7)$$

Since $|\mathcal{M}_{\mathsf{hi}}^\ell| \leq \ell$ (because $\overline{\xi}_{i,\ell} \geq 1$ for any $i \in \mathcal{M}_{\mathsf{hi}}^\ell$) and $\sum_{i\in[m]} \overline{\xi}_{i,\ell} \leq \ell$ by the LP constraint (7.5), we get:

$$\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{hi}})] \leq \mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{hi}})] \leq t_\ell \cdot \sum_{i\in\mathcal{M}_{\mathsf{hi}}} (\overline{\xi}_i + 7) \leq 8\ell t_\ell. \qquad \blacksquare$$

Lemma 8.6. *We have* $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{excep},\ell})] \leq 7\ell t_\ell$ *for any index* $\ell \in \mathsf{POS}$.

Proof. We prove the lemma by bounding the expected Top_m norm of $Y^{\mathsf{excep},\ell}$. Recall that $\mathsf{POS}^= = \{\ell \in \mathsf{POS} : 2\ell \notin \mathsf{POS} \text{ or } t_\ell = t_{2\ell}\}$. First, suppose that $\ell \in \mathsf{POS}^< \cap \mathsf{POS}^=$. By Claim 8.4(a), we have $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{excep},\ell})] \leq \mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{excep},\ell})] = \sum_{j\in J} \mathbf{E}[X_{\sigma(j),j}^{\geq t_\ell}] \leq 7\ell t_\ell$.

Next, suppose that $\ell \in \mathsf{POS} \setminus \mathsf{POS}^<$, so $t_\ell = t_{\ell/2}$. Let ℓ' be the largest index in $\mathsf{POS}^<$ that is at most ℓ. Then, $\ell' \leq \ell/2$ and $t_\ell = t_{\ell'}$. So $Y^{\mathsf{excep},\ell} = Y^{\mathsf{excep},\ell'}$. Also, $\ell' \in \mathsf{POS}^=$, since $t_\ell \leq t_{2\ell'} \leq t_{\ell'} = t_\ell$. Therefore, $\mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{excep},\ell})] = \mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{excep},\ell'})] \leq 7\ell' t_{\ell'} \leq 7\ell t_\ell$.

Finally, suppose $\ell \in \mathsf{POS}\setminus\mathsf{POS}^=$, so $t_\ell = 2t_{2\ell}$. Now, let ℓ' be the smallest index in $\mathsf{POS}^=$ that is at least ℓ. Then, $\ell' \geq 2\ell$, and $t_{\ell'/2} > t_{\ell'}$ (since $\ell'/2 \notin \mathsf{POS}^=$), so $\ell' \in \mathsf{POS}^<$. We claim that $\ell' t_{\ell'} = \ell t_\ell$. This is because for every $\ell'' \in \mathsf{POS}$, $\ell \leq \ell'' < \ell'$, we have $t_{\ell''} = 2t_{2\ell''}$, and so $\ell'' t_{\ell''} = 2\ell'' t_{2\ell''}$. Hence, we have $\mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{excep},\ell})] \leq \mathbf{E}[\mathsf{Top}_m(Y^{\mathsf{excep},\ell'})] \leq 7\ell' t_{\ell'} = 7\ell t_\ell$. $\blacksquare$

Lastly, we prove the bound on $Y^{\mathsf{low},\ell}$.

Lemma 8.7. *We have* $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{low},\ell})] \le 180\,\ell t_\ell$ *for any index* $\ell \in \mathsf{POS}$.

Proof. As with Lemma 8.5, we only need to consider $\ell \in \mathsf{POS}^<$. So fix $\ell \in \mathsf{POS}^<$. Let $W := Y^{\mathsf{low},\ell}/4t_\ell$. Observe that it suffices to show that

$$\sum_{i\in[m]} \mathbf{E}[W_i^{\ge 21}] = \sum_{i\in\mathcal{M}^\ell_{\mathsf{low}}} \mathbf{E}[W_i^{\ge 21}] \le 24\,\ell$$

holds. Then by Lemma 3.4 we immediately get $\mathbf{E}[\mathsf{Top}_\ell(W)] \le 45\,\ell$, or equivalently, $\mathbf{E}[\mathsf{Top}_\ell(Y^{\mathsf{low},\ell})] \le 180\,\ell t_\ell$. To this end, Claim 8.4(b) gives $\beta_{\lambda_{i,\ell}}(W_i) \le 20$ for any machine $i \in \mathcal{M}^\ell_{\mathsf{low}}$, where $\lambda_{i,\ell} = \min(100m, \lceil 1/\bar{\xi}_{i,\ell}\rceil) \ge 2$. Using Lemma 5.7 we get:

$$\mathbf{E}\left[W_i^{\ge 21}\right] \le 23/\lambda_{i,\ell} = 23\,\max\left(\frac{1}{100m}, \frac{1}{\lceil 1/\bar{\xi}_{i,\ell}\rceil}\right) \le 23\,\bar{\xi}_{i,\ell} + \frac{23}{100m}.$$

Summing over all machines in $\mathcal{M}^\ell_{\mathsf{low}}$ gives $\sum_{i\in\mathcal{M}^\ell_{\mathsf{low}}} \mathbf{E}[W_i^{\ge 21}] \le 24\ell$. ∎

Combining the above three lemmas gives the following result.

Theorem 8.8. *The assignment* σ *satisfies* $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^\sigma)] \le 195\,\ell t_\ell$ *for any index* $\ell \in \mathsf{POS}$.

8.3 Our Algorithm

Our $O(1)$-approximation algorithm for BerNormLB is essentially the same as the algorithm in Section 7.2.5 for the general StochNormLB problem. Recall $\mathsf{UB} := \sum_{j\in J}\left(\min_{i\in[m]} \mathbf{E}[X_{ij}]\right)$ and

$$\mathcal{T} := \Big\{\vec{t} \in \mathbb{R}^{\mathsf{POS}}_{\ge 0} : \forall \ell \in \mathsf{POS},\quad \frac{2\cdot\mathsf{UB}}{m^2} \le t_\ell < 4\cdot\mathsf{UB}, \quad t_\ell \text{ is a power of 2, and } \frac{t_\ell}{t_{2\ell}} \in \{1,2\} \text{ whenever } 2\ell \in \mathsf{POS}\Big\}.$$

Proof of Theorem 8.1. We enumerate over all guess vectors $\vec{t} \in \mathcal{T}$ and check if (LP($\vec{t}$)) is feasible. For each $\vec{t} \in \mathcal{T}$ and $\ell \in \mathsf{POS}$, define $B_\ell(\vec{t}) := \ell t_\ell$. For ease of notation, we drop the argument $\vec{t}$ when it is clear from the context. Let $b : [0, m] \to \mathbb{R}_{\ge 0}$ denote the upper envelope curve (see Definition 2.16) for the sequence $\left(B_\ell(\vec{t})\right)_{\ell\in\mathsf{POS}}$. Define $\vec{b}(\vec{t}) \in \mathbb{R}^m_{\ge 0}$ as follows: for $i \in [m]$, $\vec{b}_i := b(i) - b(i-1)$. Among all feasible $\vec{t}$ vectors, let $\vec{t'}$ be a vector that minimizes $f\big(\vec{b}(\vec{t'})\big)$. From Claim 7.13, we already have $f\big(\vec{b}(\vec{t'})\big) \le 4\,\mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma^*})]$.

Consider a feasible fractional assignment $\overline{z}$ to $(\mathsf{LP}(\vec{t'}))$. As shown in Section 8.2, we can round $\overline{z}$ to an integral assignment σ that satisfies $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})] \le O(B_\ell(\vec{t'}))$ for all $\ell \in \mathsf{POS}$ (Theorem 8.8). Therefore, by Theorem 4.15(ii), we get:

$$\mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma})] \le O(1) \cdot f(\vec{b}(\vec{t'})) \le O(1) \cdot \mathbf{E}[f(\overrightarrow{\mathsf{load}}^{\sigma^*})].$$

Now suppose that the machines are identical. We describe a simultaneous $O(1)$-approximation algorithm for all monotone symmetric norms. The binary search from Section 6.1.3 can essentially be repeated for each $\ell \in \mathsf{POS}$ to obtain a non-increasing vector $\vec{t} = (t_\ell)_{\ell \in \mathsf{POS}}$ satisfying: (i) each t_ℓ is a power of 2; (ii) for all $\ell \in \mathsf{POS}$, $t_\ell = O(\mathsf{OPT}_\ell/\ell)$, where OPT_ℓ denotes the optimal objective value when f is the Top_ℓ norm; (iii) $\sum_{j\in J} \mathbf{E}[X_j^{\ge t_\ell}] \le \ell t_\ell$ for all indices $\ell \in \mathsf{POS}$; and (iv) $\sum_{j \in J} \beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell) \le 10\,m$ for all indices $\ell \in \mathsf{POS}$. As we have seen before in Section 7.1, for any assignment σ and any index $\ell \in \mathsf{POS}$, the contribution of exceptional jobs (w.r.t. threshold t_ℓ) to the objective function is $O(\ell t_\ell)$, so it suffices to focus only on truncated jobs.

Consider a feasibility assignment LP which has budget constraints of the form:

$$\sum_{j\in J} \beta_{m/\ell}(X_j^{<t_\ell}/4t_\ell)\,\eta_{ij} \le 10$$

for all machines $i \in [m]$ and indices $\ell \in \mathsf{POS}^<$; note that we filtered redundant constraints arising from indices not in $\mathsf{POS}^<$. Observe that the uniform assignment $\overline{z}_{ij} := 1/m$ for all $i \in [m], j \in J$ is feasible to this LP, and in Claim 8.2 we established that the column sums are bounded by $O(1)$. Using Theorem 2.32 we can round $\overline{z}$ to an integral assignment σ satisfying $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})] = O(\ell t_\ell) = O(\mathsf{OPT}_\ell)$ for all $\ell \in \mathsf{POS}^<$. By repeating the proof strategy in Lemma 8.7, it is easy to see that the same bound on $\mathbf{E}[\mathsf{Top}_\ell(\overrightarrow{\mathsf{load}}^{\sigma})]$ holds for all $\ell \in \mathsf{POS}$. Since the upper bound is w.r.t. OPT_ℓ, Theorem 4.15(i) implies that σ is simultaneously an $O(1)$-approximate assignment for every monotone, symmetric norm. ■

Chapter 9

Stochastic Load Balancing with Poisson Jobs

In this chapter we consider stochastic minimum-norm load balancing with Poisson jobs, which we denote $\mathsf{PoisNormLB}$. In this setting, the processing time of a job $j \in J$ on machine $i \in [m]$ is distributed as a Poisson variable with mean λ_{ij}. Recall that the probability distribution of a Poisson variable $\mathsf{Pois}(\lambda)$ with mean $\lambda \in \mathbb{R}_{\geq 0}$ is given by:

$$\mathbf{Pr}\big[\mathsf{Pois}(\lambda) = k\big] = \frac{e^{-\lambda}\lambda^k}{k!} \text{ for any nonnegative integer } k \in \mathbb{Z}_{\geq 0}.$$

Our main result in this chapter is a novel, clean, and versatile *black-box reduction* from $\mathsf{PoisNormLB}$ to the deterministic problem, $\mathsf{MinNormLB}$, showing that, for any machine environment (i.e., unrelated/identical machines), guarantees obtained for $\mathsf{MinNormLB}$ translate to yield essentially the same guarantees for $\mathsf{PoisNormLB}$.

Theorem 9.1.
Let $\mathcal{I}^{\mathsf{Pois}} = (J, [m], \{\lambda_{ij}\}, f)$ *be an instance of* $\mathsf{PoisNormLB}$*, and let* $\mathcal{A}^{\mathsf{Det}}$ *be a* ρ*-approximation algorithm for* $\mathsf{MinNormLB}$*-instances with job-set* J*, machine-set* $[m]$*, and* $\{\lambda_{ij}\}_{i\in[m],j\in J}$ *job sizes. For any* $\varepsilon, \eta > 0$*, we can utilize* $\mathcal{A}^{\mathsf{Det}}$ *to obtain a* $\rho\big(1 + O(\varepsilon)\big)$*-approximate solution to* $\mathsf{PoisNormLB}$ *with probability at least* $1 - \eta$*, in time* $\mathrm{poly}\big(m^{1/\varepsilon}, n, \frac{1}{\eta}\big)$*. The run time also bounds the number of calls to* $\mathcal{A}^{\mathsf{Det}}$ *and the sample size.*

We emphasize that: (a) the above reduction preserves the machine environment: for instance, if we have identical machines ($\lambda_{ij} = \lambda_j$ for all i, j), we only need $\mathcal{A}^{\mathsf{Det}}$ to work for identical machines; and (b) algorithm $\mathcal{A}^{\mathsf{Det}}$ is required to work for an arbitrary monotone, symmetric norm (and not just the norm f): this generality is *crucial* for the above reduction and brings to the fore a prime benefit of working at the level of generality of monotone, symmetric norms.

We also remark that this reduction is not in fact limited to load-balancing but applies to any stochastic min-norm combinatorial optimization problem, and shows that this can be reduced to a deterministic min-norm version of the *same* combinatorial-optimization problem.

In [19], Ibrahimpur and Swamy gave a $(2+\varepsilon)$-approximation algorithm for MinNormLB on unrelated machines, and a PTAS when the machines are identical. We immediately get the following approximation guarantees for PoisNormLB by combining Theorem 9.1 and the results of [19]. We do not explicitly indicate the failure probability η below; the sample size, for a fixed ε, is $\text{poly}(m)/\eta$.

