Convex Optimization Over Integer Points

Haotian Jiang

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:
Yin Tat Lee, Chair
Thomas Rothvoss
Simon S. Du

Program Authorized to Offer Degree:
Computer Science & Engineering

University of Washington

Abstract

Convex Optimization Over Integer Points

Haotian Jiang

Chair of the Supervisory Committee:
Yin Tat Lee
Computer Science & Engineering

Many problems in discrete optimization can be succinctly encapsulated as the question of minimizing a convex function f, which captures the combinatorial structures of the problem, over integer points that are typically $\{0,1\}^n$, i.e. $\min_{x\in\{0,1\}^n} f(x)$. At a high level, this thesis focuses on the efficient solvability and approximability of this optimization problem, with the aim of uncovering general principles towards understanding and utilizing the interplay between continuous and discrete optimization for the future development of algorithm design.

In seminal work, Grötschel, Lovász, and Schrijver (Combinatorica'81, Prog. Comb. Optim.'84, Springer'88) identified a central condition for the efficient solvability of this optimization problem: that the minimizer of f lies in $\{0,1\}^n$ itself. Under this condition, Grötschel, Lovász, and Schrijver designed a unified framework using the ellipsoid method to establish the polynomial solvability of a broad range of combinatorial optimization problems.

When the integer minimizer condition fails, the problem typically becomes computationally intractable, as witnessed by the NP-Hardness of the Integer Linear Programming problem. In this case, one has to resort to solving a convex relaxation, which is typically the linear relaxation $\min_{x\in[0,1]^n} f(x)$, and then rounding the fractional solution to an integer one, where the rounding error is classically known to be related to the discrepancy of the system (Lovász, Spencer, and Vesztergombi, Eur. J. Comb.'86). Over the past decade, following a breakthrough of Bansal (FOCS'10), there has been a burst of progress in developing efficient

algorithms for fundamental discrepancy results which were once thought to be computationally intractable. Consequently, this opens up the opportunity to develop a unified and systematic framework for rounding, and more broadly for algorithm design, through the lens of discrepancy theory.

In this thesis, we make substantial progress in both of the directions discussed above. The contributions of this thesis are summarized below:

- Under the integer minimizer condition, we give a faster and unified algorithm for solving the problem $\min_{x\in\{0,1\}^n} f(x)$ based on a reduction to the Shortest Vector Problem in lattice theory, improving upon the classical work by Grötschel, Lovász, and Schrijver from the 1980s in its full generality. Consequently, we obtain the first sub-cubic strongly polynomial oracle complexity algorithms for Submodular Function Minimization in its 50 years of study. We complement our algorithms by proving stronger hardness results for Submodular Function Minimization.

- We advance the frontier for central problems in discrepancy theory and obtain new applications of them to algorithm design. In particular, we prove the Matrix Spencer Conjecture, a generalization of the seminal result of Spencer (Trans. Am. Math. Soc.'85) in discrepancy theory, up to poly-logarithmic rank; we also obtain the first poly-logarithmic discrepancy algorithms for Online Discrepancy Minimization, complementing Spencer's classical result over 40 years ago which states that random coloring cannot be improved against adaptive adversaries (Spencer, J. Comb. Theory Ser. B'77). We further demonstrate applications of these results to the problems of Quantum Random Access Codes and Online Fair Allocation.

ACKNOWLEDGMENTS

As is often said, every Ph.D. process is a journey. Diving deep into research, this journey can sometimes become lonely and even stressful. Over the course of my Ph.D., I was extremely fortunate to have received the guidance and support of many around me, who have made my Ph.D. journey at the University of Washington a much more fruitful and enjoyable experience than it might otherwise have been.

First and foremost, I would like to thank my advisor, Yin Tat Lee, who introduced me to the fantastic world of convex optimization. Yin Tat was very generous with his time and energy, offering me lots of research ideas and inspiration. He encouraged me to work independently from him and introduced me to new research collaborations and opportunities from time to time. He also provided me with sufficient funding to travel around and work with other fantastic researchers in the area. It is fair to say that my achievements today would not have been possible without the time and effort Yin Tat has devoted to me.

I would also like to thank the other members of my Ph.D. supervisory committee, including Thomas Rothvoss, Simon Du, and Rekha Thomas. Thomas was also on my Qualification Examination committee back in 2019, and he gave me countless wonderful feedback and suggestions on research throughout my Ph.D. years. Simon and Rekha devoted a lot of time and effort to supervising my progress since my General Examination in 2020.

I would like to express my gratitude to all members of the Theory Group in the Paul G. Allen School of Computer Science & Engineering at the University of Washington. There were a lot of unforgettable moments and wonderful interactions with the theory group over the past four years. I would also like to thank the Paul G. Allen School of Computer Science & Engineering and the University of Washington for offering such a wonderful Ph.D. program

and for all the support they have offered me throughout the program.

Next, I would like to thank my other academic mentors, Janardhan Kulkarni, Nikhil Bansal, and Daniel Dadush, who greatly helped me develop my taste and passion for science over the past few years. Jana introduced me to the beautiful area of discrepancy theory and mentored my summer internship with Microsoft Research in 2019 and 2021. From him, I learned the importance to not only prove good results but also prove them in a mathematically beautiful and elegant way. This is the case for most of the results presented in this thesis. Nikhil and I have been collaborating on discrepancy theory, one of the most successful research directions during my Ph.D., for almost three years now. He hosted me for two visits to the University of Michigan in 2022 where he offered me invaluable research suggestions and important tips for giving talks. Daniel gave me some of the best comments and suggestions while I was working on the paper "Minimizing Convex Functions with Integral Minimizers". It was from these comments that I gradually deepened my understanding of the topic and even some of my own proofs. Daniel also got me interested in and worked on the Matrix Spencer Conjecture, which eventually led to my fruitful line of research on this topic presented in Chapters 4 - 6 of this thesis.

I would also like to thank all my wonderful academic friends and research collaborators during my Ph.D., including Nikhil Bansal, Sébastien Bubeck, Deeparnab Chakrabarty, Daniel Dadush, Sally Dong, Sivakanth Gopi, Andrei Graur, Anupam Gupta, Venkatesan Guruswami, Arun Jambulapati, Anna Karlin, Tarun Kathuria, Janardhan Kulkarni, Yin Tat Lee, Jian Li, Daogao Liu, Yang P. Liu, Raghu Meka, Swati Padmanabhan, Victor Reis, Thomas Rothvoss, Mehtaab Sawhney, Ziv Scully, Ruoqi Shen, Aaron Sidford, Sahil Singla, Makrand Sinha, Zhao Song, Kevin Tian, Santosh S. Vempala, Sam Chiu-wai Wong, Guanghao Ye, Lichen Zhang, and Xinzhi Zhang.

Last but not least, I would like to thank my parents, Shengfeng Jiang and Ying Wang, and my wife, Han Zhang. It would not have been possible for me to come this far if it were

not for the love and support they have given me over the years.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

Chapter 1

INTRODUCTION

Optimization problems arise widely and naturally in business, artificial intelligence, engineering, and applied sciences, and they have been extensively studied in computer science, machine learning, operations research, economics, and mathematics. Over the past few decades, the most prominent and successful approach for discrete optimization is to first solve a tractable convex relaxation using continuous optimization tools and then to round the fractional solution to an integral one. This generic "relax-and-round" approach, under the broader scope of the interplay between continuous and discrete optimization, has achieved enormous and unprecedented success, leading to better and sometimes even theoretically optimal algorithms for many fundamental discrete optimization problems.

However, despite its tremendous success, a systematic framework for implementing this approach, i.e., a theory that brings new techniques and understandings together in a unified and principled manner to be conveniently and effectively applicable to classical and new optimization problems, remains largely undeveloped. The main goal of this thesis is therefore to develop systematic tools and general principles to contribute towards building this unified framework for discrete optimization.

For most problems in discrete optimization, the relax-and-round approach can be succinctly formulated as the problem of minimizing a convex function f over integer points which are typically $\{0,1\}^n$. This leads to the central problem studied in this thesis:

$$\min_{x \in \{0,1\}^n} f(x). \tag{1.1}$$

Here, the convex function f captures the combinatorial structures of the problem at hand. In

its full generality, problem (1.1) generalizes the Integer Linear Programming (ILP) problem which is known to be computationally intractable in general. This hardness result is, in particular, responsible for the NP-Hardness of many combinatorial optimization problems.

In spite of this computational bottleneck in general, clear exceptions have presented themselves for problems such as maximum matching, minimum spanning tree, and submodular function minimization, where clever efficient algorithms are known. This phenomenon leads naturally to the quest for a characterization of the efficient solvability of problem (1.1):

For which convex functions f is problem (1.1) efficiently solvable?

Convex Functions with Integer Minimizers. An outstanding answer to the above question goes back to the seminal work of Grötschel, Lovász, and Schrijver in the 1980s [GLS81, GLS84, GLS88], where they identified a fundamental condition for which problem (1.1) becomes computationally tractable:

$$\text{The minimizer of the convex function } f \text{ lies inside } \{0,1\}^n. \tag{1.2}$$

Under this condition, they showed that the ellipsoid method [Sho77, YN76], which is used to give the first polynomial time algorithm for solving Linear Programming (LP) [Kha80], can be applied to solve problem (1.1) exactly in polynomial time. Not only so, they showed that the problem can be solved in *strongly polynomial time*, with runtime depending only on the dimension n but not on the "size" of the function f. Their approach further generalizes to the setting of *rational polyhedra*, where the set of minimizers of the convex function is a polytope whose vertices are all rational vectors with bounded bit complexity. As a result, Grötschel, Lovász and Schrijver were able to prove the weakly and strongly polynomial solvability of a wide range of combinatorial optimization problems using a unified framework.

Despite the tremendous success of their general framework, the original techniques used by Grötschel, Lovász, and Schrijver were not efficient enough compared with problem-specific algorithms, where problem structures are heavily exploited, often in an ad-hoc way, to achieve

fast runtime guarantees. The lack of general principles and techniques for obtaining algorithms with fast runtimes motivates the first central question of this thesis:

> How fast can one solve problem (1.1) when it is efficiently solvable? And what is the computational limit for solving it?

Rounding. When condition (1.2) fails, problem (1.1) typically becomes computationally intractable, which unfortunately is the case for many problems in combinatorial optimization. For these problems, the main approach over the past fourty years has been to solve the relaxation of problem (1.1) with the integer constraints $x \in \{0,1\}^n$ replaced by linear constraints $x \in [0,1]^n$, for which the problem becomes efficiently solvable, and then to *round* the fractional solution to an integer one with provable guarantees on the rounding error.

Despite the prevalence and enormous success of this approach, the design of rounding algorithms over the past many decades has mostly been problem-specific. The lack of a unified way to design and analyze rounding algorithms motivates the following second central question of this thesis:

> What is the optimal rounding error for problem (1.1) when it is not computationally tractable?

In subsequent chapters of this thesis, we develop general principles and techniques for providing better answers to the two central questions above. We also discuss our general approaches in the context of some of the most classical questions within the subject, and illustrate how they imply better algorithms for these specific problems. More broadly, the goal of this thesis is to provide new insights into the efficient solvability and approximability of problem (1.1), with the aim of uncovering general principles towards understanding and utilizing the interplay between continuous and discrete optimization for the future development of algorithm design.

1.1 Convex Functions with Integer Minimizers

As mentioned earlier, in their pioneer work, Grötschel, Lovász, and Schrijver [GLS81, GLS84, GLS88] showed that under the integer minimizer condition (i.e., condition (1.2)), problem (1.1) can be solved efficiently in weakly and strongly polynomial time using the ellipsoid method. Their work not only successfully explained the polynomial time solvability of a wide range of discrete optimization problems for which efficient algorithms were known but also gave the first polynomial time algorithms for fundamental problems including submodular function minimization (SFM).

Submodular Function Minimization. Submodular function minimization has been recognized as an important problem in the field of combinatorial optimization. Submodular functions are ubiquitous in many applications, e.g., graph cut functions, matroid rank functions, set coverage functions, utility functions in economics, etc. Since the foundational work by Edmonds in 1970 [Edm70], submodular functions and submodular function minimization have served as popular modeling and optimization tools in various areas of theoretical computer science, operations research, game theory, and, more recently, machine learning [B$^+$13].

One of the most fundamental results for submodular function minimization is that it can be solved in strongly polynomial time: this was first shown by Grötschel, Lovász, and Schrijver in the 1980s [GLS81, GLS88] and then combinatorically by Iwata, Fleischer, Fujishige, and Schrijver around the year 2000 [Sch00, IFF01]. These foundational results were recognized as recipients of the Fulkerson prize[1] in 1982 and 2003. After decades of efforts since these initial endeavors [FI03, Iwa03, Vyg03, Orl09, IO09], the best strongly polynomial time algorithms by Lee, Sidford, and Wong [LSW15] and Dadush, Végh, and Zambelli [DVZ18] have $O(n^3 \log^2 n)$ oracle complexity[2] (see Table 1.1 for the history of SFM). It was even believed at that time that no algorithm can achieve sub-cubic oracle complexity even information-theoretically.