Theorem 9.2.
For any constant $\varepsilon > 0$, there is a randomized $\big(2+O(\varepsilon)\big)$-approximation algorithm for PoisNormLB *on unrelated machines.*

Theorem 9.3.
There is a randomized PTAS for PoisNormLB *on identical machines.*

Outline of our reduction. Before delving into the technical details, we discuss the chief ideas behind the reduction in Theorem 9.1. Since the sum of independent Poisson random variables is another Poisson random variable (Fact 2.2), the objective value of an assignment $\sigma : J \to [m]$ depends only the aggregate λ-vector $\Lambda^\sigma = (\Lambda_i^\sigma)_{i\in[m]}$, where $\Lambda_i^\sigma := \sum_{j:\sigma(j)=i} \lambda_{ij}$ for all $i \in [m]$. We drop σ in Λ^σ if the assignment σ is clear from the context. Overloading notation, for a vector $y \in \mathbb{R}^m_{\geq 0}$, we use $\mathsf{Pois}(y)$ to denote the random vector $\big(\mathsf{Pois}(y_1), \ldots, \mathsf{Pois}(y_m)\big)$ of *independent* Poisson random variables. Note that $\mathsf{Pois}(y)$ follows a product distribution on $\mathbb{R}^m_{\geq 0}$. We define the function $g : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ as follows: for any $y \in \mathbb{R}^m_{\geq 0}$,

$$g(y) := \mathbf{E}[f(\mathsf{Pois}(y))] = \mathbf{E}[f(\mathsf{Pois}(y_1), \ldots, \mathsf{Pois}(y_m))]. \tag{9.1}$$

With the above definitions, the goal in PoisNormLB is to find an assignment σ that minimizes $g(\Lambda^\sigma)$. The function g is *not* convex, but it satisfies a majorization inequality (Theorem 9.6): if $\mathsf{Top}_\ell(y) \leq \mathsf{Top}_\ell(y')$ for all $\ell \in [m]$, then $g(y) \leq g(y')$; this inequality is closely related to a property called *Schur convexity* that is satisfied by all symmetric convex functions. Theorem 9.6 provides a means for controlling $g(y)$ by bounding the Top_ℓ-norms of y, and is key to our approach; we give a self-contained proof of this result in Section 9.1.2.

Without loss of generality, we may assume that f is normalized so that $f(1, 0, \ldots, 0) = 1$ holds. Let σ^* be an optimal solution to the PoisNormLB-instance $\mathcal{I}^{\mathsf{Pois}}$, and let $\Lambda^* := \Lambda^{\sigma^*}$. The idea underlying our reduction is strikingly simple. Given Theorem 9.6, we aim to (ideally) find an assignment σ such that $\mathsf{Top}_\ell(\Lambda^\sigma) \leq \mathsf{Top}_\ell(\Lambda^*)$ for all $\ell \in [m]$. One of our chief insights is that *this amounts to solving a deterministic min-norm load balancing problem* with job sizes $\{\lambda_{ij}\}_{i,j}$ and the monotone, symmetric norm $h : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ given by $h(v) := \max_{\ell \in [m]} \mathsf{Top}_\ell(v)/\mathsf{Top}_\ell(\Lambda^*)$. Now σ^* yields a solution to this MinNormLB-instance of cost 1, and so solving this MinNormLB problem optimally, and utilizing Theorem 9.6, would yield the desired solution.

Additional ingredients are needed to make this idea work. We do not know the $\mathsf{Top}_\ell(\Lambda^*)$ values, and cannot "guess" these values for all $\ell \in [m]$; moreover, we cannot solve the MinNormLB problem optimally. We use the sparsification tools from Chapter 2, and, with a small loss in approximation, move to the sparse set $\mathsf{POS} = \mathsf{POS}_{m,\delta}$ (for, say, $\delta = \min\{0.5, \varepsilon\}$) and work with estimates B_ℓ of $\mathsf{Top}_\ell(\Lambda^*)$ for all $\ell \in \mathsf{POS}$; such that the norm in the MinNormLB instance is now $h(v) := \max_{\ell \in \mathsf{POS}} \mathsf{Top}_\ell(v)/B_\ell$. Now, using the algorithm $\mathcal{A}^{\mathsf{Det}}$ with the correct estimate-vector $B^* \in \mathbb{R}^{\mathsf{POS}}_{\geq 0}$ (where each B^*_ℓ overestimates $\mathsf{Top}_\ell(\Lambda^*)$ within a $(1 + \delta)$-factor), we obtain an assignment σ such that $\mathsf{Top}_\ell(\Lambda^\sigma) \leq \alpha' \mathsf{Top}_\ell(\Lambda^*)$ for all $\ell \in [m]$, where $\alpha' = \alpha\big(1 + O(\varepsilon)\big)$. Theorem 9.6 then shows that $g(\Lambda^\sigma) \leq g(\alpha' \Lambda^*)$, but we need a bound in terms of $g(\Lambda^*)$. To this end, we prove the important property that g is *subhomogeneous* (Lemma 9.7): $g(\theta y) \leq \theta \cdot g(y)$ for any scalar $\theta \geq 1$. Finally, similar to the issue that we encountered in Chapters 7 and 8, we cannot quite identify the correct B^*, but we can isolate it in a polynomial-sized set. We show how to estimate $g(y)$ using polynomially many samples (Lemma 9.8), and use this estimator to find (loosely speaking) the best solution among those computed for each candidate estimate-vector in this set. Combining these various ingredients yields Theorem 9.1.

9.1 Expected Norm of a Poisson Product Distribution

In this section, we investigate the function $g : y \mapsto \mathbf{E}[f(\mathsf{Pois}(y))]$ without referencing our load balancing or spanning tree applications. We prove several useful properties and results involving this function. Proofs in this section are not necessary to follow the reduction in Theorem 9.1, so the reader may choose to skip them. We note that these results are not new and most of them follow from elementary arguments. The most technical result in this section is a proof of Schur convexity of g, which is originally due to Rinott [37].

9.1.1 Basic Properties of $g(y)$

The following lemma — similar in spirit to Lemma 2.2 — gives convenient upper and lower bounds on $g(y)$.

Lemma 9.4. *Let $y \in \mathbb{R}^m_{\geq 0}$. We have* $\max\{f(y), 1 - \mathsf{exp}(-\mathsf{Top}_m(y))\} \leq g(y) \leq \mathsf{Top}_m(y)$.

Proof. The upper bound follows from Lemma 2.2 (recall that f is normalized) and Fact 2.1:

$$g(y) = \mathbf{E}[f(\mathsf{Pois}(y))] \leq \mathbf{E}[\mathsf{Top}_m(\mathsf{Pois}(y))] = \mathbf{E}\big[\sum_{i\in[m]} \mathsf{Pois}(y_i)\big] = \mathsf{Top}_m(y).$$

By convexity of norms, we get: $\mathbf{E}[f(\mathsf{Pois}(y))] \geq f(\mathbf{E}[\mathsf{Pois}(y)]) = f(y)$. The other lower bound uses Lemma 2.2:

$$\begin{aligned}\mathbf{E}[f(\mathsf{Pois}(y))] \geq \mathbf{E}[\mathsf{Top}_1(\mathsf{Pois}(y))] &\geq \mathbf{Pr}[\mathsf{Top}_1(\mathsf{Pois}(y)) > 0] \\ &= 1 - \prod_{i\in[m]} \mathbf{Pr}[\mathsf{Pois}(y_i) = 0] = 1 - \mathsf{exp}(-\mathsf{Top}_m(y)).\end{aligned}$$ ∎

The next result shows similarities between g and the monotone, symmetric norm f that gives rise to it.

Lemma 9.5.
Let $y, y' \in \mathbb{R}^m_{\geq 0}$. The function g has the following properties:

(a) *(monotonicity) $g(y) \geq g(y')$ whenever $y \geq y'$;*

(b) *(symmetry) $g(y) = g(y^{\downarrow})$;*

(c) *(subadditivity) $g(y + y') \leq g(y) + g(y')$; and*

(d) *(uniform continuity) $g(y) - g(y') \leq \|y - y'\|_1$.*

Proof. First of all, observe that symmetry of g follows from symmetry of f. For scalars $\lambda \geq \lambda' \geq 0$, the random variable $\mathsf{Pois}(\lambda)$ stochastically dominates $\mathsf{Pois}(\lambda')$: that is, we have $\mathbf{Pr}[\mathsf{Pois}(\lambda) \geq \theta] \geq \mathbf{Pr}[\mathsf{Pois}(\lambda') \geq \theta]$ for any $\theta \in \mathbb{R}_{\geq 0}$. Consider vectors $y, y' \in \mathbb{R}^m_{\geq 0}$

satisfying $y \geq y'$. Since $\mathsf{Pois}(y_i)$ stochastically dominates $\mathsf{Pois}(y'_i)$ for all $i \in [m]$, and f is monotone, we get that $\mathbf{E}[f(\mathsf{Pois}(y))] \geq \mathbf{E}[f(\mathsf{Pois}(y'))]$.

The triangle inequality for f implies subadditivity (or triangle inequality) for g. For scalars $\lambda_1, \lambda_2 \in \mathbb{R}_{\geq 0}$, $\mathsf{Pois}(\lambda_1+\lambda_2)$ has the same distribution as $\mathsf{Pois}(\lambda_1)+\mathsf{Pois}(\lambda_2)$ (Fact 2.2). Thus, for any $y, y' \in \mathbb{R}^m_{\geq 0}$, the distributions of $\mathsf{Pois}(y+y')$ and $\mathsf{Pois}(y)+\mathsf{Pois}(y')$ are identical. Therefore:

$$g(y + y') = \mathbf{E}[f(\mathsf{Pois}(y) + \mathsf{Pois}(y'))] \leq \mathbf{E}[f(\mathsf{Pois}(y)) + f(\mathsf{Pois}(y'))] = g(y) + g(y').$$

Finally, to prove uniform continuity of g, consider $y, y' \in \mathbb{R}^m_{\geq 0}$. Let $y'' \in \mathbb{R}^m_{\geq 0}$ be the pointwise minimum of y and y', i.e., for each $i \in [m]$, $y''_i := \min(y_i, y'_i)$. By construction, $\|y - y''\|_1 \leq \|y - y'\|_1$. Since g is monotone, $g(y) - g(y') \leq g(y) - g(y'')$ holds, and hence it suffices to work with y and y''. By subadditivity,

$$g(y) \leq g(y'') + g(y - y'') \leq g(y'') + \|y - y''\|_1 ,$$

where the second inequality follows from Lemma 9.4. ■

9.1.2 Schur Convexity of $g(y)$

We now prove the key technical result in this chapter: g is Schur-convex. This result is originally due to Rinott [37], and its proof can also be found in [31] (see Chapter 11, Proposition E.6). Since g is monotone, we state this result in a slightly stronger form.

Theorem 9.6.
Let $y, y' \in \mathbb{R}^m_{\geq 0}$. If $\mathsf{Top}_\ell(y) \leq \mathsf{Top}_\ell(y')$ for all $\ell \in [m]$, then $g(y) \leq g(y')$.

Proof. The stated inequality is related to a property called *Schur convexity* of the function g (see Chapter 3 in [31] for more details). Schur convexity is the property that the stated inequality holds when y and y' also satisfy $\mathsf{Top}_m(y) = \mathsf{Top}_m(y')$. However, since g is monotone, it suffices to restrict our attention to this case.[1] Furthermore, since g is symmetric, we may assume that $y = y^{\downarrow}$ and $y' = y'^{\downarrow}$. It is easy to show that there exists a finite sequence of vectors $y^{(0)} = y, y^{(1)}, y^{(2)}, \ldots, y^{(N)} = y'$ such that for all $k \in [N]$ we have (i) $y^{(k)} = y^{(k)\downarrow}$; (ii) for any $\ell \in [m]$, $\mathsf{Top}_\ell(y^{(k-1)}) \leq \mathsf{Top}_\ell(y^{(k)})$, and the inequality is tight for

[1] If $\mathsf{Top}_m(y) < \mathsf{Top}_m(y')$, then we can increase coordinates of y, without violating the bounds on the Top_ℓ-norms of y, until $\mathsf{Top}_m(y) = \mathsf{Top}_m(y')$.

$\ell = m$; and (iii) $y^{(k)}$ and $y^{(k-1)}$ differ in exactly two coordinates. Hence, it suffices to prove the lemma with the additional restriction that y and y' differ in exactly two coordinates (note that y and y' are nonincreasing). Let $1 \leq i < j \leq m$ such that $y_k = y'_k$ for all $k \in [m] \setminus \{i, j\}$. Thus, $y_i < y'_i$, $y_j > y'_j$, and $y_i + y_j = y'_i + y'_j$.