[1]The Fulkerson prize is the triennial award for up to best three papers in discrete mathematics.

[2]It is a standard model that submodular functions are given by querying an evaluation oracle.

Authors	Year	Oracle Complexity	Remarks
[GLS81, GLS88]	1981,88	$\widetilde{O}(n^5)$ [McC05]	first strongly
[Sch00]	2000	$O(n^8)$	first comb. strongly
[IFF01]	2000	$O(n^7 \log(n))$	first comb. strongly
[FI03]	2000	$O(n^7)$	
[Iwa03]	2002	$O(n^6 \log(n))$	
[Vyg03]	2003	$O(n^7)$	
[Orl09]	2007	$O(n^5)$	
[IO09]	2009	$O(n^5 \log(n))$	
[LSW15]	2015	$O(n^3 \log^2(n))$	previous best strongly
[LSW15]	2015	$O(n^3 \log(n))$	exponential time
[DVZ18]	2018	$O(n^3 \log^2(n))$	previous best strongly
Chapter 2	2021	$O(n^3 \log\log(n)/\log(n))$	
Chapter 2	2021	$O(n^2 \log(n))$	exponential time

Table 1.1: Strongly polynomial algorithms for submodular function minimization. The oracle complexity measures the number of calls to the evaluation oracle EO. In the case where a paper is published in both conference and journal, the year we provide is the earliest one.

A Reduction to the Shortest Vector Problem. In Chapter 2, we refute this belief by giving an algorithm with strongly polynomial oracle complexity[3] $O(n^2 \log n)$. In particular, we give a generic reduction from submodular function minimization to the shortest vector problem – the most central computational problem in lattice theory, which has been studied for more than a century since Minkowski's theorem from 1889 – that explicitly links their computational complexity for the first time. Our results naturally lead to the hypothesis that $\Theta(n^2)$ is the optimal strongly polynomial oracle complexity for SFM algorithms, and we discuss progress on proving the corresponding lower bounds in Section 1.2 and Chapter 3.

Unified Framework of Minimizing Convex Functions with Rational Minimizers. Most algorithms for submodular function minimization in the last 30 years heavily exploited the geometric and polyhedral properties of submodular functions. In contrast, our algorithm

[3]If the algorithm is required to run in polynomial time, the oracle complexity is $O(n^3 \log\log n/\log n.)$ This computational bottleneck for getting $O(n^2 \log n)$ oracle complexity in polynomial time comes from solving the shortest vector problem and is unrelated to submodular function minimization itself.

uses submodularity only in a minimal way and works much more generally beyond submodular function minimization. In fact, the only property our algorithm uses is that the set of minimizers of the convex function forms a *rational polyhedron*, a setting pioneered in the work of Grötschel, Lovász, and Schrijver [GLS81, GLS88], which they used to establish the polynomial solvability of a wide range of combinatorial optimization problems.

Despite its incredible generality, the original Grötschel-Lovász-Schrijver approach used simultaneous Diophantine approximation to perform dimension reduction; hence, it is not efficient enough to compete with problem-specific algorithms. In our algorithm, we leverage the approximate shortest vector algorithm on an auxiliary lattice to perform dimension reduction much more efficiently, and we give a potential function analysis that correctly amortizes the progress made by the algorithm. As a result, we improve the performance of the Grötschel-Lovász-Schrijver approach by a factor of n in its full generality, leading to our nearly quadratic, strongly polynomial oracle complexity result. This improves over all previous algorithms that are specifically tailored to submodular function minimization via a general algorithmic framework for solving problem (1.1) under condition (1.2).

Chapter Notes. Chapter 2 is based on my solo paper [Jia21], which appeared in the *ACM-SIAM Symposium on Discrete Algorithms (SODA21)*, and its journal version [Jia22], which appeared in the *Journal of the ACM (JACM)*. [Jia22] is a substantial strengthening of the conference version [Jia21] and obtains the full reduction to the shortest vector problem.

1.2 Lower Bounds for Submodular Function Minimization

Given the results of Chapter 2, a natural question to ask is: how tight are the results in Chapter 2, for the general problem of (1.1) and for the more specific problem of submodular function minimization? As we shall remark in Chapter 2, for the general problem of (1.1) where the function f is accessed through a separation oracle (see Definition 2.1.1), the number of separation oracle calls by the algorithm in Chapter 2 is tight up to a $O(\log n)$ factor. But restricting ourselves to the context of submodular function minimization, this question becomes much more intriguing and will be the main topic of discussion in Chapter 3.

Query Complexity Lower Bound for SFM. Despite the rich history of SFM research, obtaining *lower bounds* on the query complexity for SFM has been notoriously difficult. [Har08] described two different constructions of submodular functions whose minimization requires n-queries to an evaluation oracle; in fact, both can be minimized by querying all the n singletons. Later, [CLSW17] showed that one of the examples in [Har08] also needs $n/4$ gradient queries to the Lovász extension of the submodular function. This remained the best lower bound, until recently [GPRW20] proved a $2n$-query lower bound on SFM via a non-trivial construction of a submodular function (which can be minimized in $2n$ queries).

Parallel Lower Bound for SFM. More recently, there has been an interest in understanding the *parallel complexity* of SFM. Note that any SFM algorithm proceeds by making queries to an evaluation oracle in rounds, and the parallel complexity of SFM is the minimum number of rounds (also known as the depth) required by any *query-efficient* SFM algorithm that makes at most $\mathsf{poly}(n)$ evaluation oracle queries. All SFM algorithms above proceed in $\Omega(n)$-rounds. The best-known parallel complexity is obtained by the algorithm in [Jia22] which runs in $O(n \log n)$ rounds. On the lower bound side, [BS20] proved that any query-efficient SFM algorithm must proceed in $\Omega(\log n / \log \log n)$-rounds. This was improved in [CCK21] to an $\widetilde{\Omega}(n^{1/3})$-lower bound on the number of rounds for query-efficient SFM. The latter paper also mentioned a bottleneck of $n^{1/3}$ to their approach and left open the question of whether a nearly-linear number of rounds are needed, or whether there is a query-efficient SFM algorithm proceeding in $n^{1-\delta}$ many rounds for some absolute constant $\delta > 0$.

Improved Lower Bounds for SFM. In Chapter 3, we provide improved lower bounds for both the query complexity for SFM, and the parallel complexity for query-efficient parallel SFM. Our first main result is that any deterministic SFM algorithm requires $\Omega(n \log n)$ queries to an evaluation oracle. This result is the first super-linear query complexity lower bound in the history of SFM, lending support for the hypothesis that SFM needs quadratic queries to the evaluation oracle. Our second main result is that any parallel SFM algorithm making at most $\mathsf{poly}(n)$ queries must proceed in $\Omega(n/\log n)$ parallel rounds, which matches

the parallel complexity of the algorithm in Chapter 2 up to poly-logarithmic factors.

Chapter Notes. Chapter 3 is based on my joint work with Deeparnab Chakrabarty, Andrei Graur, and Aaron Sidford [CGJS22] which appeared in the *63rd IEEE Symposium on Foundations of Computer Science (FOCS 2022)*.

1.3 Unified Algorithms Through the Lens of Discrepancy Theory

When condition (1.2) fails, then problem (1.1) typically becomes computationally intractable. This is the case even when f is a linear function defined over a polytope (and is infinity outside of it), which already captures the NP-Hard problem of Integer Linear Programming. For such computationally intractable problems, one prominent strategy is to solve the relaxation of (1.1) with the linear constraints $x \in [0,1]^n$ instead and then round the fractional solution to an integer one. The major question then is how to bound the rounding error.

In seminal work, Lovász, Spencer, and Vesztergombi [LSV86] showed that the rounding error of a linear system is closely related to the *hereditary discrepancy* of the linear constraints, but this connection had few algorithmic consequences at that time because most fundamental discrepancy results were non-algorithmic [Spe85, Ban98]. Following a groundbreaking result of Bansal [Ban10] in 2010, the past decade has seen a surge of progress in giving efficient algorithms for fundamental discrepancy results [LM15a, Rot17, ES18, BDG19, LRR17, BDGL18, DNTTJ18], opening up an opportunity for rounding, and more generally algorithm design, through the unified lens of discrepancy theory.

Notably, discrepancy theory has recently found a wide range of applications in many different areas: coresets and sketches in machine learning [Phi09, KL19], approximation algorithms [EPR13, BRS22], graph sparsification [RR20a], randomized experiment design [HSSZ19], (quantum) random access codes [ANTSV02, HRS22, BJM22b], differential privacy [MN12, NTZ13], and many more classical applications, such as quasi-Monte Carlo, sampling, and pseudorandomness that are presented in classical textbooks [Cha00, Mat99].

Despite these significant applications, many fundamental discrepancy questions are not yet well understood, resulting in the sub-optimality of many of the current applications of

discrepancy theory to rounding and algorithm design. In Chapters 4 - 9 of this thesis, we present substantial progress on two central discrepancy problems: matrix discrepancy theory and online discrepancy theory.

Matrix Discrepancy Theory. Matrix discrepancy theory has attracted significant attention in the last decade: the classical result of Batson, Spielman, and Srivastava on graph sparsification [BSS12], Marcus, Spielman, and Srivastava's breakthrough on the Kadison-Singer problem [MSS15], and the widely-known notion of quantum random access codes [ANTSV02] all essentially concern matrix discrepancy [RR20a, HRS22]. However, fundamental matrix discrepancy questions remain unanswered, a famous example of which is the matrix Spencer conjecture [Zou12, Mek14].

Conjecture 1.3.1 (Matrix Spencer Conjecture, [Zou12, Mek14])**.** *Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m\times m}$ with each $\|A_i\|_{\mathsf{op}} \leq 1$, there exist signs $x \in \{\pm 1\}^n$ such that $\left\|\sum_{i=1}^n x_i A_i\right\|_{\mathsf{op}} \leq O(\sqrt{n} \cdot \max\{1, \sqrt{\log(m/n)}\})$. In particular, the matrix discrepancy is $O(\sqrt{n})$ for $m = n$.*

Conjecture 1.3.1 generalizes a seminal result of Spencer [Spe85] (see Theorem 4.1.1) which establishes the conjecture when all matrices A_i are diagonal. Apart from this classical result, progress on the matrix Spencer conjecture has stagnated for a long time despite tremendous effort. In Chapter 4 - 6, we describe multiple progress, in chronological order, towards resolving this conjecture. In Chapter 4, we give the first non-trivial improvement over the naïve random coloring bound for this problem, based on a general framework for proving matrix discrepancy bounds via mirror descent. In Chapter 5, we use the inverse polynomial barrier to obtain a strengthening of the result in Chapter 4 and in independent work by Hopkins, Raghavendra, and Shetty [HRS22].

Finally, in Chapter 6, we show how to resolve the matrix Spencer conjecture up to polylogarithmic rank. This result implies an almost tight $\log n - 3\log\log n$ qubit lower bound for quantum random access codes encoding n classical bits with advantage $\gg 1/\sqrt{n}$. It also demonstrates a phase transition phenomenon for quantum random access codes that is

absent in the classical world: while $(1/2)\log n$ qubits suffice to achieve $c/\sqrt{n}$ advantage for a small constant $c > 0$, it requires at least $\log n - 3\log\log n$ qubits to obtain $C/\sqrt{n}$ advantage for some large constant $C > 0$.

Online Discrepancy Theory. Consider the following online discrepancy question: vectors $v_1, v_2, \ldots, v_T \in \mathbb{R}^n$ arrive online, and upon the arrival of each vector v_t, a sign $\chi_t \in \{\pm 1\}$ must be chosen irrevocably so that the ℓ_∞-norm of the signed sum $d_t = \chi_1 v_1 + \ldots + \chi_t v_t$, also called the discrepancy, remains as small as possible. This problem was initially studied in the 1970s but did not receive much interest since the $\sqrt{T}$ discrepancy by random coloring cannot be improved against adaptive adversaries [Spe77].

It is therefore natural to ask if relaxing the power of the adversary can lead to interesting new algorithms that achieve $\mathsf{poly}(\log T)$ discrepancy. In Chapters 7 and 8, we gave the first algorithms achieving poly-logarithmic bounds for the online discrepancy problem in the stochastic[4] setting. Our results also imply poly-logarithmic bounds for online geometric discrepancy (Tusnády's problem) and online fair allocation that minimizes *envy*, a notion of fairness studied widely in the economics literature [Fol66, TV85, LMMS04, Bud11]. A simple but very powerful algorithm, known as the *self-balancing walk*, for the online discrepancy problem in the more difficult oblivious adversary setting was given by Alweiss, Liu, and Sawhney [ALS21]. Their original analysis of this algorithm is based on the notion of mean-preserving spread and is less explicit. In Chapter 9 of this thesis, we present a more direct folklore[5] proof of their result, which also appeared later in [DM21].

Chapter Notes. These chapters are based on multiple joint papers with Daniel Dadush, Victor Reis, Nikhil Bansal, Raghu Meka, Sahil Singla, and Makrand Sinha [DJR22, BJM22a, BJM22b, BJSS20, BJM+21]. Matrix discrepancy theory is studied in Chapters 4 - 6, and the three different approaches we present for this problem, one in each chapter, are based on

[4]The stochasticity assumption is sometimes necessary, even in the simple case of interval discrepancy [JKS19].

[5]To the best of our knowledge, this potential-based analysis of the self-balancing walk has been independently reconstructed multiple times.

the three papers [DJR22, BJM22a, BJM22b]. The online discrepancy problem is considered in Chapters 7 - 9, where the approaches we present are based on the papers [BJSS20] and [BJM+21], and a folklore proof of the result in [ALS21].

Part I

CONVEX FUNCTIONS WITH INTEGER MINIMIZERS

Chapter 2

MINIMIZING CONVEX FUNCTIONS WITH RATIONAL MINIMIZERS

In this chapter, we study the problem of minimizing a convex function whose minimizer is an integer (and more generally, rational) point. We present a general algorithm for this problem that improves upon the framework of Grötschel, Lovász, and Schrijver in its full generality. Our algorithm is based on a generic reduction of the problem to the Shortest Vector Problem in computational lattice theory. This chapter is based on my single-author paper [Jia21], which appeared in the *ACM-SIAM Symposium on Discrete Algorithms (SODA21)*, and its journal version [Jia22], which appeared in the *Journal of the ACM (JACM)*.

2.1 Introduction

We investigate the problem of minimizing a convex function f on $\mathbb{R}^n$ accessed through a separation oracle SO [GLS81]. When queried with a point x, the oracle returns "YES" if x minimizes f; otherwise, the oracle returns a hyperplane that separates x from the minimizer of f. An algorithm is said to be *strongly polynomial* [GLS88] for such a problem if it makes $\mathsf{poly}(n)$ calls to SO, uses $\mathsf{poly}(n)$ arithmetic operations, and the size of numbers occurring during the algorithm is polynomially bounded by n and the size of the output of the separation oracle.

Designing strongly polynomial algorithms for continuous optimization problems with certain underlying combinatorial structure is a well-studied but challenging task in general. To this date, despite tremendous effort, it remains a major open question to solve linear programming (LP) in strongly polynomial time. This problem is also widely known as Smale's 9th question [Sma98]. Despite this barrier, such algorithms are known under additional as-

sumptions: linear systems with at most two non-zero entries per row [Meg83, AC91, CM94] or per column [Vég17, OV20] in the constraint matrix, LPs with bounded entries in the constraint matrix [Tar86, VY96, DHNV20], and LPs with 0-1 optimal solutions [Chu12, Chu15].

For minimizing a general convex function f, strongly polynomial algorithms are hopeless unless f satisfies certain combinatorial properties. In this work, we study the setting where the minimizer of f is an integral point inside a box with radius[1] $R = 2^{\mathsf{poly}(n)}$. The integrality assumption on the minimizer is natural, and is general enough to encapsulate well-known problems such as submodular function minimization, where $R = 1$. Prior to our work, an elegant application of simultaneous Diophantine approximation due to Grötschel, Lovász and Schrijver [GLS84, GLS88] gives[2] a strongly polynomial algorithm that minimizes f using $O(n^2(n+\log(R)))$ calls to the separation oracle and an exponential time algorithm that finds the minimizer of f using $O(n^2 \log(nR))$ oracle calls.

In fact, Grötschel, Lovász and Schrijver's approach applies to the more general setting of rational polyhedra, which they use to derive polynomial time algorithms for a wide range of combinatorial optimization problems [GLS81, GLS88]. In the rational polyhedra setting, the set of minimizers of f is a polyhedron K^* inside a box with radius R, and the vertices of K^* are all rational vectors with LCM vertex complexity[3] bounded by at most $\varphi \geq 0$ (Definition 2.3.6). In particular, the case of integral minimizers in the previous paragraph corresponds to when $\varphi = 0$. For the more general setting of rational polyhedra, Grötschel, Lovász and Schrijver's approach implies a polynomial time algorithm that finds a vertex of K^* using $O(n^2(n + \varphi + \log(R)))$ separation oracle calls, and an exponential time algorithm

[1] It's easy to show that strongly polynomial algorithm doesn't exist if $\log(R)$ is super-polynomial (see Remark 2.1.4).

[2] The original approach by Grötschel, Lovász and Schrijver was given in the context of obtaining exact solutions to LP, but it is immediately applicable to our problem. Their approach was briefly described in [GLS84] with details given in [GLS88]. Their approach originally used the ellipsoid method which is sub-optimal in terms of oracle complexity. The oracle complexity given here uses Vaidya's cutting plane method [Vai89].

[3] Here we use a slightly different definition from Grötschel, Lovász and Schrijver's original definition of vertex complexity in [GLS81, GLS88] so that $\varphi = 0$ corresponds to the setting of integral minimizers. More details can be found in Section 2.3.

that uses $O(n^2(\varphi + \log(nR)))$ oracle calls. We refer interested readers to [GLS88, Chapter 6] for a detailed presentation of their approach. The purpose of the present chapter is to design a new method to improve the number of separation oracle calls.

A closely related problem, known as the Convex Integer Minimization problem (problem (1.1)), asks to minimize a convex function f over the set of integer points. Dadush [Dad12, Section 7.5] gave an algorithm for this problem that takes $n^{O(n)}$ time and exponential space. In fact, the Convex Integer Minimization problem generalizes integer linear programming and thus cannot be solved in sub-exponential time under standard complexity assumptions, so the integrality/rationality assumption on the minimizer of f is, in some sense, necessary for obtaining efficient algorithms.

The number of separation oracle calls made by an algorithm for minimizing a convex function f, known as the *oracle complexity*, plays a central role in black-box models of convex optimization. For weakly polynomial algorithms, it's well-known that $\Theta(n \log(nR/\epsilon))$ oracle calls is optimal, with ϵ being the accuracy parameter. The first exponential time algorithm that achieves the optimal oracle complexity is the famous center of gravity method discovered independently by Levin [Lev65] and Newman [New65]. As for polynomial time algorithms, an oracle complexity of this order was first achieved over thirty years ago by the method of inscribed ellipsoids [KTE88, NN89]. In contrast, the optimal oracle complexity for strongly polynomial algorithms is largely unknown to this date. This motivates the present chapter to place a focus on the oracle complexity aspect of our algorithms.

2.1.1 Our results

To formally state our result, we first define the notion of a separation oracle as formulated in [GLS81].

Definition 2.1.1 (Separation oracle [GLS81])**.** *Let f be a convex function on $\mathbb{R}^n$ and K^* be the set of minimizers of f. Then a (strong) separation oracle* SO *for f is one that:*

(a) when queried with a minimizer $x \in K^*$, it outputs "YES";

(b) when queried with a point $x \notin K^*$, it outputs a non-zero vector $c \in \mathbb{R}^n$ such that $\min_{y \in K^*} c^\top y > c^\top x$.

The setting of integral minimizers. The main result of the chapter in this setting is the following reduction to the Shortest Vector Problem (see Section 2.3.2) given in Theorem 2.1.2. The seemingly strong assumption ($\star$) guarantees that our algorithm finds an *integral* minimizer of f, which is crucial for our application to submodular function minimization. To find an arbitrary minimizer of f, we only need the much weaker assumption that f has an integral minimizer (see Remark 2.1.5).

Theorem 2.1.2 (Main result for integral minimizers)**.** *Given a separation oracle* SO *for a convex function f defined on $\mathbb{R}^n$, and a γ-approximation algorithm* ApproxSVP *for the shortest vector problem which takes T_{SVP} arithmetic operations. If the set of minimizers K^* of f is contained in a box of radius R and satisfies*

($\star$) all extreme points of K^* are integral,

then there is a randomized algorithm that with high probability finds an integral minimizer of f using $O(n \log(\gamma n R))$ calls to SO *and* $\mathsf{poly}(n, \log(\gamma R)) \cdot T_{\text{SVP}}$ *arithmetic operations.*

In particular, taking ApproxSVP to be the polynomial time $2^{n \log\log(n)/\log(n)}$-approximation algorithm in [AKS01] (which improves upon the celebrated LLL algorithm [LLL82] and Schnorr's block reduction algorithm [Sch87]), or the exponential time algorithms for exact SVP [AKS01, MV13, ADRSD15] give the following corollary.

Corollary 2.1.3 (Instantiations of main result)**.** *Under the same assumptions as in Theorem 2.1.2, there is a randomized algorithm that with high probability finds an integral minimizer of f using*

(a) $O(n(n \log\log(n)/\log(n) + \log(R)))$ calls to SO *and* $\mathsf{poly}(n, \log(R))$ *arithmetic operations, or*

(b) $O(n\log(nR))$ *calls to* SO *and* $\exp(O(n))\cdot\mathsf{poly}(\log(R))$ *arithmetic operations.*

More generally, for any integer $r>1$, one can use the $r^{O(n/r)}$-approximation algorithm in $2^{O(r)}\mathsf{poly}(n)$ time for SVP given in [AKS01, MV13] to obtain a smooth tradeoff between time and oracle complexity in Theorem 2.1.2, but we omit the explicit statements of these results.

Remark 2.1.4 (Assumption $(\star)$ and lower bound)**.** *Without assumption* $(\star)$*, we give a* $2^{\Omega(n)}$ *information theoretic lower bound on the number of* SO *calls needed to find an integral minimizer of* f*. Consider the unit cube* $K=[0,1]^n$ *and let* $V(K)=\{0,1\}^n$ *be the set of vertices. For each* $v\in V(K)$*, define the simplex* $\Delta(v)=\{x\in K:\|x-v\|_1<0.01\}$*. Randomly pick a vertex* $u\in V(K)$ *and consider the convex function*

$$f_u(x)=\begin{cases}0 & x\in K\setminus(\cup_{v\in V(K)\setminus\{u\}}\Delta(v))\\ \infty & \text{otherwise}\end{cases}.$$

When queried with a point $x\in\Delta(v)$ *for some* $v\in V(K)\setminus\{u\}$*, we let* SO *output a separating hyperplane* H *such that* $K\cap H\subseteq\Delta(v)$*; when queried with* $x\notin K$*, we let* SO *output a hyperplane that separates* x *from* K*. Notice that* u *is the unique integral minimizer of* f_u*, and to find* u*, one cannot do better than randomly checking vertices in* $V(K)$ *which takes* $2^{\Omega(n)}$ *queries to* SO*.*

We next argue that $\Omega(n\log(R))$ *calls to* SO *is information theoretically necessary in Theorem 2.1.2. Consider* f *with a unique integral minimizer which is a random integral point in* $B_\infty(R)\cap\mathbb{Z}^n$*, where* $B_\infty(R)$ *is the* ℓ_∞ *ball with radius* R*. In this case, one cannot hope to do better than just bisecting the search space for each call to* SO *and this strategy takes* $\Omega(n\log(R))$ *calls to* SO *to reduce the size of the search space to a constant.*

Remark 2.1.5 (A weaker assumption)**.** *As shown in the previous remark, it is impossible in general to find an integral minimizer of* f *efficiently without assumption* $(\star)$*. However, one can still find a minimizer (which is not necessarily integral) of* f *under the much weaker*

assumption that f has an integral minimizer, i.e. $K^* \cap \mathbb{Z}^n \neq \emptyset$. In this case, one can use the same algorithm as in Theorem 2.1.2 until SO *first returns "YES" and simply output the query point. The guarantees in Theorem 2.1.2 also applies to this case.*

Generalization to the rational polyhedra setting. Theorem 2.1.2 generalizes to the setting of rational polyhedra, where the set of minimizers K^* of f is a polyhedron contained in a box of radius R, and all vertices of K^* are rational vectors with LCM vertex complexity at most $\varphi \geq 0$. Roughly speaking, this means that the least common multiple of the denominators in the fractional representation of each vertex is upper bounded by 2^φ. We postpone the precise definitions of LCM vertex complexity and rational polyhedra to Section 2.3 (Definition 2.3.6 and 2.3.7). The proof of the following theorem (which also implies Theorem 2.1.2) will be given in Section 2.6.