Let $\theta_3, \dots, \theta_m \in \mathbb{R}_{\geq 0}$ be $m-2$ scalars, and let $\vec{\theta} := (\theta_3, \dots, \theta_m)$. We define a symmetric 2-dimensional function $q : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{\geq 0}$ as follows:

$$q(a, b; \theta_3, \dots, \theta_m) := \mathbf{E}[f(\mathsf{Pois}(a), \mathsf{Pois}(b), \theta_3, \dots, \theta_m)].$$

That is, $q(a, b)$ is the expected f-norm of a vector whose first two coordinates are independent Poisson variables with parameters a and b, and the rest of the coordinates are fixed. To prove the lemma, it suffices to show that q is *Schur-convex* for any choice of (arbitrary but fixed) parameters $\theta_3, \dots, \theta_m$, i.e., for any $a \geq b \geq 0$, $a' \geq b' \geq 0$ satisfying $a \geq a'$ and $a + b = a' + b'$, we have $q(a, b) \geq q(a', b')$. To see why this suffices, let $Y_k \sim \mathsf{Pois}(y_k)$ be independent Poisson random variables, for all $k \in [m] \setminus \{i, j\}$. Let $Y_{-ij} := (Y_k)_{k \in [m] \setminus \{i,j\}}$. Then

$$\begin{aligned} g(y) &= \mathbf{E}[f(\mathsf{Pois}(y_1), \dots, \mathsf{Pois}(y_m))] \\ &= \mathbf{E}_{Y_{-ij}}[\mathbf{E}[f(\mathsf{Pois}(y_i), \mathsf{Pois}(y_j), Y_{-ij})]] = \mathbf{E}_{Y_{-ij}}[q(y_i, y_j; Y_{-ij})], \end{aligned}$$

In the double expectations above, the outer expectation is with respect to the distribution for Y_{-ij}. Similarly, we have $g(y') = \mathbf{E}_{Y_{-ij}}[q(y'_i, y'_j; Y_{-ij})]$. Therefore, proving Schur convexity for q implies Schur convexity for g. Since q is symmetric, a necessary and sufficient condition for q to be Schur-convex is given by the Schur-Ostrowski criterion (see [31], Chapter 3, Theorem A.4):

$$\forall\, a, b \in \mathbb{R}_{\geq 0},\ a \geq b \implies \frac{\partial q}{\partial a} \geq \frac{\partial q}{\partial b}. \qquad \text{(Schur-Ostrowski criterion)}$$

Fix some parameters $\theta_3, \dots, \theta_m \geq 0$. By definition,

$$q(a, b; \vec{\theta}) = \mathbf{E}\big[f(\mathsf{Pois}(a), \mathsf{Pois}(b), \vec{\theta})\big] = e^{-(a+b)} \cdot \sum_{k, \ell \geq 0} \frac{a^k}{k!} \frac{b^\ell}{\ell!} f(k, \ell, \vec{\theta})$$

Taking the partial derivative with respect to a gives:

$$\frac{\partial q}{\partial a} = e^{-(a+b)} \cdot \sum_{k, \ell \geq 0} \frac{a^k}{k!} \frac{b^\ell}{\ell!} \big(f(k+1, \ell, \vec{\theta}) - f(k, \ell, \vec{\theta})\big)$$

Similarly, taking the partial derivative with respect to b gives:

$$\frac{\partial q}{\partial b} = e^{-(a+b)} \cdot \sum_{k,\ell \geq 0} \frac{a^k}{k!} \frac{b^\ell}{\ell!} \big(f(k, \ell + 1, \vec{\theta}) - f(k, \ell, \vec{\theta}) \big)$$

So, Schur-Ostrowski criterion is equivalent to checking if the following expression is non-negative for all $a \geq b \geq 0$:

$$\sum_{k,\ell \geq 0} \frac{a^k}{k!} \frac{b^\ell}{\ell!} \big(f(k+1, \ell, \vec{\theta}) - f(k, \ell + 1, \vec{\theta}) \big) \overset{?}{\geq} 0.$$

In the above series, terms corresponding to $k = \ell$ vanish because f is symmetric. The residual series can be expressed as $\sum_{k > \ell \geq 0} \big(T_{k,\ell} / k!\, \ell! \big)$ where

$$T_{k,\ell} := a^k b^\ell \big(f(k+1, \ell, \vec{\theta}) - f(k, \ell + 1, \vec{\theta}) \big) + a^\ell b^k \big(f(\ell + 1, k, \vec{\theta}) - f(\ell, k+1, \vec{\theta}) \big).$$

By symmetry of f, for any $k > \ell \geq 0$,

$$T_{k,\ell} = a^\ell b^\ell (a^{k-\ell} - b^{k-\ell}) \big(f(k+1, \ell, \vec{\theta}) - f(k, \ell + 1, \vec{\theta}) \big).$$

Nonnegativity of $T_{k,\ell}$ easily follows because $a \geq b$ and $k > \ell$: $f(k+1, \ell, \vec{\theta}) \geq f(k, \ell + 1, \vec{\theta})$ since for any $r \in [m]$, we have $\mathsf{Top}_r((k+1, \ell, \vec{\theta})) \geq \mathsf{Top}_r((k, \ell + 1, \vec{\theta}))$. This completes the proof of Schur convexity of g. ■

9.1.3 Subhomogeneity of $g(y)$

The last property of g that will be useful to us is subhomogeneity. Note that norms are homogeneous.

Lemma 9.7 (Subhomogeneity)**.** *For any $y \in \mathbb{R}^m_{\geq 0}$ and scalar $\theta \geq 1$, we have $g(\theta y) \leq \theta \cdot g(y)$.*

Proof. We prove the lemma for rational θ. The proof for general θ then follows from a continuity argument. Let $\theta = a/b$ for integers $a > b \geq 1$. (If $a = b$, there is nothing to be shown.) Observe that $g(\theta y) = g(az)$ and $g(y) = g(bz)$, where $z = y/b$. Thus, for the rational case, it suffices to prove that $b \cdot g(az) \leq a \cdot g(bz)$ holds for all $z \in \mathbb{R}^m_{\geq 0}$ and integers $a > b \geq 1$.

Fix some $z \in \mathbb{R}^m_{\geq 0}$. Let $Z^{(0)}, Z^{(1)}, \ldots, Z^{(a-1)}$ be a independent random vectors that are identically distributed copies of $\mathsf{Pois}(z)$ (thus each $Z_i^{(j)}$ is an independent $\mathsf{Pois}(z_i)$ random variable). For any $i \in [m]$, the distribution of $\mathsf{Pois}(az_i)$ is identical to that of $\sum_{j=0}^{a-1} Z_i^{(j)}$ (Fact 2.2), therefore

$$g(az) = \mathbf{E}[f(\mathsf{Pois}(az))] = \mathbf{E}\Big[f\Big(\sum_{j=0}^{a-1} Z^{(j)}\Big)\Big].$$

Also, for any subset $S \subseteq \{0, 1, \ldots, a\}$ with $|S| = b$, we have $\mathbf{E}[f(\sum_{j \in S} Z^{(j)})] = g(bz)$. Define size-$b$ index sets $S_k := \{(k+j) \bmod a : j = 0, \ldots, b-1\}$, for $k = 0, 1, \ldots, a-1$. Note that each $j \in \{0, \ldots, a-1\}$ is contained in exactly b sets in $\{S_k\}_k$. Therefore:

$$g(az) = \mathbf{E}\Big[f\Big(\sum_{j=0}^{a-1} Z^{(j)}\Big)\Big] = \mathbf{E}\Big[f\Big(\tfrac{1}{b} \cdot \sum_{k=0}^{a-1} \sum_{j \in S_k} Z^{(j)}\Big)\Big] \leq \frac{1}{b} \cdot \sum_{k=0}^{a-1} \mathbf{E}\Big[f\Big(\sum_{j \in S_k} Z^{(j)}\Big)\Big] = \frac{a}{b} \cdot g(bz).$$

Now suppose that θ is irrational. Fix some $y \in \mathbb{R}^m_{\geq 0}$ and consider an increasing sequence of rationals $\{\theta_r\}_{r \geq 1}$ that converge to θ. From the rational case, the inequality $g(\theta_r y) \leq \theta_r g(y) \leq \theta g(y)$ holds for all $r \in \{1, 2, \ldots\}$. Since g is continuous, we have $g(\theta y) = \sup_{r \geq 1} g(\theta_r y) \leq \theta g(y)$. ■

9.1.4 Estimator for $g(y)$

In the proof of Theorem 9.1, we work with multiple guesses of $\{\mathsf{Top}_\ell(\Lambda^*)\}_{\ell \in \mathsf{POS}}$, and each such guess gives rise to an assignment of jobs to machines. The output of our algorithm (for PoisNormLB) is the assignment with the smallest g-objective. Since we are not aware of any results that allow us to directly compare $g(y)$ and $g(y')$ for arbitrary $y, y' \in \mathbb{R}_{\geq 0}$, we instead use a sampling-based estimator for our purposes. We remark that this estimator is the reason why our reduction from PoisNormLB to MinNormLB leads to probabilistic guarantees. For technical reasons that will be clear shortly, in the following result we assume that the Top_ℓ norm of y is bounded away from 0.

Lemma 9.8. *Let $\varepsilon, \eta \in (0, 1/2]$ be some error parameters, and $y \in \mathbb{R}^m_{\geq 0}$ be such that $\mathsf{Top}_m(y) \geq \varepsilon$. Define $N := 2m^2/\varepsilon^3\eta$. Using at most N independent samples from $\mathsf{Pois}(y)$, we can compute an estimate γ satisfying:*

$$\mathbf{Pr}\big[(1-\varepsilon)g(y) \leq \gamma \leq (1+\varepsilon)g(y)\big] \geq 1 - \eta.$$

Proof. Let $x^{(1)}, \ldots, x^{(N)}$ be N independent samples from $\mathsf{Pois}(y)$, and let γ denote the sample f-average $\big(\sum_{j=1}^{N} f(x^{(j)})\big)/N$. We show that γ is the desired estimator by using Chebyshev's concentration inequality. To this end, we need a bound on the variance of the real-valued random variable $f(\mathsf{Pois}(y))$. We have:

$$\begin{aligned}
\mathrm{Var}\big[f(\mathsf{Pois}(y))\big] &\leq \mathbf{E}[f^2(\mathsf{Pois}(y))] \\
&\leq \mathbf{E}[\mathsf{Top}_m^2(\mathsf{Pois}(y))] && \text{(By Lemma 2.2)} \\
&= \mathrm{Var}\big[\mathsf{Top}_m(\mathsf{Pois}(y))\big] + \big(\mathbf{E}[\mathsf{Top}_m(\mathsf{Pois}(y))]\big)^2 \\
&= \mathrm{Var}\big[\mathsf{Pois}(\mathsf{Top}_m(y))\big] + \big(\mathbf{E}[\mathsf{Pois}(\mathsf{Top}_m(y))]\big)^2 && \text{(By Fact 2.2)} \\
&= \mathsf{Top}_m(y)(1 + \mathsf{Top}_m(y)). && \text{(By Fact 2.1)}
\end{aligned}$$

Since γ is an average of $f(\mathsf{Pois}(y))$ over N independent samples, we get that $\mathrm{Var}[\gamma] = \mathrm{Var}\big[f(\mathsf{Pois}(y))\big]/N$. By Chebyshev's inequality (Lemma 2.27):

$$\mathbf{Pr}[|\gamma - g(y)| > \varepsilon g(y)] \leq \frac{\mathrm{Var}[\gamma]}{\varepsilon^2 g^2(y)} = \frac{1}{N\varepsilon^2} \cdot \frac{\mathsf{Top}_m(y)}{g(y)} \cdot \frac{1 + \mathsf{Top}_m(y)}{g(y)} \leq \frac{2m^2}{N\varepsilon^3} = \eta.$$

We use Lemma 9.4 to justify the penultimate inequality used above. Since $\mathsf{Top}_m(y)/g(y) \leq m$ holds for any $y \in \mathbb{R}^m_{\geq 0}$, the required inequality holds whenever $\mathsf{Top}_m(y) \geq 1$. Now suppose that $\mathsf{Top}_m(y) \in [\varepsilon, 1]$ holds. Then we have $1 + \mathsf{Top}_m(y) \leq 2$ and $g(y) \geq f(y) \geq \mathsf{Top}_m(y)/m \geq \varepsilon/m$, and hence the required inequality follows. ■

Estimating the value of $g(y)$ when $\mathsf{Top}_m(y)$ is close to zero is much simpler: we simply use $\mathsf{Top}_m(y)$ as an approximation for $g(y)$.

Lemma 9.9. *Let $\varepsilon \in (0, 1/2]$ be an error parameter, and $y \in \mathbb{R}^m_{\geq 0}$ be such that $\mathsf{Top}_m(y) \leq \varepsilon$. We have $g(y) \leq \mathsf{Top}_m(y) \leq (1 + 2\varepsilon)g(y)$.*

Proof. We have already proved the lower bound. For the upper bound, we use Lemma 9.4 and the assumption that $\mathsf{Top}_m(y) \leq \varepsilon \leq 1/2$ holds:

$$g(y) \geq 1 - \exp(-\mathsf{Top}_m(y)) \geq \mathsf{Top}_m(y)(1 - \mathsf{Top}_m(y)) \geq \mathsf{Top}_m(y)/(1 + 2\varepsilon).$$

In the penultimate inequality we use $1 - e^{-z} \geq z - z^2$, which holds for any $z \in \mathbb{R}_{\geq 0}$, and in the last inequality we use $(1 - \varepsilon)(1 + 2\varepsilon) \geq 1$. ■

9.2 Reduction from **PoisNormLB** to **MinNormLB**

Using the tools from Section 9.1, we give a proof of our main theorem in this chapter. Recall that the $\mathsf{PoisNormLB}$-instance is given by $\mathcal{I}^{\mathsf{Pois}} = (J, [m], \{\lambda_{ij}\}, f)$, and we have access to a ρ-approximation algorithm $\mathcal{A}^{\mathsf{Det}}$ for $\mathsf{MinNormLB}$-instances with job-set J, machine-set $[m]$, and $\{\lambda_{ij}\}_{i\in[m],j\in J}$ job sizes. We are also given two error parameters $\varepsilon, \eta > 0$, and the goal is to obtain a $\rho\big(1 + O(\varepsilon)\big)$-approximate solution for $\mathcal{I}^{\mathsf{Pois}}$ with probability at least $1 - \eta$. Recall that $\Lambda^* = \Lambda^{\sigma^*}$ denotes the aggregate λ-vector for the optimal assignment σ^*.

Proof of Theorem 9.1. Set $\varepsilon' = \delta = \min\{0.5, \varepsilon\}$. Let σ^{sum} be the assignment that minimizes the expected sum of machine loads, so $\sigma^{\mathsf{sum}}(j) = \arg\min_i \lambda_{ij}$ for all jobs $j \in J$. Let $\Lambda^{\mathsf{sum}} := \Lambda^{\sigma^{\mathsf{sum}}}$, and $\mathsf{UB} := \mathsf{Top}_m(\Lambda^{\mathsf{sum}}) = \sum_{j\in J}\big(\min_{i\in[m]} \lambda_{ij}\big)$. Note $\mathsf{Top}_m(\Lambda^*) \geq \mathsf{UB}$.

If $\mathsf{UB} \leq \varepsilon'$, then we claim that σ^{sum} is a $(1 + 2\varepsilon')$-approximate solution to $\mathcal{I}^{\mathsf{Pois}}$. This follows from Lemma 9.4: (i) $g(\Lambda^{\mathsf{sum}}) \leq \mathsf{Top}_m(\Lambda^{\mathsf{sum}}) = \mathsf{UB}$; and (ii) noting that $\mathsf{UB} \leq \varepsilon' \leq 1/2$:

$$g(\Lambda^*) \geq 1 - \mathsf{exp}(-\mathsf{Top}_m(\Lambda^*)) \geq 1 - \mathsf{exp}(-\mathsf{UB}) \geq \mathsf{UB}(1 - \mathsf{UB}) \geq \mathsf{UB}/(1 + 2\varepsilon').$$

Now suppose that $\mathsf{UB} \geq \varepsilon'$; note that $\mathsf{Top}_m(\Lambda^\sigma) \geq \varepsilon'$ for every assignment σ. We have $\mathsf{Top}_1(\Lambda^*) \geq \mathsf{Top}_m(\Lambda^*)/m \geq \mathsf{UB}/m$. Also, $\mathsf{Top}_m(\Lambda^*) \leq m\mathsf{Top}_1(\Lambda^*)$, and we have $\mathsf{Top}_1(\Lambda^*) \leq f(\Lambda^*) \leq g(\Lambda^*) \leq g(\Lambda^{\mathsf{sum}}) \leq \mathsf{UB}$; here, we have used Lemma 2.2, Lemma 9.4 (twice), and the optimality of σ^*. Thus we obtain $\mathsf{Top}_m(\Lambda^*) \leq m \cdot \mathsf{UB}$.

We recall the iterative definition of $\mathsf{POS}_{m,\delta} \subseteq [m]$: include the index 1 in $\mathsf{POS}_{m,\delta}$; as long as the largest index $\ell \in \mathsf{POS}_{m,\delta}$ is such that $\lceil(1+\delta)\ell\rceil \leq m$, include $\lceil(1+\delta)\ell\rceil$ (which is larger than ℓ) in $\mathsf{POS}_{m,\delta}$. For notational convenience, let $\mathsf{POS} := \mathsf{POS}_{m,\delta}$. Consider the non-increasing vector u which is $(\Lambda^*_\ell)_{\ell\in\mathsf{POS}}$ with its coordinates listed in *decreasing* order of ℓ. We apply Lemma 2.21 (b) on u, taking $\mathcal{L} = \mathsf{POS}$, and upper and lower bounds $m \cdot \mathsf{UB}$ and UB/m respectively, to obtain a $\mathrm{poly}(m^{1/\delta})$-size set $\mathcal{T} \subset \mathbb{R}^{\mathsf{POS}}_{\geq 0}$ containing a vector B^* such that $\mathsf{Top}_\ell(\Lambda^*) \leq B^*_\ell \leq (1+\delta)\mathsf{Top}_\ell(\Lambda^*)$ for all $\ell \in \mathsf{POS}$.

For each $B \in \mathcal{T}$, we do the following. Let $h_B : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ be the monotone, symmetric norm defined as $h_B(v) := \max_{\ell\in\mathsf{POS}} \mathsf{Top}_\ell(v)/B_\ell$. We run $\mathcal{A}^{\mathsf{Det}}$ on the $\mathsf{MinNormLB}$ instance with $\{\lambda_{ij}\}_{i\in[m],j\in J}$ job sizes and norm h_B, to obtain an assignment σ^B. Let $\Lambda^B = \Lambda^{\sigma^B}$. We use Lemma 9.8 to compute an estimate γ^B satisfying:

$$\mathbf{Pr}[(1-\varepsilon')g(\Lambda^B) \leq \gamma^B \leq (1+\varepsilon')g(\Lambda^B)] \geq 1 - \frac{\eta}{|\mathcal{T}|}.$$

We output the assignment $\sigma^{B'}$ with the smallest estimator-value $\gamma^{B'}$ among all $B' \in \mathcal{T}$.

We argue that $\sigma^{B'}$ is a $\rho\big(1+O(\varepsilon)\big)$-approximate assignment for $\mathcal{I}^{\mathsf{Pois}}$ with probability at least $1-\eta$. By the union bound, with probability at least $1-\eta$, the inequality $(1-\varepsilon')g(\Lambda^B) \leq \gamma^B \leq (1+\varepsilon')g(\Lambda^B)$ holds for all $B \in \mathcal{T}$; we assume that this event happens. Consider the correct guess $B^* \in \mathcal{T}$. Since $h_{B^*}(\Lambda^*) \leq 1$, the solution σ^{B^*} satisfies:

$$\mathsf{Top}_\ell(\Lambda^{B^*}) \leq \rho B^*_\ell \leq \rho(1+\delta)\mathsf{Top}_\ell(\Lambda^*) \text{ for all } \ell \in \mathsf{POS}.$$

Lemma 2.12 gives a slightly weaker bound for all Top_ℓ norms of Λ^{B^*}:

$$\mathsf{Top}_\ell(\Lambda^{B^*}) \leq \rho(1+\delta)^2\mathsf{Top}_\ell(\Lambda^*) \text{ for all } \ell \in [m].$$

By majorization (Theorem 9.6) and subhomogeneity (Lemma 9.7):

$$g(\Lambda^{B^*}) \leq \rho(1+\delta)^2 g(\Lambda^*).$$

Accounting for the error due to the γ^B estimates, yields

$$\mathbf{E}\Big[f\Big(\overrightarrow{\mathsf{load}}^{\sigma^{B'}}\Big)\Big] = g(\Lambda^{B'}) \leq \frac{1+\varepsilon'}{1-\varepsilon'} \cdot g(\Lambda^{B^*}) \leq \rho(1+13\varepsilon')g(\Lambda^*).$$

■

Chapter 10

Stochastic Minimum Norm Spanning Tree

We now apply our framework to devise approximation algorithms for *stochastic minimum norm spanning tree*, which we abbreviate to StochNormTree. We state the problem and mention our main results in Section 10.1. We design an $O(1)$-approximation algorithm for StochNormTree with Top_ℓ norms in Section 10.2, and generalize this result to arbitrary monotone symmetric norms in Section 10.3. Our results on StochNormTree easily carry-over to three natural extensions: *stochastic min-norm matroid basis*, *stochastic min-norm degree-bounded spanning tree*, and *stochastic min-norm traveling salesperson problem*; we discuss these extensions in Section 10.4.

10.1 Problem Statement

In an instance of StochNormTree, we are given an undirected graph $G = (V, E)$ with stochastic edge-weights and we are interested in a low-weight solution that connects the vertices of this graph to each other. More precisely, for each edge $e \in E$, we are given a nonnegative random variable X_e that denotes its weight. Edge weights are independent. A feasible solution to this problem is a spanning tree $T \subseteq E$ of G. This induces a random weight vector $Y^T = (X_e)_{e \in T}$; note that Y^T follows a product distribution on $\mathbb{R}_{\geq 0}^{n-1}$ where $n := |V|$. The objective in StochNormTree is to find a spanning tree T that minimizes $\mathbf{E}[f(Y^T)]$ for a given monotone, symmetric norm $f : \mathbb{R}_{\geq 0}^{n-1} \to \mathbb{R}_{\geq 0}$.

As an aside, we note that the deterministic version of this problem, wherein each edge e has a fixed weight $w_e \geq 0$, can be solved optimally. This is because a minimum-weight spanning tree (MST) T^* simultaneously minimizes $\sum_{e \in T}(w_e - \theta)^+$ among all spanning trees T of G, for all $\theta \geq 0$. Since $\mathsf{Top}_\ell(x) = \min_{\theta \in \mathbb{R}_{\geq 0}}\{\ell\theta + \sum_i (x_i - \theta)^+\}$ (see Lemma 2.4), it follows that T^* minimizes every Top_ℓ norm, and so by Theorem 2.7, it is *simultaneously* optimal for *every* monotone, symmetric norm. Furthermore, due to Theorem 9.6 (also see Theorem 9.1), T^* is also optimal for the stochastic setting when edge-weight variables follow Poisson distributions. Also note that when f is the Top_{n-1} norm, StochNormTree reduces to the minimum-weight spanning tree problem with (deterministic) edge-weights given by $w_e = \mathbf{E}[X_e]$.

10.1.1 Our Results

Our main results in this chapter are the following.

Theorem 10.1.
There is an $O(1)$-approximation algorithm for stochastic minimum norm spanning tree (i.e., with arbitrary edge-weight distributions and an arbitrary monotone, symmetric norm).

We get improved approximation guarantees for the Top_ℓ case.

Theorem 10.2.
Let $\varepsilon > 0$ be a constant. For any $\ell \in [n-1]$, there is a simple $(2+\varepsilon)$-approximation algorithm for stochastic Top_ℓ-norm spanning tree.
Furthermore, for the stochastic bottleneck spanning tree problem (i.e., the $\ell = 1$ case), the approximation guarantee can be improved to $e/(e-1) + \varepsilon$.

10.1.2 Overview

We give a high-level overview of the main ideas used in our proofs. Let T^* denote an optimal spanning tree. Let $Y^* := Y^{T^*}$ and $\mathsf{OPT} := \mathbf{E}[f(Y^*)]$ denote the induced cost vector and the optimal objective value, respectively. We drop the superscript T in Y^T when the tree T is clear from the context. We use Y_e to denote the coordinate in Y corresponding to edge e; we will ensure that Y_e is used only when $e \in T$. Also note that since f is symmetric, the order in which the edge-variables Y_e appear in Y is immaterial. We recall some notation and definitions from Chapters 2 and 3 that are frequently used in this chapter. The domain of f has dimension $n-1$, so we reserve $\mathsf{POS} = \{1, 2, 4, \ldots, 2^{\lfloor \log_2(n-1) \rfloor}\}$ to refer to the set of powers of 2 that are at most $n-1$. [1] The $\tau_\ell(Y)$ order statistic (see Definition 3.7) that is a proxy for the ℓ^{th} largest coordinate of Y is given by:

$$\tau_\ell(Y) := \inf\Big\{\theta \in \mathbb{R}_{\geq 0} : \sum_{e\in T} \mathbf{Pr}[Y_e > \theta] < \ell\Big\} \tag{10.1}$$

The $\gamma_\ell(Y)$ proxy function (see Definition 3.8) that serves as a constant-factor approximation to $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ is given by:

$$\gamma_\ell(Y) := \ell\tau_\ell(Y) + \sum_{e\in T} \mathbf{E}\left[\left(Y_e - \tau_\ell(Y)\right)^+\right] \tag{10.2}$$

The cost vector Y^T is inherently less complex than the load vector in load balancing, in that each coordinate Y_e^T is an "atomic" random variable whose distribution we can directly access. Thus, our approach is guided by Theorems 4.4 and 4.16, which show that to obtain an approximation guarantee for stochastic f-norm spanning tree, it suffices to find a spanning tree T such that: (i) the τ_ℓ statistics of Y^T are "comparable" to those of Y^*; and (ii) the total mass of $\{X_e\}_{e\in T}$ that lies above the $\tau_1(Y^*)$ threshold is bounded by the corresponding total mass for $\{X_e\}_{e\in T^*}$.

Our approximation strategy is based on solving an LP-relaxation of $\mathsf{StochNormTree}$ to obtain a fractional spanning tree whose "cost" is $O(\mathsf{OPT})$, which we then round using the iterative rounding framework (Theorem 2.32) to obtain an $O(1)$-approximate (integral) spanning tree. More concretely, our LP consists of constraints that define the spanning-tree polytope and a collection of side constraints that impose that the τ_ℓ-statistic of our solution be bounded by t_ℓ, where t_ℓ is a guess for $\tau_\ell(Y^*)$; from the expression in (10.1), it should be clear that the τ_ℓ constraints are easy to linearize. We control the contribution from $\sum_{e\in T}\mathbf{E}[(X_e-\tau_1)^+]$ by moving it to the objective function. For algorithmic purposes, we include τ_ℓ constraints only for indices $\ell \in \mathsf{POS}$. This is done for two reasons: first, we are not aware of any efficient procedure to simultaneously guess $\tau_\ell(Y^*)$ for all $\ell \in [m]$; and second, the approximation-quality of the rounded solution depends on the maximum column-sum in the LP-constraint matrix, so it makes sense to work with a sparse collection of indices that yield the desired approximation guarantee. We give all the details in Section 10.3.

Our approximation algorithm for the Top_ℓ case is substantially simpler because we only have to work with a single τ_ℓ statistic. We remark that the objective function now has a term of the form $\sum_{e\in T}\mathbf{E}[(X_e-\tau_\ell)^+]$ since we are controlling $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ by controlling the $\gamma_\ell(Y)$ proxy function. We start with the simpler Top_ℓ case.