Theorem 2.1.6 (Main result for rational polyhedra)**.** *Given a separation oracle* SO *for a convex function f defined on $\mathbb{R}^n$, and a γ-approximation algorithm* ApproxSVP *for the shortest vector problem which takes T_{SVP} arithmetic operations. If the set of minimizers K^* of f is a rational polyhedron contained in a box of radius R and has LCM vertex complexity at most $\varphi \geq 0$, then there is a randomized algorithm that with high probability finds a vertex of K^* using $O(n(\varphi + \log(\gamma n R)))$ calls to* SO *and* $\mathsf{poly}(n, \varphi, \log(\gamma R)) \cdot T_{\text{SVP}}$ *arithmetic operations.*

2.1.2 Application to Submodular Function Minimization

Submodular function minimization (SFM) has been recognized as an important problem in the field of combinatorial optimization. Classical examples of submodular functions include graph cut functions, set coverage function, and utility functions from economics. Since the seminal work by Edmonds in 1970 [Edm70], SFM has served as a popular tool in various fields such as theoretical computer science, operations research, game theory, and machine learning. For a more comprehensive account of the rich history of SFM, we refer interested readers to the excellent surveys [McC05, Iwa08].

The formulation of SFM we consider is the standard one: we are given a submodular function f defined over subsets of an n-element ground set. The values of f are integers, and are evaluated by querying an evaluation oracle that takes time EO. Since the breakthrough work by Grötschel, Lovász, Schrijver [GLS81, GLS88] that the ellipsoid method can be used to construct a strongly polynomial algorithm for SFM, there has been a vast literature on obtaining better strongly polynomial algorithms (see Table 1.1). These include the very first combinatorial strongly polynomial algorithms constructed by Iwata, Fleischer and Fujishige [IFF01] and Schrijver [Sch00]. Very recently, a major improvement was made by Lee, Sidford and Wong [LSW15] using an improved cutting plane method. Their algorithm achieves the state-of-the-art oracle complexity of $O(n^3 \log^2(n))$ for strongly polynomial algorithms. A simplified variant of this algorithm achieving the same oracle complexity was given in [DVZ18].

The authors of [LSW15] also noted that $O(n^3 \log(n))$ oracle calls are information theoretically sufficient for SFM ([LSW15, Theorem 71]), but were unable to give an efficient algorithm achieving such an oracle complexity. They asked as open problems ([LSW15, Section 16.1]):

(a) whether there is a strongly polynomial algorithm achieving the $O(n^3 \log(n))$ oracle complexity;

(b) whether one could further (even information theoretically) remove the extraneous $\log(n)$ factor from the oracle complexity.

The significance of these questions stem from their belief that $\Theta(n^3)$ is the tight oracle complexity for strongly polynomial algorithms for SFM (see [LSW15, Section 16.1] for a more detailed discussion).

We answer both these open questions affirmatively in the following Theorem 2.1.7, which follows from applying Corollary 2.1.3 to the Lovász extension $\hat{f}$ of the function f, together with the standard fact that a separation oracle for $\hat{f}$ can be implemented using n calls to

the evaluation oracle ([LSW15, Theorem 61]). We provide details on these definitions and the proof of Theorem 2.1.7 in Section 2.7.

Theorem 2.1.7 (Submodular function minimization)**.** *Given an evaluation oracle* EO *for a submodular function f defined over subsets of an n-element ground set, there exist*

(a) a strongly polynomial algorithm that minimizes f using $O(n^3 \log\log(n)/\log(n))$ calls to EO*, and*

(b) an exponential time algorithm that minimizes f using $O(n^2 \log(n))$ calls to EO*.*

To the best of our knowledge, the results in Theorem 2.1.7 represent the first algorithms that achieve $o(n^3)$ oracle complexity for SFM, even information theoretically. The first result in Theorem 2.1.7 breaks the natural $O(n^3)$ barrier for the oracle complexity of strongly polynomial algorithms. The second result pushes the information theoretic oracle complexity for exact SFM down to nearly quadratic.

Our algorithm is conceptually simpler than the algorithms given in [LSW15, DVZ18]. Moreover, while most of the previous strongly polynomial algorithms for SFM vastly exploit different combinatorial structures of submodularity, our result is achieved via a very general algorithm and uses the structural properties of submodular functions in a minimal way.

2.2 Proof Overview

Without loss of generality, we may assume that f has a unique minimizer x^* in Theorem 2.1.2 and 2.1.6. To justify this statement, suppose the set of minimizers K^* of f satisfies assumption $(\star)$. Let $x^* \in K^*$ be the unique lexicographically minimal minimizer, i.e. every other minimizer $x \in K^*$ satisfies $x_i > x_i^*$ for the smallest coordinate $i \in [n]$ in which $x_i \neq x_i^*$. Whenever SO is queried at a minimizer $y \in K^*$ and outputs "YES", our algorithm continues to minimize the linear objective $e_i^\top x$, where $i \in [n]$ is the smallest index such that the ith standard orthonormal basis vector e_i is not orthogonal to the current working subspace, by

pretending that SO returns[4] the vector $-e_i$ (until its search set contains a single point). Equivalently, our algorithm minimizes the linear objectives $e_1^\top x, \cdots, e_n^\top x$ in the given order inside K^*, and this optimization problem has the unique solution x^*. We make the assumption that f has a unique minimizer x^* in the rest of this chapter.

For simplicity, we further assume in the subsequent discussions that $x^* \in \{0,1\}^n$, i.e. $R = 1$ in the setting of integral minimizer, which does not change the problem inherently.

On a high level, our algorithm maintains a convex search set K that contains the integral minimizer x^* of f, and iteratively shrinks K using the cutting plane method; as the volume of K becomes small enough, our algorithm finds a hyperplane P that contains all the integral points in K and recurse on the lower-dimensional search set $K \cap P$. The assumption that x^* is integral guarantees that $x^* \in K \cap P$. This natural idea was previously used in [GLS84, GLS88] to handle rational polytopes that are not full-dimensional and in [LSW15] to argue that $O(n^3 \log(n))$ oracle calls is information theoretically sufficient for SFM. The main technical difficulties in efficiently implementing such an idea are two-fold:

(a) we need to find the hyperplane P that contains $K \cap \mathbb{Z}^n$;

(b) we need to carefully control the amount $\mathsf{vol}(K)$ is shrunk so that progress is not lost.

The second difficulty is key to achieving a small oracle complexity and deserves some further explanation. To see why shrinking K arbitrarily might result in a loss of progress, it's instructive to consider the following toy example: suppose an algorithm starts with the unit cube $K = [0,1]^n$ and x^* lies on the hyperplane $K_1 = \{x : x_1 = 0\}$; suppose the algorithm obtains, in its ith call to SO, the halfspace $H_i = \{x : x_1 \le 2^{-i}\}$. After T calls to SO, the algorithm obtains the refined search set $K \cap H_T$ with volume 2^{-T}. However, when the algorithm reduces the dimension and recurses on the hyperplane K_1, the $(n-1)$-dimensional

[4]Note that this implementation of the separation oracle for the lexicographically minimal minimizer x^* does not quite satisfy the conditions in Definition 2.1.1. In particular, even when x^* is queried, the separation oracle for finding x^* might not realize it unless the current working subspace is trivial (i.e. 0-dimensional). However, all our results and proofs still hold under this slightly weaker implementation of the separation oracle.

volume of the search set again becomes 1, and the progress made by the algorithm in shrinking the volume of K is entirely lost. In contrast, the correct algorithm can reduce the dimension after only one call to SO when it's already clear that $x^* \in K_1$.

The Grötschel-Lovász-Schrijver Approach. For the moment, let's take K to be an ellipsoid. Such an ellipsoid can be obtained by Vaidya's volumetric center cutting plane method[5] [Vai89]. One natural idea to find the hyperplane comes from the following geometric intuition: when the ellipsoid K is "flat" enough in one direction, then all of its integral points lie on a hyperplane P. To find such a hyperplane P, Grötschel, Lovász and Schrijver [GLS84, GLS88] gave an elegant application of simultaneous Diophantine approximation. We explain the main ideas behind this application in the following. We refer interested readers to [GLS88, Chapter 6] for a more comprehensive presentation of their approach and its implications to finding exact LP solutions.

For simplicity, we assume K is centered at 0. Let a be the unit vector parallel to the shortest axis of K and $\mu_{\min}$ be the Euclidean length of the shortest axis of K. Approximating the vector a using the efficient simultaneous Diophantine approximation algorithm by Lenstra, Lenstra and Lovász [LLL82], one obtains an integral vector $v \in \mathbb{Z}^n$ and a positive integer $q \in \mathbb{Z}$ such that

$$\|qa - v\|_\infty < 1/3n \qquad \text{and} \qquad 0 < q < 2^{2n^2}.$$

This implies that for any integral point $x \in K \cap \{0,1\}^n$,

$$|v^\top x| \leq |qa^\top x| + \frac{1}{3n} \cdot \|x\|_1 \leq q \cdot \mu_{\min} + 1/3.$$

When $\mu_{\min} < 2^{-3n^2}$, the integral inner product $v^\top x$ has to be 0 and therefore all integral

[5]Perhaps a more natural candidate is the ellipsoid method developed in [YN76, Sho77, Kha80]. This method, however, shrinks the volume of K by a factor of $O(n)$ slower than Vaidya's method. In fact, the Grötschel-Lovász-Schrijver approach [GLS84] originally used the ellipsoid method which results in an oracle complexity of $O(n^4)$ for their polynomial time algorithm.

points in K lie on the hyperplane $P = \{x : v^\top x = 0\}$. An efficient algorithm immediately follows: we first run the cutting plane method until the shortest axis of K has length $\mu_{\min} \approx 2^{-3n^2}$, then apply the above procedure to find the hyperplane P on which we recurse.

To analyze the oracle complexity of this algorithm, one naturally uses $\mathsf{vol}(K)$ as the potential function. An amortized analysis using such a volume potential previously appeared, for example, in [DVZ20] for finding maximum support solutions in the linear conic feasibility problem. Roughly speaking, each cutting plane step (corresponding to one oracle call) decreases $\mathsf{vol}(K)$ by a constant factor; each dimension reduction step increases $\mathsf{vol}(K)$ by roughly $1/\mu_{\min} \approx 2^{3n^2}$. As there are n dimension reduction steps before the problem becomes trivial, the total number of oracle calls is thus $O(n^3)$. The exponential time oracle complexity bound of $O(n^2 \log(n))$ can be obtained similarly by using Dirichlet's approximation theorem on simultaneous Diophantine approximation (e.g. [Cas71, Section 1.10]) instead.

One might wonder if the oracle complexity upper bound for their polynomial time algorithm can be improved using a better analysis. However, there is some fundamental issue in getting such an improvement. In particular, the upper bound of $2^{O(n^2)}$ on q in efficient simultaneous Diophantine approximation corresponds to the $2^{O(n)}$-approximation factor of the Shortest Vector Problem in lattices, first obtained by Lenstra, Lenstra and Lovász [LLL82]. Despite forty years of effort, this approximation factor has only been improved slightly to $2^{n \log\log(n)/\log n}$ for polynomial time algorithms [AKS01].

Lattices to the Rescue: A Reduction to the Shortest Vector Problem. To bypass the previous bottleneck and prove Theorem 2.1.2, we give a reduction to the Shortest Vector Problem directly. We give a new method to find the hyperplane for dimension reduction based on an approximately shortest vector of certain lattice, and analyze its oracle complexity via a novel potential function that captures simultaneously the volume of the search set K and the density of the lattice. The change in the potential function after dimension reduction is analyzed through a high dimensional slicing lemma. The details for this algorithm and its analysis are given in Section 2.5 and 2.6.

Finding the hyperplane. We maintain a polytope K (which we assume to be centered at 0 for simplicity) using an efficient implementation of the center of gravity method due to Bertsimas and Vempala [BV04]. The following sandwiching condition is standard in convex geometry

$$E(\mathsf{Cov}(K)^{-1}) \subseteq K \subseteq 2n \cdot E(\mathsf{Cov}(K)^{-1}), \tag{2.1}$$

where $\mathsf{Cov}(K)$ is the covariance matrix of the uniform distribution over K. Sufficiently good approximation to $\mathsf{Cov}(K)$ can be obtained efficiently by sampling from K [BV04] so we ignore any computational issue for now.

To find a hyperplane P that contains all integral points in K, it suffices to preserve all the integral points in the outer ellipsoid $E = 2n \cdot E(\mathsf{Cov}(K)^{-1})$ on the RHS of (2.1). Let $x \in E \cap \mathbb{Z}^n$ be an arbitrary integral point. For any vector v,

$$|v^\top x| \leq \|v\|_{\mathsf{Cov}(K)} \cdot \|x\|_{\mathsf{Cov}(K)^{-1}} \leq 2n \cdot \|v\|_{\mathsf{Cov}(K)}. \tag{2.2}$$

As long as $\|v\|_{\mathsf{Cov}(K)} < 1/10n$ and $v^\top x$ is an integer, we can conclude that $v^\top x = 0$ and this implies that all integral points in K lie on the hyperplane $P = \{x : v^\top x = 0\}$. Note that by (2.2), such a vector v with small $\|v\|_{\mathsf{Cov}(K)}$ essentially controls the ellipsoid width $\mathsf{width}_E(v) := \max_{x \in E} v^\top x - \min_{x \in E} v^\top x$.

One might attempt to guarantee that $v^\top x$ is integral by choosing v to be an integral vector. However, this idea has a fundamental flaw: as the algorithm reduces the dimension by restricting on a subspace W, the set of integral points on W might become much *sparser*. As such, one needs $\mathsf{vol}(K)$ to be very small to guarantee that $\|v\|_{\mathsf{Cov}(K)} < 1/10n$ and this results in a very large oracle complexity.

To avoid this issue, we take $v = \Pi_W(z) \neq 0$ as the projection of some integral point $z \in \mathbb{Z}^n$ on W, where W is the subspace on which K lies. Since $z - v \in W^\perp$, we have $v^\top x = z^\top x$ and this guarantees that $v^\top x$ is integral. For the general case where K is not

centered at 0, a simple rounding procedure computes the desired hyperplane. We postpone the details of constructing the hyperplane to Lemma 2.4.1.

How do we find a vector $v \in \Pi_W(\mathbb{Z}^n) \setminus \{0\}$ that satisfies $\|v\|_{\mathsf{Cov}(K)} < 1/10n$? This is where lattices come into play. In particular, since $\Lambda = \Pi_W(\mathbb{Z}^n)$ forms a lattice, we can apply any γ-approximation algorithm for the Shortest Vector Problem. If the shortest non-zero vector in Λ has $\mathsf{Cov}(K)$-norm at most $1/10\gamma n$, then we can find a non-zero vector v that satisfies $\|v\|_{\mathsf{Cov}(K)} < 1/10n$.

The algorithm. This new approach for finding the hyperplane immediately leads to the following algorithm: we run the approximate center of gravity method for one step to decrease the volume of the polytope K by a constant factor; then we run the γ-approximation algorithm for SVP to find a non-zero vector v for dimension reduction. If $\|v\|_{\mathsf{Cov}(K)} \geq 1/10n$, then we continue to run the cutting plane method; otherwise, we use the above procedure to find a hyperplane P containing all integral points in K, update the polytope K to be $K \cap P$ and recurse.

Potential function analysis. To analyze such an algorithm, one might attempt to use $\mathsf{vol}(K)$ as the potential function as in the Grötschel-Lovász-Schrijver approach. However, one quickly realizes that $\mathsf{vol}(K \cap P)/\mathsf{vol}(P)$ can be as large as $\|v\|_2 / \|v\|_{\mathsf{Cov}(K)}$. While it's expectable that $\|v\|_{\mathsf{Cov}(K)}$ is not too small since we are frequently checking for a short lattice vector, one has no control over $\|v\|_2$ in general.

Key to our analysis is the potential function $\Phi = \mathsf{vol}(K) \cdot \det(\Lambda)$ that measures simultaneously the volume of K and the covolume $\det(\Lambda)$ of the lattice Λ. Essentially, this potential function controls the lattice width $\min_{v \in \Lambda \setminus \{0\}} \mathsf{width}_E(v)$ of the outer ellipsoid E. In fact, Minkowski's first theorem (Theorem 2.3.4) implies that there always exists a vector $v \in \Lambda \setminus \{0\}$ such that $\mathsf{width}_E(v) \leq \mathsf{poly}(n) \cdot \Phi^{1/n}$, and thus the potential function would never get too small before dimension reduction takes place.

Continuing with the analysis via the potential function Φ, while $\mathsf{vol}(K)$ increases by $\|v\|_2 / \|v\|_{\mathsf{Cov}(K)}$ after the dimension reduction, standard fact on lattice projection (Fact 2.3.2)

shows that the covolume of the lattice decreases by a factor of $\|v\|_2$. The decrease in the covolume of the lattice thus elegantly cancels out the increase in $\mathsf{vol}(K)$, leading to an overall increase in the potential of at most $1/\|v\|_{\mathsf{Cov}(K)} = O(\gamma n)$. It follows that the total increase in the potential over all n dimension reduction steps is at most $(\gamma n)^n$. Note that each cutting plane step still decreases the potential function by a constant factor since the lattice is unchanged. Therefore, the total number of oracle calls is at most $O(n \log(\gamma n))$.

High dimensional slicing lemma for consecutive dimension reduction steps. The argument above ignores a slight technical issue: while we can guarantee that $\|v\|_{\mathsf{Cov}(K)} \geq 1/\gamma n$ after cutting plane steps by checking for short non-zero lattice vectors, it's not clear why $\|v\|_{\mathsf{Cov}(K)}$ cannot be too small after a sequence of dimension reduction steps. It turns out that this can happen only when $\mathsf{Cov}(K)$ becomes much smaller (e.g. the hyperplane P is far from the centroid of K) after dimension reduction, in which case $\mathsf{vol}(K)$ as well as the potential also become much smaller.

To formally analyze the change in the potential function after a sequence of k consecutive dimension reduction steps, we note that the polytope K (which we assume to be isotropic for simplicity) becomes a "slice" $K \cap W$ and the lattice Λ becomes the projected lattice $\Pi_W(\Lambda)$, where W is a subspace. One can show using standard convex geometry tools that $\mathsf{vol}(K \cap W)/\mathsf{vol}(K)$ is at most $k^{O(k)}$, and via Minkowski's first theorem that $\det(\Pi_W(\Lambda))/\det(\Lambda)$ is at most $\sqrt{k}^k/\lambda_1(\Lambda)^k$, where $\lambda_1(\Lambda)$ is the Euclidean length of the shortest non-zero vector in Λ. We leave the details of this high dimensional slicing lemma to Lemma 2.4.2. Since we know that $\lambda_1(\Lambda) \geq 1/\gamma n$ in the first dimension reduction step, the potential function increases by a factor of at most $(\gamma n)^{O(k)}$ over a sequence of k consecutive dimension reduction steps. This gives a more precise analysis of the $O(n \log(\gamma n))$ oracle complexity.

2.3 Preliminaries

2.3.1 Notations

We use $\mathbb{R}_+$ to denote the set of non-negative real numbers. For any positive integer n, we use $[n]$ to denote the set $\{1, \cdots, n\}$. Given a real number $a \in \mathbb{R}$, the floor of a, denoted as $\lfloor a \rfloor$, is the largest integer that is at most a. Define the closest integer to a, denoted as $\lceil a \rfloor$, to be $\lceil a \rfloor := \lfloor a + 1/2 \rfloor$. Given an integer $\varphi \geq 0$ and $a \in \mathbb{R}$, we use $\lceil a \rfloor_\varphi$ to denote the closest rational number to a with denominator at most 2^φ. Given integers $a_1, \cdots, a_m$ which are not all 0, we denote $\mathsf{gcd}(a_1, \cdots, a_m)$ their greatest common divisor. Given non-zero integers $a_1, \cdots, a_m$, we denote $\mathsf{lcm}(a_1, \cdots, a_m)$ their least common multiple.

For any $i \in [n]$, we denote e_i the ith standard orthonormal basis vector of $\mathbb{R}^n$. We use $B_p(R)$ to denote the ℓ_p-ball of radius R in $\mathbb{R}^n$ and $B_p = B_p(1)$ the unit ℓ_p-ball. For any set of vectors $V \subseteq \mathbb{R}^n$, we use $\mathsf{span}\{V\}$ to denote the linear span of vectors in V. Throughout, a subspace W is a linear subspace of $\mathbb{R}^n$ with $0 \in W$; an affine subspace W is a translation of a subspace of $\mathbb{R}^n$ (and thus might not pass through the origin). Given a subspace W, we denote $W^\perp$ the orthogonal complement of W and $\Pi_W(\cdot)$ the orthogonal projection onto the subspace W. Given a PSD matrix $A \in \mathbb{R}^{n \times n}$ and a subspace $V \subseteq \mathbb{R}^n$, we say A has full rank on V if $\mathsf{rank}(A) = \dim(V)$ and the eigenvectors corresponding to non-zero eigenvalues of A form an orthogonal basis of V.

Given a subspace $V \subseteq \mathbb{R}^n$ and a PSD matrix $A \in \mathbb{R}^{n \times n}$ that has full rank on V, the function $\langle \cdot, \cdot \rangle_A$ given by $\langle x, y \rangle_A = x^\top A y$ defines an inner product on V. The inner product $\langle \cdot, \cdot \rangle_A$ induces a norm on V, i.e. $\|x\|_A = \sqrt{\langle x, x \rangle_A}$ for any $x \in V$, which we call the A-norm. Given a point $x_0 \in \mathbb{R}^n$ and a PSD matrix $A \in \mathbb{R}^{n \times n}$, we use $E(x_0, A)$ to denote the (might not be full-rank) ellipsoid given by $E(x_0, A) := \{x \in x_0 + W_A : (x - x_0)^\top A (x - x_0) \leq 1\}$, where W_A is the subspace spanned by eigenvectors corresponding to non-zero eigenvalues of A. When the ellipsoid is centered at 0, we use the short-hand notation $E(A)$ to denote $E(0, A)$.

2.3.2 Lattices

Given a set of linearly independent vectors $b_1, \cdots, b_k \in \mathbb{R}^n$, denote $\Lambda(b_1, \cdots, b_k) = \{\sum_{i=1}^k \lambda_i b_i, \lambda_i \in \mathbb{Z}\}$ the lattice generated by $b_1, \cdots, b_k$. Here, k is called the rank of the lattice. A lattice is said to have full-rank if $k = n$. Any set of k linearly independent vectors that generates the lattice $\Lambda = \Lambda(b_1, \cdots, b_k)$ under integer linear combinations is called a basis of Λ. In particular, the set $\{b_1, \cdots, b_k\}$ is a basis of Λ. Different basis of a full-rank lattice are related by unimodular matrices, which are integer matrices with determinant ± 1.

Given a basis $B \in \mathbb{R}^{n \times k}$, the fundamental parallelepiped of $\Lambda = \Lambda(B)$ is the polytope $\mathcal{P}(B) := \{\sum_{i=1}^k \lambda_i b_i : \lambda_i \in [0, 1), \forall i \in [k]\}$. The determinant of the lattice (also known as the covolume), denoted as $\det(\Lambda)$, is defined to be the volume of the fundamental parallelepiped, which is independent of the basis. We also define the notion of dual lattices below.

Definition 2.3.1 (Dual lattice)**.** *Given a lattice $\Lambda \subseteq \mathbb{R}^n$, the dual lattice Λ^* is the set of all vectors $x \in \mathsf{span}\{\Lambda\}$ such that $\langle x, y \rangle \in \mathbb{Z}$ for all $y \in \Lambda$.*

We refer interested readers to standard textbooks (e.g. [Sch98]) for a more comprehensive introduction to lattice theory.

Lattice Projection and Intersection with Subspaces. The following standard facts on lattice projection follow from Gram-Schmidt orthogonalization.

Fact 2.3.2 (Lattice projection)**.** *Let Λ be a full-rank lattice in $\mathbb{R}^n$ and W be a linear subspace such that $\dim(\mathsf{span}\{\Lambda \cap W\}) = \dim(W)$. Then we have*

$$\det(\Lambda) = \det(\Lambda \cap W) \cdot \det(\Pi_{W^\perp}(\Lambda)).$$

Fact 2.3.3 (Dual of lattice projection)**.** *Let Λ be a full-rank lattice in $\mathbb{R}^n$ and W be a linear subspace such that $\dim(\mathsf{span}\{\Lambda \cap W\}) = \dim(W)$. Then we have the following duality*

$$(\Pi_W(\Lambda))^* = \Lambda^* \cap W.$$

Minkowski's First Theorem. Minkowski's first theorem [Min53] asserts the existence of a non-zero lattice point in a symmetric convex set with large enough volume. An important consequence of it is the following upper bound on $\lambda_1(\Lambda, A)$, the length of the shortest non-zero vector in lattice Λ under A-norm.