[1] In other chapters, we used m to refer to the dimension, and defined POS to be the set of powers of 2 that are at most m.

10.2 Stochastic Spanning Tree with Top_ℓ Norms

Fix an index $\ell \in [m]$ for the rest of this section, and let $\varepsilon > 0$ be a constant. We now describe an LP-based algorithm that finds a spanning tree T satisfying $\gamma_\ell(Y^T) \leq (1 + O(\varepsilon))\gamma_\ell(Y^*)$. Since $\gamma_\ell(Y^*) \leq 2\mathbf{E}[\mathsf{Top}_\ell(Y^*)]$ (see Theorem 3.9), this would imply that T is a $(2 + O(\varepsilon))$-approximate solution to the given instance of stochastic Top_ℓ-norm spanning tree.

We use the following linear description of the spanning-tree polytope:

$$\mathcal{Q}^{\mathsf{tree}} := \left\{z \in \mathbb{R}^E_{\geq 0} : \quad z(E) = |V| - 1, \qquad z(A) \leq n - \mathsf{comp}(A) \quad \forall A \subseteq E\right\}.$$

In the above, the z_e variables indicate if an edge e belongs to the spanning tree, $z(A)$ denotes $\sum_{e \in A} z_e$, and $\mathsf{comp}(A)$ denotes the number of connected components in the subgraph (V, A). It is well-known that the above polytope is the spanning-tree polytope of G, i.e., it is the convex hull of indicator vectors of spanning trees of G. We use the matroid base-polytope characterization for two reasons: (i) our arguments can be generalized verbatim to the more general setting with an arbitrary matroid; and (ii) we can directly invoke Theorem 2.32 to round fractional LP solutions to integral solutions.

LP Relaxation. Let $t_\ell \in \mathbb{R}_{\geq 0}$ be a guess for $\tau_\ell(Y^*)$. The constraints of our LP encode that $z \in \mathcal{Q}^{\mathsf{tree}}$ and that the τ_ℓ-statistic of the cost vector induced by z is bounded by t_ℓ. The objective function captures the contribution from cost vector entries that lie above the t_ℓ threshold.

$$\mathsf{Tree\text{-}OPT}\big(\ell, t_\ell\big) := \min \quad \sum_{e \in E} \mathbf{E}[(X_e - t_\ell)^+] z_e$$

$$(\mathsf{Tree\text{-}LP}(\ell, t_\ell)) \qquad \text{s.t.} \quad \sum_{e \in E} \mathbf{Pr}[X_e > t_\ell] z_e \leq \ell \tag{10.3}$$

$$z \in \mathcal{Q}^{\mathsf{tree}}. \tag{10.4}$$

The following results are straightforward.

Claim 10.3. *For any scalar $t_\ell \geq \tau_\ell(Y^*)$, $(\mathsf{Tree\text{-}LP}(\ell, t_\ell))$ is feasible, and its optimal objective value $\mathsf{Tree\text{-}OPT}\big(\ell, t_\ell\big)$ is at most $\sum_{e \in T^*} \mathbf{E}\big[\big(X_e - \tau_\ell(Y^*)\big)^+\big]$.*

Proof. Consider the solution z^* induced by the optimal solution T^*, i.e., z^* is the indicator vector of T^*. Constraint (10.3) is satisfied because $t_\ell \geq \tau_\ell(Y^*)$. The objective value of z^* is $\sum_{e \in T^*} \mathbf{E}\big[\big(X_e - t_\ell\big)^+\big]$, which is at most the stated bound because $t_\ell \geq \tau_\ell(Y^*)$. ∎

Theorem 10.4. *Let $t_\ell \in \mathbb{R}_{\geq 0}$ be a scalar such that (Tree-LP(ℓ, t_ℓ)) is feasible. Let T be a minimum-weight spanning tree in the graph $G = (V, E)$ with (deterministic) edge-weights $w_e := \mathbf{E}[(X_e - t_\ell)^+]$, and let $Y = Y^T$ denote the cost vector induced by T in the given* StochNormTree *instance. Then, $\mathbf{E}[\mathsf{Top}_\ell(Y)] \leq \gamma_\ell(Y) \leq \ell t_\ell + \mathsf{Tree\text{-}OPT}\big(\ell, t_\ell\big)$.*

Proof. By Lemma 3.11(i), we get $\gamma_\ell(Y) \leq \ell t_\ell + \sum_{e \in T} \mathbf{E}[(X_e - t_\ell)^+]$. Since T is a min-weight spanning tree, the second term in the above expression is bounded by $\mathsf{Tree\text{-}OPT}\big(\ell, t_\ell\big)$. ∎

We remark that the above proof only uses the feasibility of (Tree-LP(ℓ, t_ℓ)), and the integrality of the spanning-tree polytope.

10.2.1 Our Algorithm and Analysis

Given Claim 10.3 and Theorem 10.4, it is clear that if we work with a guess t_ℓ that is sufficiently close to $\tau_\ell(Y^*)$, then we obtain a solution with expected Top_ℓ cost close to $\ell\tau_\ell(Y^*) + \mathsf{Tree\text{-}OPT}\big(\ell, \tau_\ell(Y^*)\big) \leq \gamma_\ell(Y^*)$, and thus a near-optimal solution to the given instance of StochNormTree.

It is easy to obtain (polytime computable) lower and upper bounds on $\tau_\ell(Y^*)$, and thereby find a polynomial-size set $\mathcal{T}$ containing a good guess t_ℓ^* of $\tau_\ell(Y^*)$ (Claim 10.7). In order to select a suitable $t_\ell \in \mathcal{T}$, we need an estimator for $\gamma_\ell(Y)$, where Y is the cost vector induced by the spanning tree T mentioned in Theorem 10.4. We argue that, roughly speaking, $\ell t_\ell + \mathsf{Tree\text{-}OPT}\big(\ell, t_\ell\big)$ can be used as such an estimator; in particular, for $t_\ell = t_\ell^*$, this quantity is quite close to $\gamma_\ell(Y^*)$. Putting everything together yields the following algorithm.

Description of our algorithm. Define UB to be the weight of a minimum-weight spanning tree in G with edge-weights given by (the deterministic quantity) $w_e := \mathbf{E}[X_e]$. Define

$$\mathcal{T} := \Big\{t_\ell \in \mathbb{R}_{\geq 0} : \frac{2\varepsilon\,\mathsf{UB}}{(1+\varepsilon)n^2} \leq t_\ell \leq 2(1+\varepsilon)\,\mathsf{UB},\ t_\ell \text{ is a power of } (1+\varepsilon)\Big\} \tag{10.5}$$

Let $t'_\ell \in \mathcal{T}$ be such that (Tree-LP(ℓ, t'_ℓ)) is feasible and the quantity $\ell t'_\ell + \mathsf{Tree\text{-}OPT}\big(\ell, t'_\ell\big)$ is minimized. Note here that (Tree-LP(ℓ, t'_ℓ)) can be solved in polytime since one can efficiently separate over $\mathcal{Q}^{\mathsf{tree}}$. Return the minimum-weight spanning tree T in the graph $G = (V, E)$ with (deterministic) edge-weights given by $w_e := \mathbf{E}[(X_e - t'_\ell)^+]$.

Analysis. We first argue that UB can be used to obtain both upper and lower bounds on the optimum (Lemma 10.5), and hence that $\mathcal{T}$ contains a good guess of $\tau_\ell(Y^*)$ (Claim 10.7).

Lemma 10.5. *Let* UB *denote the weight of a minimum-weight spanning tree in* G *with edge-weights given by (the deterministic quantity)* $w_e := \mathbf{E}[X_e]$. *The following inequality holds for any monotone, symmetric norm* f*:*

$$\frac{\mathsf{UB}}{n-1} \le \mathbf{E}[f(Y^*)] \le \mathsf{UB}.$$

Proof. This follows from Lemma 2.2. For more details, see the proof of Lemma 6.10 which was used in the load balancing setting. ∎

Lemma 10.6. *We have* $\tau_\ell(Y^*) \le \gamma_\ell(Y^*) \le 2\,\mathsf{UB}$.

Proof. Follows from Theorem 3.9: $\tau_\ell(Y^*) \le \gamma_\ell(Y^*) \le 2\,\mathbf{E}[\mathsf{Top}_\ell(Y^*)] \le 2\,\mathsf{UB}$. ∎

Claim 10.7. *We have: (i)* $|\mathcal{T}| = \mathrm{poly}(n, 1/\varepsilon)$*; and (ii) there exists a scalar* $t^*_\ell \in \mathcal{T}$ *that satisfies* $\tau_\ell(Y^*) \le t^*_\ell \le \max\big\{(1+\varepsilon)\tau_\ell(Y^*), 2\varepsilon\,\mathsf{UB}/n^2\big\}$.

Proof. The first claim is straightforward: $|\mathcal{T}| \le 1 + \lceil \log_{1+\varepsilon}\big((1+\varepsilon)^2 n^2/\varepsilon\big)\rceil = O(\log n/\varepsilon^2)$. For the second claim, consider the smallest scalar t^*_ℓ that is a power of $(1+\varepsilon)$ and satisfies $t^*_\ell \ge \tau_\ell(Y^*)$. Note $t^*_\ell \le 2(1+\varepsilon)\,\mathsf{UB}$ by Lemma 10.6. If $t^*_\ell \in \mathcal{T}$, then t^*_ℓ is the desired scalar. Otherwise, the smallest scalar in $\mathcal{T}$ satisfies the claim. ∎

Now, we show that if t^*_ℓ is a good estimate of $\tau_\ell(Y^*)$, then $\ell t^*_\ell + (\mathsf{Tree\text{-}LP}(\ell, t^*_\ell))$ is a good estimate of $\gamma_\ell(Y^*)$.

Lemma 10.8. *Let* t^*_ℓ *be such that* $\tau_\ell(Y^*) \le t^*_\ell \le \max\big\{(1+\varepsilon)\tau_\ell(Y^*), 2\varepsilon\,\mathsf{UB}/n^2\big\}$. *Then,* $\ell t^*_\ell + \mathsf{Tree\text{-}OPT}\big(\ell, t^*_\ell\big) \le (1+O(\varepsilon))\gamma_\ell(Y^*)$.

Proof. By the definition of γ_ℓ and the bounds in Claim 10.3 and Lemma 10.5, we get:

$$\begin{aligned}
\ell t^*_\ell + \mathsf{Tree\text{-}OPT}\big(\ell, t^*_\ell\big) &\le (1+\varepsilon)\ell\tau_\ell(Y^*) + \frac{2\varepsilon \cdot \mathsf{UB} \cdot \ell}{n^2} + \sum_{e \in T^*} \mathbf{E}\big[\big(X_e - \tau_\ell(Y^*)\big)^+\big] \\
&\le \gamma_\ell(Y^*) + \varepsilon\ell\tau_\ell(Y^*) + 2\varepsilon\,\mathsf{UB}/n \\
&\le (1+3\varepsilon)\gamma_\ell(Y^*).
\end{aligned}$$

∎

We now finish the proof of Theorem 10.2.

Proof of Theorem 10.2. Let $t'_\ell \in \mathcal{T}$ and T be as defined in our algorithm (see Algorithm 10.2.1). Let $Y = Y^T$ denote the cost vector induced by the spanning tree T in the given $\mathsf{StochNormTree}$ instance. By Theorem 10.4, we have $\mathbf{E}[\mathsf{Top}_\ell(Y)] \le \ell t'_\ell + \mathsf{Tree\text{-}OPT}(\ell, t'_\ell)$. Since t'_ℓ minimizes $\ell t'_\ell + \mathsf{Tree\text{-}OPT}(\ell, t'_\ell)$ among all feasible guesses, by Claim 10.7 and Lemma 10.8 we get:

$$\mathbf{E}[\mathsf{Top}_\ell(Y)] \le (1+3\varepsilon)\gamma_\ell(Y^*) \le \begin{cases} (2+O(\varepsilon))\mathbf{E}[\mathsf{Top}_\ell(Y^*)] \text{ if } \ell > 1 \\ \left(\frac{e}{e-1} + O(\varepsilon)\right)\mathbf{E}[\mathsf{Top}_1(Y^*)] \text{ if } \ell = 1. \end{cases} \qquad \blacksquare$$

10.3 Stochastic Spanning Tree with Monotone Symmetric Norms

We now consider general $\mathsf{StochNormTree}$ where $f : \mathbb{R}^{n-1}_{\ge 0} \to \mathbb{R}_{\ge 0}$ is an arbitrary monotone, symmetric norm. Our approximation algorithm for this setting is based on rounding a fractional LP-solution whose cost vector has τ_ℓ statistics that are comparable to $\tau_\ell(Y^*)$ for all $\ell \in \mathsf{POS} = \{1, 2, \ldots, 2^{\lfloor \log_2(n-1) \rfloor}\}$. Like in the Top_ℓ case, we use the LP-objective to control the contribution from the coordinates larger than τ_1 to the expected f-norm.

LP Relaxation. Let $\vec{t} \in \mathbb{R}^{\mathsf{POS}}_{\ge 0}$ be a nonincreasing vector. We define the linear relaxation (Tree-LP($\vec{t}$)) as follows.

$$\text{(Tree-LP}(\vec{t})) \qquad \begin{aligned} \mathsf{Tree\text{-}OPT}(\vec{t}) := \min \quad & \sum_{e \in E} \mathbf{E}[(X_e - t_1)^+] z_e \\ \text{s.t.} \quad & \sum_{e \in E} \mathbf{Pr}[X_e > t_\ell] z_e \le \ell \qquad \forall \ell \in \mathsf{POS} \qquad (10.6) \\ & z \in \mathcal{Q}^{\mathsf{tree}}. \qquad (10.7) \end{aligned}$$

Claim 10.3 applied to all $\ell \in \mathsf{POS}$ gives the following.