Theorem 2.3.4 (Consequence of Minkowski's first theorem, [Min53])**.** *Let Λ be a full-rank lattice in $\mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ be a positive definite matrix. Then*

$$\lambda_1(\Lambda, A) \leq \sqrt{n} \cdot \det(A^{1/2})^{1/n} \cdot \det(\Lambda)^{1/n}.$$

The Shortest Vector Problem and the Lenstra-Lenstra-Lovász Algorithm. Given a lattice Λ and a PSD matrix A that has full rank on $\mathsf{span}\{\Lambda\}$, the Shortest Vector Problem (SVP) asks to find a shortest non-zero vector in Λ under A-norm[6], whose length is denoted as $\lambda_1(\Lambda, A)$. SVP is one of the most fundamental computational problems in lattice theory and is known to be NP-hard. For this problem, the celebrated Lenstra-Lenstra-Lovász (LLL) algorithm [LLL82] finds in polynomial time a $2^{n/2}$-approximation to $\lambda_1(\Lambda, A)$. Building on top of a block-reduction algorithm by Schnorr [Sch87], Ajtai, Kumar and Sivakumar [AKS01] obtained the current best polynomial time approximation factor of $2^{n \log\log(n)/\log(n)}$ for SVP.

Theorem 2.3.5 ([AKS01])**.** *Given a basis $b_1, \cdots, b_n \in \mathbb{Z}^n$ for lattice Λ and a positive definite matrix $A \in \mathbb{Z}^{n\times n}$. Let $D \in \mathbb{Z}$ be such that $\|b_i\|_A^2 \leq D$ for any $i \in [n]$. Then there exists an algorithm that outputs in $\mathsf{poly}(n, \log(D))$ arithmetic operations a vector b_1' such that*

$$\|b_1'\|_A \leq 2^{n \log\log(n)/\log(n)} \cdot \lambda_1(\Lambda, A).$$

Moreover, the integers occuring in the algorithm have bit sizes at most $\mathsf{poly}(n, \log(D))$.

In fact, for any integer $r > 1$, [AKS01] gave a $2^{O(r)}\mathsf{poly}(n)$-time $r^{O(n/r)}$-approximation algorithm for SVP, allowing a smooth tradeoff between time and approximation quality.

[6]Equivalently, one could think of finding an approximately shortest vector under the Euclidean norm in the lattice $A^{1/2}\Lambda$.

For solving SVP exactly, the state-of-the-art is a deterministic $\widetilde{O}(2^{2n})$-time and $\widetilde{O}(2^n)$-space algorithm given by Micciancio and Voulgaris [MV13], and a randomized $2^{n+o(n)}$-time and space algorithm due to Aggarwal et al. [ADRSD15]. We refer to these excellent papers and the references therein for a comprehensive account of the rich history of SVP.

Rational Polyhedra. We start with the definition of the LCM vertex complexity of a rational vector.

Definition 2.3.6 (LCM vertex complexity)**.** *Given a rational vector $a = (p_1/q_1, \cdots, p_n/q_n)$, where integers p_i and $q_i \geq 1$ are coprime for all $i \in [n]$, we define its LCM vertex complexity to be the smallest integer $\varphi \geq 0$ such that the 1-dimensional lattice $L_a := \{a^\top z : z \in \mathbb{Z}^n\}$ is a sub-lattice of $\mathbb{Z}/q$ for some positive integer $q \leq 2^\varphi$.*

In particular, the number q above is $\mathsf{lcm}(q_1, \cdots, q_n)$. When $\mathsf{gcd}(p_1, \cdots, p_n) = 1$, by Bézout's identity, we in fact have that $L_a = \mathbb{Z}/q$. We next formally define the notion of rational polyhedra with bounded LCM vertex complexity.

Definition 2.3.7 (Rational polyhedra with bounded LCM vertex complexity)**.** *A bounded convex set $K \subseteq \mathbb{R}^n$ is a rational polyhedron with LCM vertex complexity at most $\varphi \geq 0$ if K is a polyhedron and the LCM vertex complexity of every vertex of K is at most φ.*

For convenience, we define the set of all rational vectors with bounded LCM vertex complexity.

Definition 2.3.8 (Rational vectors with bounded LCM vertex complexity)**.** *For any integer $\varphi \geq 0$, we define S^n_φ the set of all rational vectors in $\mathbb{R}^n$ with LCM vertex complexity at most φ.*

Remark 2.3.9 (Different definitions)**.** *We remark that our definition of LCM vertex complexity in Definition 2.3.6 is different from the standard definition of vertex complexity in the literature used by Grötschel, Lovász and Schrijver [GLS88], who defined the vertex complexity of a rational vector a to be its binary description length, i.e. bit complexity. The*

LCM vertex complexity of a rational vector as in Definition 2.3.6 is always smaller than its bit complexity, and in fact might be much smaller. The reason we deviate from Grötschel, Lovász and Schrijver's more standard notion of vertex complexity is that Definition 2.3.6 allows a slightly cleaner presentation of the results and proofs in this chapter. In particular, one can obtain the results and proofs in the setting of integral minimizers by taking $\varphi = 0$.

2.3.3 Convex Geometry

A function $g : \mathbb{R}^n \to \mathbb{R}_+$ is log-concave if its support $\mathsf{supp}(g)$ is convex and $\log(g)$ is concave on $\mathsf{supp}(g)$. An integrable function $g : \mathbb{R}^n \to \mathbb{R}_+$ is a density function, if $\int_{\mathbb{R}^n} g(x)dx = 1$. The centroid of a density function $g : \mathbb{R}^n \to \mathbb{R}_+$ is defined as $\mathsf{cg}(g) = \int_{\mathbb{R}^n} g(x)xdx$; the covariance matrix of the density function g is defined as $\mathsf{Cov}(g) = \int_{\mathbb{R}^n} g(x)(x - \mathsf{cg}(g))(x - \mathsf{cg}(g))^\top dx$. A density function $g : \mathbb{R}^n \to \mathbb{R}_+$ is isotropic, if its centroid is 0 and its covariance matrix is the identity matrix, i.e. $\mathsf{cg}(g) = 0$ and $\mathsf{Cov}(g) = I$.

A typical example of a log-concave distribution is the uniform distribution over a convex body $K \subseteq \mathbb{R}^n$. Given a convex body K in $\mathbb{R}^n$, its volume is denoted as $\mathsf{vol}(K)$. The centroid (resp. covariance matrix) of K, denoted as $\mathsf{cg}(K)$ (resp. $\mathsf{Cov}(K)$), is defined to be the centroid (resp. covariance matrix) of the uniform distribution over K. A convex body K is said to be isotropic if the uniform density over it is isotropic. Any convex body can be put into its isotropic position via an affine transformation.

Sometimes we will be working with a bounded convex set $K \subseteq W$, where W is an affine subspace that might not be full dimensional. For convenience, we extend the definitions above to this case by first applying a linear transformation and then restricting to W so that K becomes full-dimensional.

Theorem 2.3.10 (Brunn's principle)**.** *Let K be a convex body and W be a subspace in $\mathbb{R}^n$. Then the function $g_{K,W} : W^\perp \to \mathbb{R}_+$ defined as $g_{K,W}(x) := \mathsf{vol}(K \cap (W + x))$ is log-concave on its support.*

Theorem 2.3.11 (Property of log-concave density, Theorem 5.14 of [LV07]). *Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an isotropic log-concave density function. Then we have $f(x) \le 2^{8n} n^{n/2}$ for every x.*

We also need the following result from [KLS95].

Theorem 2.3.12 (Ellipsoidal approximation of convex body, [KLS95]). *Let K be an isotropic convex body in $\mathbb{R}^n$. Then,*

$$\sqrt{\frac{n+1}{n}} \cdot B_2 \subseteq K \subseteq \sqrt{n(n+1)} \cdot B_2,$$

where B_2 is the unit Euclidean ball in $\mathbb{R}^n$.

The following lemma is an immediate consequence of Theorem 2.3.12.

Lemma 2.3.13 (Stability of covariance). *Let K be a convex body in $\mathbb{R}^n$ and $x \in K$ satisfies $\|x - \mathsf{cg}(K)\|_{\mathsf{Cov}(K)^{-1}} \le 0.1$. Let H be a halfspace such that $x \in H$, then we have*

$$\frac{1}{5n^2} \cdot \mathsf{Cov}(K) \preceq \mathsf{Cov}(K \cap H) \preceq n^2 \cdot \mathsf{Cov}(K).$$

Proof. Without loss of generality, we may assume that K is in isotropic position, in which case the condition that $\|x - \mathsf{cg}(K)\|_{\mathsf{Cov}(K)^{-1}} \le 0.1$ becomes $\|x\|_2 \le 0.1$. Theorem 2.3.12 then gives

$$\sqrt{\frac{n+1}{n}} \cdot B_2 \subseteq K \subseteq \sqrt{n(n+1)} \cdot B_2.$$

Let halfspace H_1 be the translation of halfspace H such that x lies on its boundary hyperplane H_1'. Note that $K \cap H_1 \subseteq K \cap H$. Let $x' := \Pi_{H_1'}(\mathsf{cg}(K))$ be the orthogonal projection of $\mathsf{cg}(K) = 0$ onto the hyperplane H_1'. Then,

$$\|x'\|_2 \le \|x - 0\|_2 \le 0.1.$$

This shows that the hyperplane H_1' is at Euclidean distance at most 0.1 from 0. It then follows that $\sqrt{\frac{n+1}{n}}B_2 \cap H_1$ contains a ball of radius at least

$$\frac{1}{2} \cdot \left(\sqrt{\frac{n+1}{n}} - 0.1 \right) \geq 0.45\sqrt{\frac{n+1}{n}} \geq \sqrt{\frac{n+1}{5n}},$$

where the last inequality uses $\sqrt{5} \times 0.45 \geq 1$. Since we have $\sqrt{\frac{n+1}{n}}B_2 \cap H_1 \subseteq K \cap H_1 \subseteq K \cap H$, this implies that $K \cap H$ contains a ball of radius $\sqrt{\frac{n+1}{5n}}$, and is contained in a ball of radius $\sqrt{n(n+1)}$. Consider the ellipsoid $E_{K\cap H} = \{y : y^\top \mathsf{Cov}(K \cap H)^{-1} y \leq 1\}$. Then Theorem 2.3.12 implies that

$$\mathsf{cg}(K \cap H) + \sqrt{\frac{n+1}{n}} \cdot E_{K\cap H} \subseteq K \cap H \subseteq \mathsf{cg}(K \cap H) + \sqrt{n(n+1)} \cdot E_{K\cap H}.$$

We thus have $\frac{1}{\sqrt{5n}} \cdot B_2 \subseteq E_{K\cap H} \subseteq n \cdot B_2$, and the statement of the lemma follows immediately. □

We note that some of these convex geometry tools have previously been used, for example, to find the densest sub-lattice in arbitary norm [DM13].

2.3.4 Cutting Plane Methods

Cutting plane methods optimize a convex function f by maintaining a convex set K that contains the minimizer of f, which gets refined iteratively using the separating hyperplanes returned by the separation oracle. One of the most classical cutting plane methods is the center of gravity method, discovered independently by Levin [Lev65] and Newman [New65].

Theorem 2.3.14 (Center of gravity method [Lev65, New65])**.** *Given a separation oracle* SO *for a convex function* f *defined on* $\mathbb{R}^n$ *with minimizers* K^**, and a convex body* $K \subseteq \mathbb{R}^n$ *containing* K^**. If* $\mathsf{cg}(K)$ *doesn't minimize* f*, then the convex body* K' *returned by CenterOfGravity*(SO, K) *above contains* K^* *and satisfies* $\mathsf{vol}(K') \leq (1 - 1/e) \cdot \mathsf{vol}(K)$.

Algorithm 1

procedure CENTEROFGRAVITY(SO, K)
  Query SO at $\mathsf{cg}(K)$
  if SO outputs "YES" **then**
    Return "YES"
  else
    Let c be the output of SO
    Return $K' := K \cap \{x : c^\top x \geq c^\top \mathsf{cg}(K)\}$
  end if
end procedure

The center of gravity method is not efficient as it involves computing the centroid of convex bodies. Using sampling techniques to estimate $\mathsf{cg}(K)$ and $\mathsf{Cov}(K)$, an efficient implementation of the center of gravity method was given in [BV04]. We start with the definition of ϵ-approximate centroid and covariance.

Definition 2.3.15 (ϵ-approximate centroid and covariance)**.** *Let $0 < \epsilon < 1$ be a parameter. Given a convex body $K \subseteq \mathbb{R}^n$, we call $x_K \in \mathbb{R}^n$ an ϵ-approximate centroid of K if $\|x_K - \mathsf{cg}(K)\|_{\mathsf{Cov}(K)^{-1}} \leq \epsilon$. We call PSD matrix $\Sigma_K \in \mathbb{R}^{n \times n}$ an ϵ-approximate covariance matrix if $(1-\epsilon) \cdot \mathsf{Cov}(K) \preceq \Sigma_K \preceq (1+\epsilon) \cdot \mathsf{Cov}(K)$.*

Constructing ϵ-approximate centroids and covariance matrices via sampling for well-rounded convex bodies appeared in the works of [KLS97, ALPTJ10, SV13]. The formulation of the following theorem is from [JLLV21, Lemma 2.5 and Theorem 2.7] together with the standard fact that the uniform distribution over a convex body is log-concave.