Claim 10.9. *Consider a nonincreasing vector $\vec{t}$ satisfying $t_\ell \ge \tau_\ell(Y^*)$ for all $\ell \in \mathsf{POS}$. Then,* (Tree-LP($\vec{t}$)) *is feasible, and its optimal objective value* $\mathsf{Tree\text{-}OPT}(\vec{t})$ *is at most* $\sum_{e \in T^*} \mathbf{E}\left[\left(X_e - \tau_1(Y^*)\right)^+\right]$.

10.3.1 LP-Rounding Strategy

Let $\vec{t} \in \mathbb{R}_{\geq 0}^{\mathsf{POS}}$ be a nonincreasing vector such that (Tree-LP($\vec{t}$)) is feasible. Let $\bar{z}$ be an optimal fractional spanning tree solution to (Tree-LP($\vec{t}$)), and let Tree-OPT$(\vec{t})$ be the optimal value of (Tree-LP($\vec{t}$)). We remark that $\bar{z}$ can be computed in polynomial time by using a separation oracle for the spanning-tree polytope (more generally, matroid base-polytope); note that separating over constraints (10.6) is trivial.

The procedure for rounding $\bar{z}$ is more involved now, compared to that for a single Top_ℓ norm, since we now need to ensure that the τ_ℓ statistics of the cost-vector Y, of the induced tree, are comparable to the t_ℓ's for all $\ell \in \mathsf{POS}$. This involves rounding while roughly preserving constraints (10.6). A key property of the LP-constraint matrix that we exploit in our rounding scheme is that the column-sums can be bounded by $O(1)$ if we scale constraint (10.6) by ℓ; we perform the scaling because we can afford to incur an additive $O(\ell)$ violation in (10.6) for any index $\ell \in \mathsf{POS}$.

Define the parameter $\nu := \sum_{\ell \in \mathsf{POS}} \frac{1}{\ell} < 2$. Using Theorem 2.32, with the parameter ν, we can round $\bar{z}$ to an integral spanning tree T with the following guarantees.

Lemma 10.10. *The cost vector $Y = Y^T$ satisfies the following: (i) $\sum_{e \in T} \mathbf{E}[(Y_e - t_1)^+] \leq$ Tree-OPT$(\vec{t})$; and (ii) for any index $\ell \in \mathsf{POS}$, we have $\tau_{3\ell}(Y) \leq t_\ell$.*

Proof. We first show that the hypothesis of Theorem 2.32 holds with the violation parameter ν. For any edge $e \in E$ and index $\ell \in \mathsf{POS}$, we trivially have $\mathbf{Pr}[X_e > t_\ell] \in [0,1]$. Therefore, $\sum_{\ell \in \mathsf{POS}} \mathbf{Pr}[X_e > t_\ell]/\ell \leq \sum_{\ell \in \mathsf{POS}} 1/\ell = \nu$.

The first claim follows directly from Theorem 2.32(a). For the second claim, we use Theorem 2.32(b): we have that T satisfies $\sum_{e \in T} \mathbf{Pr}[X_e > t_\ell] \leq \ell(1+\nu) < 3\ell$ for every index $\ell \in \mathsf{POS}$. By definition, $\tau_{3\ell}(Y) \leq t_\ell$. ∎

The above result combined with Theorem 4.16 yields a bound on $\mathbf{E}[f(Y)]$ involving Tree-OPT$(\vec{t})$ and $\{t_\ell\}_{\ell \in \mathsf{POS}}$. For notational convenience, define $t_\ell := 0$ for indices $\ell > n-1$. We need the following notion of *expansion* of a vector in $\mathbb{R}_{\geq 0}^{\mathsf{POS}}$.

Definition 10.11. *For a nonincreasing vector $\vec{t}$, its expansion $\vec{t'} \in \mathbb{R}_{\geq 0}^{n-1}$ is given by: for any index $i \in [n-1]$, $t'_i := t_{2^{\lceil \log_2 i \rceil}}$.*

Theorem 10.12. *Let $\vec{t'} \in \mathbb{R}_{\geq 0}^{n-1}$ denote the expansion of $\vec{t} \in \mathbb{R}_{\geq 0}^{\mathsf{POS}}$. We have*

$$\mathbf{E}[f(Y)] \leq 4.026 \cdot \Big(3t_1 + \text{Tree-OPT}(\vec{t}) + 6f(\vec{t'})\Big).$$

Proof. We observe that the hypothesis of Theorem 4.16 holds if we take $W = \vec{t'}$, $\lambda = 1$, $B_1 = \textsf{Tree-OPT}(\vec{t})$, $\alpha = 3$ and $\beta = 1$; here, we use that $\tau_\ell(W) = \tau_\ell(\vec{t'}) = t'_\ell$ for all indices $\ell \in [m]$. Therefore, Theorem 4.16 gives:

$$\mathbf{E}[f(Y)] \leq 4.026 \cdot \{3t_1 + \textsf{Tree-OPT}(\vec{t}) + 6f(\vec{t'})\} \qquad \blacksquare$$

10.3.2 Our Algorithm and Analysis

We repeat the approximation strategy that we used for the $\textsf{Top}_\ell$ case (see Section 10.2.1). Suppose that we have a guess-vector $\vec{t}$ that is coordinate-wise close to $(\tau_\ell(Y^*))_{\ell \in \textsf{POS}}$. Then by Claim 10.9, we can say that $t_1 + \textsf{Tree-OPT}(\vec{t}) \leq \gamma_1(Y^*)$. Theorem 10.12 then shows that our cost is $O(\gamma_1(Y^*) + f(\vec{t'}))$, where $\vec{t'}$ is the expansion of $\vec{t}$ (see Definition 10.11). Now, recall that Theorem 4.4 gives $f(\gamma_1(Y^*), \tau_2(Y^*), \ldots, \tau_m(Y^*)) = O(\mathbf{E}[f(Y^*)])$, so we will be done if we can bound $\gamma_1(Y^*) + f(\vec{t'})$ in terms of $f(\gamma_1(Y^*), \tau_2(Y^*), \ldots, \tau_m(Y^*))$. This is not hard to do: since f is normalized, we have $\gamma_1(Y^*) \leq f(\gamma_1(Y^*), 0, \ldots, 0)$, and since f is monotone and $t'_1 \leq \gamma_1(Y^*)$, we can say that $f(\gamma_1(Y^*), 0, \ldots, 0) + f(\vec{t'}) \leq 2\, f(\gamma_1(Y^*), \tau_2(Y^*), \ldots, \tau_m(Y^*))$.

As in the $\textsf{Top}_\ell$ setting, we can find a polynomial-size set $\mathcal{T}$ containing a suitable guess vector $\vec{t}$ (see Claim 10.13). In the $\textsf{Top}_\ell$ setting, we used the expression $\ell t_\ell + \textsf{Tree-OPT}(\ell, t_\ell)$ as a proxy for the expected $\textsf{Top}_\ell$-norm of the tree obtained from an optimal LP solution; now Theorem 4.4 suggests that we use the expression $f(t'_1 + \textsf{Tree-OPT}(\vec{t}), t'_2, t'_3, \ldots, t'_{n-1})$, where $\vec{t'} = (t'_1, \ldots, t'_{n-1})$ is the expansion of $\vec{t}$, and we indeed use this to identify a suitable guess vector from $\mathcal{T}$ (see Lemma 10.14). Combining these ingredients yields our algorithm, which we now state and proceed to analyze.

Description of our algorithm. Define $\textsf{UB}$ to be the weight of a minimum-weight spanning tree in G with edge-weights given by (the deterministic quantity) $w_e := \mathbf{E}[X_e]$. Define

$$\mathcal{T} := \left\{ \vec{t} \in \mathbb{R}_{\geq 0}^{\textsf{POS}} : \vec{t} = \vec{t}^{\downarrow}, \forall\, \ell \in \textsf{POS},\ \frac{2\,\textsf{UB}}{n^2} \leq t_\ell \leq 4\,\textsf{UB},\ t_\ell \text{ is a power of } 2 \right\} \tag{10.8}$$

Let $\vec{t} \in \mathcal{T}$ be such that ($\textsf{Tree-LP}(\vec{t})$) is feasible and $f(t'_1 + \textsf{Tree-OPT}(\vec{t}), t'_2, t'_3, \ldots, t'_{n-1})$ is minimized, where $\vec{t'} = (t'_1, \ldots, t'_{n-1})$ is the expansion of $\vec{t}$ (see Definition 10.11). Let $\bar{z}$ be an optimal spanning tree solution to ($\textsf{Tree-LP}(\vec{t})$), and $\textsf{Tree-OPT}(\vec{t})$ denote the corresponding LP-objective value; again, note that ($\textsf{Tree-LP}(\vec{t})$) can be solved in polytime. Round $\bar{z}$ to obtain a spanning tree T as described in Section 10.3.1. Return T.

Analysis. Claim 10.13 shows that $\mathcal{T}$ is a polynomial-size set containing a suitable guess vector $\vec{t}$, and (the proof of) Lemma 10.14 shows that for the expansion $\vec{t'} = (t'_1, t'_2, \ldots, t'_{n-1})$ of $\vec{t}$, $f(t'_1 + \mathsf{Tree\text{-}OPT}(\vec{t}), t'_2, t'_3, \ldots, t'_{n-1})$ is $O(OPT)$. Combining this with Theorem 10.12 then yields Theorem 10.1.

Claim 10.13. *We have: (i) $|\mathcal{T}| = \text{poly}(n)$; and (ii) there exists a nonincreasing vector $\vec{t} \in \mathcal{T}$ satisfying the following inequality for all indices $\ell \in \mathsf{POS}$:*

$$\tau_\ell(Y^*) \le t_\ell \le \max\Big\{2\tau_\ell(Y^*), \frac{4\,\mathsf{UB}}{n^2}\Big\}.$$

Proof. Since there are $O(\log n)$ indices in POS and $O(\log n)$ unique powers of 2 in the range $[2\,\mathsf{UB}/n^2, 4\,\mathsf{UB}]$, Claim 2.20 gives:

$$|T| \le (2e)^{\max(O(\log_2 n), |\mathsf{POS}|)} = O(\text{poly}(n)).$$

For the second claim, consider the nonincreasing vector $\vec{t} \in \mathbb{R}_{\ge 0}^{\mathsf{POS}}$ where for each $\ell \in \mathsf{POS}$, t_ℓ is the smallest power of 2 that is at least $\max\{\tau_\ell(Y^*), 2\,\mathsf{UB}/n^2\}$. By definition, for every $\ell \in \mathsf{POS}$ we have $t_\ell \le 2 \cdot \max\{\tau_\ell(Y^*), 2\,\mathsf{UB}/n^2\}$, so it remains to show that $\vec{t} \in \mathcal{T}$, and this follows from Lemma 10.6. ■

Lemma 10.14. *Let $\vec{t} \in \mathbb{R}_{\ge 0}^{\mathsf{POS}}$ be such that for all indices $\ell \in \mathsf{POS}$, we have $\tau_\ell(Y^*) \le t_\ell \le \max\big\{2\tau_\ell(Y^*), 4\,\mathsf{UB}/n^2\big\}$. Let $\vec{t'} = (t'_1, \ldots, t'_{n-1})$ denote the expansion of $\vec{t}$, i.e., for any index $i \in [n-1]$, $t'_i := t_{2^{\lceil \log_2 i \rceil}}$. Then,*

$$f(t'_1 + \mathsf{Tree\text{-}OPT}(\vec{t}), t'_2, t'_3, \ldots, t'_{n-1}) \le 8\,\mathbf{E}[f(Y^*)] = 8\,\mathsf{OPT}.$$

Proof. We first establish some bounds on t'_i for an arbitrary index $i \in [n-1]$. Define $\ell_i := 2^{\lceil \log_2 i \rceil}$. Note that either $\ell_i > n-1$ or $\ell_i \in \mathsf{POS}$. We get:

$$t'_i = t_{\ell_i} \le 2\tau_{\ell_i}(Y^*) + \frac{4\,\mathsf{UB}}{n^2}.$$

Next, by Claim 10.9, we have $\mathsf{Tree\text{-}OPT}(\vec{t}) \le \sum_{e \in T^*} \mathbf{E}[(X_e - \tau_1(Y^*))^+]$. Observe that the non-increasing vector $(t'_1 + \mathsf{Tree\text{-}OPT}(\vec{t}), t'_2, t'_3, \ldots, t'_{n-1})$ is coordinate-wise dominated by the vector $2v + v'$, where

$$v := (\gamma_1(Y^*), \tau_2(Y^*), \tau_3(Y^*), \ldots, \tau_m(Y^*)) \text{ and } v' := (4\,\mathsf{UB}/n^2, \ldots, 4\,\mathsf{UB}/n^2).$$

Since f is a monotone norm, we obtain:

$$\begin{aligned} f(t'_1 + \mathsf{Tree\text{-}OPT}(\vec{t}), t'_2, t'_3, \ldots, t'_{n-1}) &\le 2f(v) + f(v') \\ &\le 2f(\gamma_1(Y^*), \tau_2(Y^*), \ldots, \tau_m(Y^*)) + 4\,\mathsf{UB}/n \\ &\qquad\qquad \text{(By Lemma 2.2)} \\ &\le 8\,\mathbf{E}[f(Y^*)] \quad \text{(By Theorem 4.4 and Lemma 10.5)} \end{aligned}$$

■

Proof of Theorem 10.1. Let $\vec{t} \in \mathcal{T}$, and T be the spanning tree that we define in Algorithm 10.3.2. Let $Y = Y^T$ denote the cost vector induced by T in the given $\mathsf{StochNormTree}$ instance. By Theorem 10.12, we have

$$\mathbf{E}[f(Y)] \le 4.026 \cdot \Big(3t_1 + \mathsf{Tree\text{-}OPT}(\vec{t}) + 6f(\vec{t'})\Big),$$

where $\vec{t'} \in \mathbb{R}^{n-1}_{\ge 0}$ denotes the expansion of $\vec{t}$. Since f is normalized, and also a monotone, symmetric norm, we get:

$$\mathbf{E}[f(Y)] \le 4.026 \cdot 9 \cdot f(t'_1 + \mathsf{Tree\text{-}OPT}(\vec{t}), t'_2, t'_3, \ldots, t'_{n-1}) \le 290 \cdot \mathsf{OPT}.$$

In the final inequality, we use Claim 10.13, Lemma 10.14 and the fact that $\vec{t}$ minimizes $f(t'_1 + \mathsf{Tree\text{-}OPT}(\vec{t}), t'_2, t'_3, \ldots, t'_{n-1})$ among all feasible guess vectors. ■

10.4 Extensions

We discuss three natural extensions of $\mathsf{StochNormTree}$.

10.4.1 Stochastic Min-Norm Matroid Basis

Our results extend quite seamlessly to the *stochastic minimum norm matroid basis* problem (denoted $\mathsf{StochNormMatBasis}$), which is the generalization of stochastic minimum-norm spanning tree, where we replace spanning trees by bases of an arbitrary matroid. More precisely, the setup is that we are given a matroid $\mathcal{M} = (\mathcal{U}, \mathcal{I})$ specified via its rank function r (equivalently, we have access to an independence oracle for the matroid), and a monotone, symmetric norm $f : \mathbb{R}^{\mathsf{r}(\mathcal{M})}_{\ge 0} \to \mathbb{R}_{\ge 0}$. Each element $e \in \mathcal{U}$ has a stochastic weight X_e. The goal is to find a basis of $\mathcal{M}$ whose induced random weight vector has the minimum expected f-norm.

Our algorithms and analyses from Sections 10.2 and 10.3 extend essentially as is here. The only changes in the algorithm are that we replace the spanning-tree polytope $\mathcal{Q}^{\mathsf{tree}}$ in (Tree-LP($\vec{t}$)) by the base polytope of $\mathcal{M}$, and (i) for Top_ℓ norms, we now of course find the min-weight basis of $\mathcal{M}$; and (ii) for an arbitrary monotone, symmetric norm, we invoke Theorem 2.32 with the base polytope of $\mathcal{M}$.

Theorem 10.15.
There is an $O(1)$-approximation algorithm for stochastic minimum norm matroid basis (i.e., with arbitrary edge-weight distributions and an arbitrary monotone, symmetric norm).
Furthermore, for StochNormMatBasis *with a* Top_ℓ *norm, we obtain approximation guarantees of $(2+\varepsilon)$ for general ℓ, and $\big(e/(e-1)+\varepsilon\big)$ for $\ell = 1$, where $\varepsilon > 0$ is a constant.*

10.4.2 Stochastic Min-Norm Degree-Bounded Spanning Tree

In the *stochastic minimum norm degree-bounded spanning tree* problem, we are given an instance of StochNormTree along with upper bounds $\{d_v\}_{v\in V}$ on the vertex degrees. Feasible solutions to this problem are given by spanning trees that respect the degree constraints, i.e., T must satisfy $|\delta(v) \cap T| \le d(v)$ for all vertices $v \in V$, where $\delta(v)$ denotes the set of edges in G that are incident to v. The goal is to find a degree-bounded spanning tree T that minimizes $\mathbf{E}[f(Y^T)]$. We call this problem StochNormDegBndTree.

The classical setting of (deterministic) minimum-weight degree-bounded spanning tree problem, wherein each edge e has a fixed weight $w_e \in \mathbb{R}_{\ge 0}$ and f is $\mathsf{Top}_{|V|-1}$-norm, is already NP-hard because it generalizes the path version of the *traveling salesperson problem* (TSP). There is a tight approximation result for this setting: if the instance is feasible, we can find a tree of total weight at most the optimum that violates the degree bounds by at most an additive 1 [41].

Our approximation strategy for StochNormTree can be suitably modified to give a bi-criteria approximation algorithm for StochNormDegBndTree where we violate the degree bound at each vertex by at most an additive $O(1)$ factor; a violation in the degree constraint is unavoidable even in the deterministic setting with $f = \mathsf{Top}_{|V|-1}$-norm. Concretely, we append (Tree-LP($\vec{t}$)) by including degree-constraints of the form $\sum_{e\in\delta(v)} z_e \le d(v)$ for all vertices v. Note that the maximum column sum increases by at most 2 because each edge-variable participates in at most two degree constraints. The main change in the analysis occurs at Lemma 10.10: the rounded (approximately degree-bounded) spanning tree satisfies $\tau_{5\ell}(Y) \leq t_\ell$ for all indices $\ell \in \mathsf{POS}$. Rest of the proofs carryover with minor changes in the constants.

Our main result on StochNormDegBndTree is the following.

Theorem 10.16.
In polynomial time, we can detect that the StochNormDegBndTree *instance is infeasible, or obtain a spanning tree T satisfying:*

(i) $\mathbf{E}[f(Y^T)] = O(\mathbf{E}[f(Y^*)])$, *where Y^* is an optimal degree-bounded spanning tree.*

(ii) *For any vertex $v \in V$, the violation in the degree constraint at v is at most 4, i.e., $|\delta(v) \cap T| \leq d(v) + 4$.*

10.4.3 Stochastic Min-Norm Traveling Salesperson Problem

In this section, we describe a stochastic min-norm optimization problem that arises from TSP. In an instance of the *stochastic minimum norm traveling salesperson problem* (denoted StochNormTSP), we are given an undirected graph $G = (V, E)$, stochastic edge-weights $\{X_e\}_{e \in E}$, and a monotone, symmetric norm $f : \mathbb{R}_{\geq 0}^{|E|} \to \mathbb{R}_{\geq 0}$; note that, unlike in StochNormTree, the norm f is $|E|$-dimensional. A multiedge-set K (consisting of edges in E) is said to be *feasible* if the multigraph $H = (V, K)$ is connected and *Eulerian* (i.e., the degree of each vertex in H is even). We use $\chi^K = (\chi_e^K)_{e \in E} \in \mathbb{Z}_{\geq 0}^E$ to denote the characteristic vector of K, where χ_e^K is the number of copies of edge e that appear in K. Each solution K induces an $|E|$-dimensional cost vector Y^K, where $Y_e^K := \chi_e^K \cdot X_e$. The objective in StochNormTSP is to find a feasible solution K that minimizes $\mathbf{E}[f(Y^K)]$. Let OPT denote the objective value of an optimal solution.

Consider the classical setting of TSP, where each edge e has a fixed weight w_e, and f is the $\mathsf{Top}_{|E|}$-norm. There is a simple 2-approximation algorithm for this problem via a connection to the min-weight spanning tree problem: Compute a minimum w-weight spanning tree T in G, and take $K := T \sqcup T$ to be two disjoint copies of T. (Any TSP solution is spanning, so $\mathsf{OPT} \geq w(T) = \sum_{e \in T} w_e$, and $w(K) = 2w(T) \leq 2\mathsf{OPT}$.) Since the dimension of f in StochNormTree and StochNormTSP instances are different, extending the above approximation strategy to solve StochNormTSP requires a minor technical change. Given some $f : \mathbb{R}_{\geq 0}^{|E|} \to \mathbb{R}_{\geq 0}$ in a StochNormTSP instance, we use the same norm while solving the StochNormTree instance that arises in the reduction. Since the Y^T-vector that arises in StochNormTree is $(|V| - 1)$-dimensional, we pad it with sufficiently many 0's before evaluating the f-norm. Now consider deterministic min-norm TSP. Recall from Section 10.1 that a min-weight spanning tree simultaneously minimizes *all* Top_ℓ norms (of the induced weight vector), so using majorization inequality (Theorem 2.7) it is not hard to see that the above 2-approximation strategy extends verbatim to the setting with arbitrary monotone, symmetric norms. We can further extend this strategy to the stochastic setting to obtain the following result.

Theorem 10.17.
Suppose that we are given black-box access to an α-approximation algorithm $\mathcal{A}$ for StochNormTree*. Then, there is a 2α-approximation algorithm for* StochNormTSP *that uses a single application of $\mathcal{A}$.*

Chapter 11

Conclusions and Future Work

In this thesis, we introduce the model of stochastic minimum-norm optimization, and present a framework for designing approximation algorithms for problems in this model. A key component of our framework is a structural result showing that if f is a monotone, symmetric norm, and Y follows a product distribution on $\mathbb{R}^m_{\geq 0}$, then $\mathbf{E}[f(Y)] = \Theta\big(f(\mathbf{E}[Y^{\downarrow}])\big)$; in particular, this shows that $\mathbf{E}[f(Y)]$ can be controlled by controlling $\mathbf{E}[\mathsf{Top}_\ell(Y)]$ for all $\ell \in [m]$ (or all $\ell \in \{1, 2, 4, \ldots, 2^{\lfloor \log_2 m \rfloor}\}$). Enroute to proving this result, we develop various deterministic proxies to reason about expected Top_ℓ-norms, which also yield a deterministic proxy for $\mathbf{E}[f(Y)]$. We utilize our framework to develop approximation algorithms for stochastic min-norm load balancing (StochNormLB) and stochastic min-norm spanning tree (StochNormTree). We obtain $O(1)$-approximation algorithms for StochNormTree, and StochNormLB with (i) an arbitrary monotone, symmetric norm and Bernoulli job sizes (BerNormLB), and (ii) Top_ℓ norms and arbitrary job-size distributions ($\mathsf{StochTop}_\ell\mathsf{LB}$).

We also give strong approximation guarantees for StochNormLB with Poisson jobs via a reduction to (deterministic) min-norm load balancing, where we only lose a $(1+\varepsilon)$-factor in approximation. Here, our key observation is that the expected f-norm of a Poisson product distribution (i.e., the function $g(y)$ from Definition 9.1) is Schur convex.

The most pressing question left open by our work is developing a constant-factor approximation algorithm for the general case of StochNormLB, where both the monotone symmetric norm and the job-size distributions are arbitrary; currently, we only have an $O(\log\log m/ \log\log\log m)$-approximation via a "reduction" to the $O(\log m)$-dimensional vector scheduling problem. We remark that unless we use any additional properties of the vector-scheduling instance that we obtain in this reduction, the current approach is unlikely to yield any meaningful improvements. Recall that, under some complexity theoretic assumptions, [38] shows that d-dimensional vector scheduling (even on identical machines) is hard to approximate to a factor $O((\log d)^{1-\varepsilon})$. Another satisfactory outcome would be to rule out $O(1)$-approximation for $\mathsf{StochNormLB}$ by leveraging hardness results for vector scheduling.

In this work, we have sought to optimize approximation ratios of our algorithms while keeping the exposition clean. Still, the $O(1)$-approximation guarantees that we obtain for $\mathsf{StochTop}_\ell\mathsf{LB}$, $\mathsf{BerNormLB}$ and the general version of $\mathsf{StochNormTree}$ are in the hundreds; this is true even of prior work on minimizing expected makespan [23, 10]. We identify two possible places where significant improvements might be possible. First, in Theorem 4.1, we show that $\mathbf{E}[f(Y)] \leq 7.634 \cdot f(\mathbf{E}[Y^{\downarrow}])$. We do not know if a much smaller constant upper bound—say, even 2—is possible; recall that in Remark 4.1 we gave an example showing that $\mathbf{E}[f(Y)]$ can be larger than $1.214 \cdot f(\mathbf{E}[Y^{\downarrow}])$. Of course, any improvement in this upper bound would directly translate to improved approximation factors for all stochastic min-norm optimization results that are based on our framework. Second, in Lemma 5.10, we show that for any composite random variable $S = \sum_j Z_j$ that is a sum of independent $[0, 1]$-bounded random variables, and an effective-size parameter $\lambda \geq 1$, the following volume inequality holds:

$$\mathbf{E}[S^{\geq 1}] \geq \Big(\sum_j \beta_\lambda(Z_j/4) - 6\Big)/4\lambda.$$

We conjecture that the above inequality can be tightened to:

$$\mathbf{E}[S^{\geq 1}] \geq \Big(\sum_j \beta_\lambda(Z_j) - 1\Big)/\lambda.$$

This would roughly lead to a factor 10 improvement in the approximation ratios for $\mathsf{StochTop}_\ell\mathsf{LB}$ and $\mathsf{BerNormLB}$. Observe that both Theorem 4.1 and Lemma 5.10 are fundamental mathematical questions bereft of any computational concerns.

Other Stochastic Min-Norm Optimization Problems

We describe two other stochastic min-norm optimization problems arising from Clustering and Bipartite Perfect Matching applications. In both problems, we want to find a solution that minimizes the expected f-norm of the induced cost vector for a given monotone, symmetric norm f.