Theorem 2.3.16 (Approximate centroid and covariance by sampling, [KLS97, ALPTJ10, SV13])**.** *Let parameters $0 < \epsilon < 1$ and $0 < \delta < 1/2$. Given a convex body $K \subseteq \mathbb{R}^n$ specified by m constraints, a point $x \in K$ and a PSD matrix $A \in \mathbb{R}^{n \times n}$ such that the following sandwiching condition holds*

$$x + E(A) \subseteq K \subseteq x + 2^{\mathsf{poly}(n)} \cdot E(A), \tag{2.3}$$

then there is a randomized algorithm that uses $m \cdot \mathsf{poly}(n, 1/\epsilon, \log(1/\delta))$ *arithmetic operations to compute, with probability at least* $1-\delta$*, an* ϵ*-approximate centroid* x_K *and an* ϵ*-approximate covariance matrix* Σ_K *of* K*.*

Since approximate centroid and covariance matrix of a convex body give a sandwiching condition as in (2.3), [BV04] obtained the following efficient implementation of the center of gravity method. The theorem below comes from directly using Theorem 2.3.16 in the algorithmic framework of [BV04].

Theorem 2.3.17 (Approximate center of gravity method, [BV04])**.** *Let parameters* $0 < \epsilon < 0.01$ *and* $0 < \delta < 1/2$*. Given a separation oracle* SO *for a convex function* f *defined on* $\mathbb{R}^n$ *with minimizers* K^**, a polytope* K *with* m *constraints containing* K^**, an* ϵ*-approximate centroid* $x_K \notin K^*$ *and an* ϵ*-approximate covariance matrix* Σ_K *of* K*, there exists a randomized algorithm* $\mathsf{RandomWalkCG}(\mathsf{SO}, K, x_K, \Sigma_K, \epsilon, \delta)$ *that makes one call to* SO *and an extra* $m \cdot \mathsf{poly}(n, 1/\epsilon, \log(1/\delta))$ *arithmetic operations to return a polytope* K'*, a point* $x_{K'} \in K'$ *and a PSD matrix* $\Sigma_{K'}$ *such that the following hold with probability at least* $1-\delta$*:*

(a) $K^* \subseteq K'$ *and* K' *is the intersection of* K *with a constraint output by* SO *at* x_K*,*

(b) $\mathsf{vol}(K') \leq \frac{2}{3} \cdot \mathsf{vol}(K)$*,*

(c) $x_{K'}$ *is an* ϵ*-approximate centroid of* K'*, and*

(d) $\Sigma_{K'}$ *is an* ϵ*-approximate covariance matrix of* K'*.*

2.4 Technical Lemmas

In this section, we prove a few technical lemmas which are key to our result.

2.4.1 Dimension Reduction that Preserves Low-Complexity Rational Points

Recall from Definition 2.3.8 that S^n_φ is the set of rational vectors with LCM vertex complexity at most $\varphi \geq 0$.

Lemma 2.4.1 (Dimension reduction that preserves low-complexity rational points)**.** *Given an affine subspace $W = x_0 + W_0$, where W_0 is a linear subspace of $\mathbb{R}^n$ and $x_0 \in \mathbb{R}^n$ is a fixed point, and an ellipsoid $E = E(x_0, A)$ that has full rank on W. Given a vector $v \in \Pi_{W_0}(\mathbb{Z}^n) \setminus \{0\}$ with $\|v\|_{A^{-1}} < 1/2^{2\varphi+1}$, where $\varphi \geq 0$ is an integer, then there exists a hyperplane $P \not\supseteq W$ such that $E \cap S^n_\varphi \subseteq P \cap W$. In particular, let $z \in \mathbb{Z}^n$ be such that $v = \Pi_{W_0}(z)$, then P can be taken as*

$$P = \{x : v^\top x = (v - z)^\top x_0 + \lceil z^\top x_0 \rfloor_\varphi\}.$$

Proof. Clearly we have $E \cap S^n_\varphi \subseteq W$ since $E \subseteq W$. It therefore suffices to show that the hyperplane P given in the lemma statement satisfies $P \not\supseteq W$ and $E \cap S^n_\varphi \subseteq P$.

Since $v \in W_0 \setminus \{0\}$ and W_0 is a translation of W, we have $P \not\supseteq W$. If $E \cap S^n_\varphi = \emptyset$, then the lemma statement trivially holds. We may therefore assume $E \cap S^n_\varphi \neq \emptyset$ in the following. Then for any rational vectors $x_1, x_2 \in E \cap S^n_\varphi$, we have

$$\begin{aligned} |v^\top (x_1 - x_2)| &\leq \|v\|_{A^{-1}} \cdot \|x_1 - x_2\|_A \\ &< \frac{1}{2^{2\varphi+1}} \cdot (\|x_1 - x_0\|_A + \|x_2 - x_0\|_A) \leq \frac{1}{2^{2\varphi}}. \end{aligned}$$

Since $x_1, x_2 \in W \cap S^n_\varphi$, we have $x_1 - x_2 \in W_0 \cap S_{2\varphi}$. As $v = \Pi_{W_0}(z)$ where $z \in \mathbb{Z}^n$, we have

$$v^\top (x_1 - x_2) = z^\top (x_1 - x_2) \in \mathbb{Z}/q,$$

for some positive integer $q \leq 2^{2\varphi}$. It then follows that $v^\top x_1 = v^\top x_2$. Finally, we note that for any rational vector $x_1 \in E \cap S^n_\varphi$, we have

$$|z^\top (x_1 - x_0)| = |v^\top (x_1 - x_0)| \leq \|v\|_{A^{-1}} \cdot \|x_1 - x_0\|_A < \frac{1}{2^{2\varphi+1}}.$$

Since $z^\top x_1 \in \mathbb{Z}/q'$ for some $q' \le 2^\varphi$, we have $z^\top x_1 = \lceil z^\top x_0 \rfloor_\varphi$. Therefore, we have

$$v^\top x_1 = \lceil z^\top x_0 \rfloor_\varphi + (v-z)^\top x_1 = \lceil z^\top x_0 \rfloor_\varphi + (v-z)^\top x_0,$$

where the last equality is because $v - z \in W_0^\perp$ and $x_1 - x_0 \in W_0$. This finishes the proof of the lemma. □

We remark here that the rounding $\lceil \cdot \rfloor_\varphi$ in the construction of the hyperplane P can be efficiently computed using the continued fraction method (e.g. [Sch98, Corollary 6.3a]).

2.4.2 High Dimensional Slicing Lemma

Lemma 2.4.2 (High dimensional slicing lemma)**.** *Let K be a convex body and L be a full-rank lattice in $\mathbb{R}^n$. Let W be an $(n-k)$-dimensional linear subspace of $\mathbb{R}^n$ such that $\dim(L \cap W) = n-k$. Then we have*

$$\frac{\mathsf{vol}(K \cap W)}{\det(L \cap W)} \le \frac{\mathsf{vol}(K)}{\det(L)} \cdot \frac{k^{O(k)}}{\lambda_1(L^*, K)^k},$$

where L^* is the dual lattice, and $\lambda_1(L^*, K)$ is the shortest non-zero vector in L^* under the norm $\|\cdot\|_{\mathsf{Cov}(K)}$.

Proof. Note that $\mathsf{vol}(K \cap W)/\det(L \cap W)$, $\mathsf{vol}(K)/\det(L)$, and $\lambda_1(L^*, K)$ are preserved when applying the same linear transformation to K and L simultaneously. We can therefore rescale K and L such that $\mathsf{Cov}(K) = I$. We may further assume that $K \cap W \neq \emptyset$ as otherwise $\mathsf{vol}(K \cap W) = 0$ and the statement trivially holds.

We first upper bound $\mathsf{vol}(K \cap W)$ in terms of $\mathsf{vol}(K)$. To this end, we apply a translation on K to obtain K_0 such that $\mathsf{cg}(K_0) = 0$, i.e. K_0 is in isotropic position, and it suffices to upper bound the cross-sectional volume $\mathsf{vol}(K_0 \cap (W + x))$ for an arbitrary $x \in W^\perp$. By identifying $W^\perp$ with $\mathbb{R}^k$, we note that the function $f(x)$ defined as $f(x) := \mathsf{vol}(K_0 \cap (W + x))/\mathsf{vol}(K_0)$ is a log-concave density function on $\mathbb{R}^k$ by Brunn's principle (Theorem 2.3.10). Furthermore, $f(x)$ is isotropic since K_0 is in isotropic position. It thus follows from Theorem 2.3.11 that

$f(x) \le k^{O(k)}$, for any $x \in \mathbb{R}^k$. Note that $K = K_0 + \mathsf{cg}(K)$, we obtain from taking $x = -\mathsf{cg}(K)$ that

$$\frac{\mathsf{vol}(K \cap W)}{\mathsf{vol}(K)} \le k^{O(k)}. \tag{2.4}$$

We next upper bound $\det(L)$ in terms of $\det(L \cap W)$. Note that

$$\det(L) = \det(L \cap W) \cdot \det(\Pi_{W^\perp}(L)) = \frac{\det(L \cap W)}{\det(L^* \cap W^\perp)}, \tag{2.5}$$

where the first equality follows from Fact 2.3.2, and the second equality is due to Fact 2.3.3. By Minkowski's first theorem (Theorem 2.3.4), we have

$$\lambda_1(L^*) \le \lambda_1(L^* \cap W^\perp) \le \sqrt{k} \cdot (\det(L^* \cap W^\perp))^{1/k}.$$

Combine this with the earlier equation (2.5) gives

$$\det(L) \le \frac{\det(L \cap W) \cdot \sqrt{k}^k}{\lambda_1(L^*)^k} \tag{2.6}$$

It then follows from (2.4) and (2.6) that

$$\frac{\mathsf{vol}(K \cap W)}{\mathsf{vol}(K)} \cdot \frac{\det(L)}{\det(L \cap W)} \le \frac{k^{O(k)}}{\lambda_1(L^*)^k}.$$

This finishes the proof of the lemma. □

2.5 Meta Algorithm

In this section, we present a simple meta algorithm (Algorithm 2) that achieves the oracle complexity in Theorem 2.1.6. While this meta algorithm requires computing the centroids and covariance matrices of polytopes and is therefore not efficient, its oracle complexity analysis contains most of the key insights of this paper. We give an efficient (but more

complicated) implementation of this meta algorithm and prove Theorem 2.1.6 in Section 2.6.

Theorem 2.5.1 (Oracle Complexity in Theorem 2.1.6)**.** *Given a separation oracle* SO *for a convex function f defined on $\mathbb{R}^n$, and a γ-approximation algorithm* ApproxSVP *for the shortest vector problem. If the set of minimizers K^* of f is a rational polyhedron contained in a box of radius R and has LCM vertex complexity at most $\varphi \geq 0$, then there is a randomized algorithm that with high probability finds a vertex of K^* using $O(n(\varphi + \log(\gamma n R)))$ calls to* SO.

2.5.1 The Meta Algorithm

By the argument in the beginning of Section 2.2, we may assume without loss of generality that f has a unique minimizer $x^* \in S^n_\varphi$. We therefore describe our algorithm under this assumption.

Our meta algorithm maintains an affine subspace W, a polytope $K \subseteq W$ containing the rational minimizer x^* of f, and a lattice Λ. It also maintains the centroid x_K and covariance matrix Σ_K of the polytope K. In the beginning, the affine subspace $W = \mathbb{R}^n$, polytope $K = B_\infty(R)$ and lattice $\Lambda = \mathbb{Z}^n$. In each iteration of the algorithm (i.e. each while loop), the algorithm uses the γ-approximation algorithm ApproxSVP to find a short non-zero vector $v \in \Lambda$ under Σ_K-norm. If the vector v satisfies $\|v\|_{\Sigma_K} \geq \frac{1}{10n2^{2\varphi}}$, then the algorithm runs the center of gravity method (Theorem 2.3.14) for one more step, and updates x_K and Σ_K to be the centroid and covariance matrix of the new polytope K. We remark that the criterion for performing the cutting plane step comes from the convex geometry fact that $K \subseteq x_K + 2n \cdot E(\Sigma_K^{-1})$ (Theorem 2.3.12).