- In stochastic k-clustering, we have a point-set $\mathcal{X}$, a distance function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$, and stochastic demands $\{u_p\}_{p \in \mathcal{X}}$. Any clustering of points in $\mathcal{X}$ around k cluster centers $q_1, \ldots, q_k$ induces a $|\mathcal{X}|$-dimensional connection-cost vector, where the connection cost of point $p \in \mathcal{X}$ is $u_p \cdot \min_{i \in [k]} d(p, q_i)$.

- In stochastic bipartite perfect matching, we have a bipartite graph $G = (V, E)$ whose edges have stochastic weights. The random cost vector induced by a perfect matching M consists of edge-weight variables that participate in the matching.

  We remark that even the deterministic case of this problem is open.

Optimization Under Correlated Uncertainty

A crucial assumption in our stochastic min-norm optimization model is that the induced cost vectors follow product distributions. Thus, a natural direction for future work is to drop this assumption. To the best of our knowledge, prior work on this subject has focused on determining the *price of correlation*, which measures the loss in approximation quality incurred by ignoring correlations. We describe two concrete problems in this setting.

- **Stochastic Load Balancing with Correlated Jobs:** What is the best approximation guarantee that we can obtain for stochastic makespan minimization on identical machines when the job-size variables $\{X_j\}_{j\in J}$ are allowed to have arbitrary correlations? We remark that, assuming P $\neq$ NP, an $O(1)$-approximation is already ruled out. This is because d-dimensional vector scheduling can be cast as a special case of this problem. Next, we give a lower bound example showing that the price of correlation for this problem is a super-constant. Consider an instance with m^2 jobs and m (identical) machines. Suppose that the job-set can be partitioned into m independent groups, where each group has m negatively-associated Bernoulli jobs that take size 1 with probability $1/m$, and size 0 otherwise. We further assume that the sum of all variables within each group is always 1. Processing each group of jobs on a unique machine gives an assignment whose expected makespan is 1. If all jobs were independent, then it is not hard to argue that any assignment has expected makespan $\Omega(\log m/\log\log m)$.

- **Stochastic Single-Sink Unsplittable Flow:** In an instance of this problem, we have an undirected capacitated graph $G = (V, E, u : E \to \mathbb{R}_{>0})$, k source vertices $s_1, \ldots, s_k$ with (independent) stochastic demands $D_1, \ldots, D_k$, respectively, and a single sink vertex t. For each $i \in [k]$, we want to route an unsplittable (s_i, t)-flow of value D_i along a fixed (s_i, t)-path P_i. Any feasible choice of $P_1, \ldots, P_k$ induces a (random) congestion vector Y on the edges of G, where $Y_e := \left(\sum_{i\in[k]:P_i\ni e} D_i\right)/u_e$. The goal is to find a solution that minimizes the expected maximum congestion $\mathbf{E}[\mathsf{Top}_1(Y)]$ (or more generally, $\mathbf{E}[f(Y)]$ for a monotone symmetric norm f). Observe that, unlike stochastic load balancing with correlated jobs, the demands in this problem are independent random variables, but the correlations in the induced congestion vector are due to edges in a flow-path P_i having positive correlations. We also remark that the restricted assignment case (of load balancing), where each job can only be processed on a specified subset of machines, can be cast as an instance of the min-congestion single-sink unsplittable flow problem, so the stochastic version of the latter problem is more complex than what we have studied so far.

Price of Uncertainty in Stochastic Min-Norm Optimization

In our stochastic min-norm optimization model, we require that the solution be computed by only knowing distributional information about the underlying costs. This obliviousness restriction can sometimes lead to poor-quality solutions, so we consider relaxations of this notion. The goal here is to understand the trade-offs between making a decision before and after seeing a random variable's realization.

- **Probing models:** Suppose that for each variable j with cost random variable X_j, we are given a price $p_j \in \mathbb{R}_{\geq 0}$ to force a realization of X_j (according to its distribution function). Given a budget B for probing, how much can we save in the objective function by using (adaptive or non-adaptive) algorithms for stochastic min-norm optimization? As B varies from 0 to ∞, we interpolate between fully oblivious and fully aware models of stochastic min-norm optimization.

  Another line of research comes from weaker notions of probing. A standard assumption in probing models is that we learn the actual realization of a random variable by paying the price for probing. Many optimization decisions often only require partial information about the randomness: for example, whether or not a random variable is less (or greater) than θ for some suitable θ. Can we work with weaker notions of probing where we pay a smaller price to only partially collapse the random variable onto its support?

- **Two-stage models:** Consider the standard model of two-stage stochastic optimization where we have distributional information in the first stage and actual realizations in the second stage. Suppose that we are required to partially fix a "super-solution" in the first stage and fully fix a solution (consistent with the first-stage super-solution) in the second stage.

  We describe two flavors of problems in this model. In the two-stage stochastic load balancing problem, we are given an additional integer parameter k. In the first stage, for each job j we must fix a subset M_j of k machines. After job-sizes are revealed in the second stage, we must assign each job j to a machine in M_j. As k varies from 1 through m, we interpolate between fully oblivious and fully aware models of stochastic min-norm load balancing.

In the two-stage stochastic spanning tree problem we are given nonnegative prices $\{p_e\}_{e\in E}$ on the edges. In the first stage, we are allowed to buy any subgraph (V, F) of G by paying a price of $p(F)$. In the second stage, edge-weights are realized and we must pick a spanning tree from the subgraph (V, F). The overall goal in this problem is to minimize the sum of first-stage price $p(F)$ and the expected norm-objective in the second stage.

References

[1] Soroush Alamdari and David B. Shmoys. A Bicriteria Approximation Algorithm for the k-Center and k-Median Problems. In *Proceedings of the 15th International Workshop on Approximation and Online Algorithms*, pages 66–75, 2017.

[2] Ali Aouad and Danny Segev. The Ordered k-Median Problem: Surrogate Models and Approximation Algorithms. *Mathematical Programming*, 177:55–83, 2019.

[3] Jaroslaw Byrka, Krzysztof Sornat, and Joachim Spoerhase. Constant-Factor Approximation for Ordered k-Median. In *Proceedings of the 50th Symposium on Theory of Computing*, pages 620–631, 2018.

[4] Deeparnab Chakrabarty and Chaitanya Swamy. Interpolating between k-Median and k-Center: Approximation Algorithms for Ordered k-Median. In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming*, pages 29:1–29:14, 2018.

[5] Deeparnab Chakrabarty and Chaitanya Swamy. Approximation Algorithms for Minimum Norm and Ordered Optimization Problems. In *Proceedings of the 51st Symposium on Theory of Computing*, pages 126–137, 2019.

[6] Deeparnab Chakrabarty and Chaitanya Swamy. Simpler and Better Algorithms for Minimum-Norm Load Balancing. In *Proceedings of the 27th European Symposium on Algorithms*, pages 27:1–27:12, 2019.

[7] Chandra Chekuri and Sanjeev Khanna. On Multidimensional Packing Problems. *SIAM Journal on Computing*, 33(4):837–851, 2004.

[8] Anindya De, Sanjeev Khanna, Huan Li, and Hesam Nikpey. An Efficient PTAS for Stochastic Load Balancing with Poisson Jobs. In *Proceedings of the 47th International Colloquium on Automata, Languages, and Programming*, pages 37:1–37:18, 2020.

[9] Ashish Goel and Piotr Indyk. Stochastic Load Balancing and Related Problems. In *Proceedings of the 40th Foundations of Computer Science*, pages 579–586, 1999.

[10] Anupam Gupta, Amit Kumar, Viswanath Nagarajan, and Xiangkun Shen. Stochastic Load Balancing on Unrelated Machines. In *Proceedings of the 29th Symposium on Discrete Algorithms*, pages 1274–1285, 2018.

[11] Anupam Gupta, Amit Kumar, Viswanath Nagarajan, and Xiangkun Shen. Stochastic Load Balancing on Unrelated Machines. *Mathematics of Operations Research*, 46(1):115–133, 2021.

[12] Varun Gupta, Benjamin Moseley, Marc Uetz, and Qiaomin Xie. Greed Works–Online Algorithms for Unrelated Machine Stochastic Scheduling. *Mathematics of Operations Research*, 45(2):497–516, 2020.

[13] Godfrey H. Hardy, John E. Littlewood, and George Pólya. *Inequalities*. Cambridge University Press, 1934.

[14] David G. Harris and Aravind Srinivasan. The Moser–Tardos Framework with Partial Resampling. *Journal of the ACM*, 66(5):1–45, 2019.

[15] Wassily Hoeffding. On the Distribution of the Number of Successes in Independent Trials. *The Annals of Mathematical Statistics*, 27(3):713–721, 1956.

[16] Joseph Y. Hui. Resource Allocation for Broadband Networks. *IEEE Journal on Selected Areas in Communications*, 6(9):1598–1608, 1988.

[17] Sharat Ibrahimpur and Chaitanya Swamy. Approximation Algorithms for Stochastic Minimum-Norm Combinatorial Optimization. In *Proceedings of the 61st Foundations of Computer Science*, pages 966–977, 2020.

[18] Sharat Ibrahimpur and Chaitanya Swamy. Approximation Algorithms for Stochastic Minimum Norm Combinatorial Optimization. *CoRR*, abs/2010.05127, 2020.

[19] Sharat Ibrahimpur and Chaitanya Swamy. Minimum-Norm Load Balancing Is (Almost) as Easy as Minimizing Makespan. In *Proceedings of the 48th International Colloquium on Automata, Languages, and Programming*, pages 81:1–81:20, 2021.

[20] Sharat Ibrahimpur and Chaitanya Swamy. A Simple Approximation Algorithm for Vector Scheduling and Applications to Stochastic Min-Norm Load Balancing. In *Proceedings of the 5th Symposium on Simplicity in Algorithms*, pages 247–256, 2022.

[21] Sungjin Im, Nathaniel Kell, Janardhan Kulkarni, and Debmalya Panigrahi. Tight Bounds for Online Vector Scheduling. *SIAM Journal on Computing*, 48(1):93–121, 2019.

[22] Sungjin Im, Benjamin Moseley, and Kirk Pruhs. Stochastic Scheduling of Heavy-tailed Jobs. In *Proceedings of the 32nd International Symposium on Theoretical Aspects of Computer Science*, volume 30, pages 474–486, 2015.

[23] Jon M. Kleinberg, Yuval Rabani, and Éva Tardos. Allocating Bandwidth for Bursty Connections. *SIAM Journal on Computing*, 30(1):191–217, 2000.

[24] Gilbert Laporte, Stefan Nickel, and Francisco Saldanha da Gama. *Location Science*. Springer, 2019.

[25] Rafał Latała. Estimation of Moments of Sums of Independent Real Random Variables. *The Annals of Probability*, 25(3):1502–1513, 1997.

[26] Jian Li and Amol Deshpande. Maximizing Expected Utility for Stochastic Combinatorial Optimization Problems. In *Proceedings of the 52nd Symposium on Foundations of Computer Science*, pages 797–806, 2011.

[27] Jian Li and Amol Deshpande. Maximizing Expected Utility for Stochastic Combinatorial Optimization Problems. *Mathematics of Operations Research*, 44(1):354–375, 2019.

[28] Jian Li and Yu Liu. Approximation Algorithms for Stochastic Combinatorial Optimization Problems. *Journal of the Operations Research Society of China*, 4(1):1–47, 2016.

[29] Jian Li and Wen Yuan. Stochastic Combinatorial Optimization via Poisson Approximation. In *Proceedings of the 45th Symposium on Theory of Computing*, pages 971–980, 2013.

[30] André Linhares, Neil Olver, Chaitanya Swamy, and Rico Zenklusen. Approximate Multi-Matroid Intersection via Iterative Refinement. *Mathematical Programming*, 183(1):397–418, 2020.

[31] Albert W Marshall, Ingram Olkin, and Barry Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer Series in Statistics. 2011.

[32] Adam Meyerson, Alan Roytman, and Brian Tagiku. Online Multidimensional Load Balancing. In *Proceedings of the 16th APPROX*, pages 287–302, 2013.

[33] Rolf H Möhring, Andreas S Schulz, and Marc Uetz. Approximation in Stochastic Scheduling: The Power of LP-Based Priority Policies. *Journal of the ACM*, 46(6):924–942, 1999.

[34] Marco Molinaro. Stochastic ℓ_p Load Balancing and Moment Problems via the L-function Method. In *Proceedings of the 30th Symposium on Discrete Algorithms*, pages 343–354, 2019.

[35] Stefan Nickel and Justo Puerto. *Location Theory: A Unified Approach*. Springer Berlin Heidelberg, 2005.

[36] Michael Pinedo. Offline Deterministic Scheduling, Stochastic Scheduling, and Online Deterministic Scheduling. In *Handbook of Scheduling*. Chapman and Hall/CRC, 2004.

[37] Yossef Rinott. Multivariate Majorization and Rearrangement Inequalities with Some Applications to Probability and Statistics. *Israel Journal of Mathematics*, 15(1):60–77, 1973.

[38] Sai Sandeep. Almost Optimal Inapproximability of Multidimensional Packing Problems. In *Proceedings of the 62nd Foundations of Computer Science*, pages 245–256, 2021.

[39] Sai Sandeep. Almost Optimal Inapproximability of Multidimensional Packing Problems. *CoRR*, abs/2101.02854, 2021.

[40] David B. Shmoys and Éva Tardos. An Approximation Algorithm for the Generalized Assignment Problem. *Mathematical Programming*, 62(1-3):461–474, 1993.

[41] Mohit Singh and Lap Chi Lau. Approximating Minimum Bounded Degree Spanning Trees to within One of Optimal. In *Proceedings of the 39th Symposium on Theory of Computing*, pages 661–670, 2007.