If, on the other hand, that $\|v\|_{\Sigma_K} < \frac{1}{10n2^{2\varphi}}$, then the algorithm uses Lemma 2.4.1 to find a hyperplane P that contains $K \cap S^n_\varphi$, where we recall from Definition 2.3.8 that S^n_φ is the set of all rational vectors in $\mathbb{R}^n$ with LCM vertex complexity at most φ. Specifically, the hyperplane $P = \{x : v^\top x = (v - z)^\top x_K + \lceil z^\top x_K \rfloor_\varphi\}$ for some integral vector $z \in \mathbb{Z}^n$ such that $v = \Pi_{W_0}(z)$ and $W_0 = -x_K + W$ is the translation of W that passes through

the origin. One may find such a vector $z \in \mathbb{Z}^n$ efficiently by solving the closest vector problem $\min_{z\in\mathbb{Z}^n} \|z - v\|_{P_{W_0}}$, where P_{W_0} is the projection matrix onto the subspace W_0. As mentioned earlier, the rounding $\lceil\cdot\rfloor_\varphi$ can also be performed efficiently using the continued fraction method. After constructing the hyperplane P, the algorithm then recurses on the lower-dimensional affine subspace $W \cap P$, updates K to be $K \cap P$, and updates x_K and Σ_K to be the centroid and covariance matrix of the new polytope $K \cap P$. The algorithm obtains a new lattice with rank reduced by one by projecting the current lattice Λ onto P_0, a translation of P that passes through the origin.

The above procedure stops when $\dim(W) = 0$, in which case K contains a unique rational point x^* which will be the output of the algorithm. Note that when $\dim(W) = 1$, the algorithm reduces to a binary search on the segment $K \subseteq W$. A formal description of the algorithm is given in Algorithm 2.

We remark that Algorithm 2 is not efficient since it requires the computation of the centroid and covariance matrix in Line 8 and 13. Line 8 can easily be made efficient using the approximate center of gravity method as in Theorem 2.3.17. However, it is not clear how to efficiently implement Line 13 since we do not know an ellipsoid satisfying condition (2.3) in Theorem 2.3.16, and thus approximate centroid and covariance matrix might not be efficiently computable by sampling. We address this computational issue in the next section.

2.5.2 Oracle Complexity Analysis

We start by proving the correctness of Algorithm 2.

Lemma 2.5.2 (Correctness of MetaALG)**.** *Assuming the conditions in Theorem 2.5.1 and that f has a unique minimizer $x^* \in S^n_\varphi$, Algorithm 2 finds x^*.*

Proof. Note that in the beginning of each iteration, we have $K \subseteq W$ and $\Lambda \subseteq W_0$, where W_0 is the translation of W that passes through the origin. We first argue that the lattice Λ is in fact the orthogonal projection of $\mathbb{Z}^n$ onto the subspace W_0, i.e. $\Lambda = \Pi_{W_0}(\mathbb{Z}^n)$. This is

Algorithm 2

procedure MetaALG(SO, R, φ)
  Affine subspace $W \leftarrow \mathbb{R}^n$, polytope $K \leftarrow B_\infty(R)$, lattice $\Lambda \leftarrow \mathbb{Z}^n$
  Centroid $x_K \leftarrow \mathsf{cg}(K)$, covariance matrix $\Sigma_K \leftarrow \mathsf{Cov}(K)$ ▷ $x_K + E(\Sigma_K^{-1})/2 \subseteq K \subseteq x_K + 2n \cdot E(\Sigma_K^{-1})$
  while $\dim(W) > 0$ **do**
    $v \leftarrow \text{ApproxSVP}(\Lambda, \Sigma_K)$ ▷ $v \in \Lambda \setminus \{0\}$
    if $\|v\|_{\Sigma_K} \geq \frac{1}{10n2^{2\varphi}}$ **then**
      $K \leftarrow \text{CenterOfGravity}(\mathsf{SO}, K)$
      $x_K \leftarrow \mathsf{cg}(K)$, $\Sigma_K \leftarrow \mathsf{Cov}(K)$
    else
      Find $z \in \mathbb{Z}^n$ such that $v = \Pi_{W_0}(z)$ ▷ Subspace $W_0 = -x_K + W$
      Construct $P \leftarrow \{y : v^\top y = (v - z)^\top x_K + \lceil z^\top x_K \rfloor_\varphi\}$
      $W \leftarrow W \cap P$, $K \leftarrow K \cap P$ ▷ Dimension reduction
      $x_K \leftarrow \mathsf{cg}(K)$, $\Sigma_K \leftarrow \mathsf{Cov}(K)$
      Construct hyperplane $P_0 \leftarrow \{y : v^\top y = 0\}$
      $\Lambda \leftarrow \Pi_{P_0}(\Lambda)$ ▷ Lattice projection
    end if
  end while
  Return unique point $x^* \in K$
end procedure

required for Lemma 2.4.1 to be applicable. Clearly $\Lambda = \Pi_{W_0}(Z)$ holds in the beginning of the algorithm since $\Lambda = \mathbb{Z}^n$ and $W = \mathbb{R}^n$. Notice that the CenterOfGravity procedure in Line 7 keeps Λ and W the same. Each time we reduce the dimension in Line 11-15, we have

$$\Pi_{W_0 \cap P_0}(\mathbb{Z}^n) = \Pi_{W_0 \cap P_0}(\Pi_{W_0}(\mathbb{Z}^n)) = \Pi_{W_0 \cap P_0}(\Lambda),$$

where the first equality follows because $W_0 \cap P_0$ is a subspace of W_0. Since $\Pi_{P_0}(\Lambda) = \Pi_{W_0 \cap P_0}(\Lambda)$ as $v \in W_0$, this shows that the invariant $\Lambda = \Pi_{W_0}(\mathbb{Z}^n)$ holds throughout the algorithm.

We now prove that Algorithm 2 finds the unique minimizer $x^* \in S^n_\varphi$. Note that in the beginning of the algorithm, we have $x^* \in K$. Since CenterOfGravity in Line 7 always preserves $x^* \in K$, we only need to prove that dimension reduction in Line 11-15 preserves $x^* \in K$. In the following, we show the stronger statement that each dimension reduction iteration in Line 11-15 preserves all rational points in $K \cap S^n_\varphi$.

Since Algorithm 2 maintains $x_K = \mathsf{cg}(K)$ and $\Sigma_K = \mathsf{Cov}(K)$ in every iteration, an immediate application of Theorem 2.3.12 gives the following sandwiching condition:

$$x_K + E(\Sigma_K^{-1})/2 \subseteq K \subseteq x_K + 2n \cdot E(\Sigma_K^{-1}). \tag{2.7}$$

Now we proceed to show that each dimension reduction iteration preserves all rational points in $K \cap S^n_\varphi$. By the RHS of (2.7), we have $K \cap S^n_\varphi \subseteq (x_K + 2n \cdot E(\Sigma_K^{-1})) \cap S^n_\varphi$. Since $\|v\|_{\Sigma_K} < \frac{1}{10n2^{2\varphi}}$ is satisfied in a dimension reduction iteration, Lemma 2.4.1 shows that all rational points in $(x_K + 2n \cdot E(\Sigma_K^{-1})) \cap S^n_\varphi$ lie on the hyperplane given by $P = \{y : v^\top y = (v - z)^\top x_K + \lceil z^\top x_K \rfloor_\varphi\}$. Thus we have $K \cap S^n_\varphi \subseteq K \cap P$ and this finishes the proof of the lemma. □

Next, we prove the oracle complexity upper bound of Algorithm 2 in Theorem 2.5.1.

Lemma 2.5.3 (Oracle complexity of MetaALG)**.** *Assuming the conditions in Theorem 2.5.1*

and that f has a unique minimizer $x^* \in S^n_\varphi$, Algorithm 2 makes at most $O(n(\varphi + \log(\gamma n R)))$ calls to SO.

Proof. We note that the oracle is only called when CenterOfGravity is invoked in Line 7, and each run of CenterOfGravity makes one call to SO according to Theorem 2.3.14. To upper bound the total number of runs of CenterOfGravity, we consider the potential function

$$\Phi = \log(\mathsf{vol}(K) \cdot \det(\Lambda)).$$

In the beginning, $\Phi = \log(\mathsf{vol}(B_\infty(R)) \cdot \det(I)) = n \log(R)$. Each time CenterOfGravity is called in Line 7, we have from Theorem 2.3.14 that the volume of K decreases by at least a constant factor, so the potential function decreases by at least $\Omega(1)$ additively.

To analyze the change in the potential function after dimension reduction, we consider a maximal sequence of consecutive dimension reduction iterations $t_0 + 1, \cdots, t_0 + k$, i.e. CenterOfGravity is invoked in iteration t_0 and $t_0 + k + 1$, while every iteration in $t_0 + 1, \cdots, t_0 + k$ decreases the dimension by one. We shall use superscript (i) to denote the corresponding notations in the beginning of iteration $t_0 + i$, for any integer $i \geq 0$. In particular, in the beginning of iteration $t_0 + 1$, we have a convex body $K^{(1)} \subseteq K^{(0)} \subseteq W^{(0)} = W^{(1)}$, and after the sequence of dimension reduction iterations, we reach a convex body $K^{(k+1)} = K^{(1)} \cap W^{(k+1)} \subseteq K^{(0)} \cap W^{(k+1)}$. The lattice changes from $\Lambda^{(0)} = \Lambda^{(1)} \subseteq W_0^{(1)}$ to $\Lambda^{(k+1)} = \Pi_{W_0^{(k+1)}}(\Lambda^{(1)}) = \Pi_{W_0^{(k+1)}}(\Lambda^{(0)})$, where we recall that subspaces $W_0^{(i)}$ are translations of the affine subspaces $W^{(i)}$ that pass through the origin. Note that the potential at the beginning of this maximal sequence of dimension reduction iterations is

$$e^{\Phi^{(0)}} = \mathsf{vol}(K^{(0)}) \cdot \det(\Lambda^{(0)}) = \frac{\mathsf{vol}(K^{(0)})}{\det((\Lambda^{(0)})^*)}.$$

The potential after this sequence of dimension reduction iterations is

$$\begin{aligned} e^{\Phi^{(k+1)}} &= \mathsf{vol}(K^{(k+1)}) \cdot \det(\Lambda^{(k+1)}) = \mathsf{vol}(K^{(1)} \cap W^{(k+1)}) \cdot \det(\Pi_{W_0^{(k+1)}}(\Lambda^{(0)})) \\ &= \frac{\mathsf{vol}(K^{(1)} \cap W^{(k+1)})}{\det((\Pi_{W_0^{(k+1)}}(\Lambda^{(0)}))^*)} = \frac{\mathsf{vol}(K^{(1)} \cap W^{(k+1)})}{\det((\Lambda^{(0)})^* \cap W_0^{(k+1)})} \leq \frac{\mathsf{vol}(K^{(0)} \cap W^{(k+1)})}{\det((\Lambda^{(0)})^* \cap W_0^{(k+1)})}, \end{aligned}$$

where the last equality follows from the duality $(\Pi_{W_0^{(k+1)}}(\Lambda^{(0)}))^* = (\Lambda^{(0)})^* \cap W_0^{(k+1)}$ in Fact 2.3.3. Since $W^{(k+1)}$ is a translation of the subspace $W_0^{(k+1)}$, we can apply Lemma 2.4.2 by taking $L = (\Lambda^{(0)})^*$ to obtain

$$e^{\Phi^{(k+1)}} \leq e^{\Phi^{(0)}} \cdot \frac{k^{O(k)}}{\lambda_1(\Lambda^{(0)}, K^{(0)})^k}, \tag{2.8}$$

where $\lambda_1(\Lambda^{(0)}, K^{(0)})$ is the shortest non-zero vector in $\Lambda^{(0)}$ under the norm $\|\cdot\|_{\mathsf{Cov}(K^{(0)})}$. As CenterOfGravity is invoked in iteration t_0, we have $\left\|v^{(0)}\right\|_{\Sigma_K^{(0)}} \geq \frac{1}{10n2^{2\varphi}}$ for the output vector $v^{(0)} \in \Lambda^{(0)} \setminus \{0\}$. Since the ApproxSVP procedure is γ-approximation and that $\Sigma_K^{(0)} = \mathsf{Cov}(K^{(0)})$, this implies that $\lambda_1(\Lambda^{(0)}, K^{(0)}) \geq \frac{\Omega(1)}{\gamma n 2^{2\varphi}}$. It then follows that

$$e^{\Phi^{(k+1)}} \leq e^{\Phi^{(0)}} \cdot (\gamma n 2^{\varphi})^{O(k)}.$$

This shows that after a sequence of k dimension reduction iterations, the potential increases additively by at most $O(k \log(\gamma n 2^{\varphi}))$. As there are at most n dimension reduction iterations, the total amount of potential increase due to dimension reduction iterations is thus at most $O(n \log(\gamma n 2^{\varphi}))$.

Finally we note that whenever the potential becomes smaller than $-10n \log(20n\gamma 2^{2\varphi})$, Minkowski's first theorem (Theorem 2.3.4) shows the existence of a non-zero vector $v \in \Lambda$ with $\|v\|_{\Sigma_K} < \frac{1}{20n\gamma 2^{2\varphi}}$. This implies that the γ-approximation algorithm ApproxSVP for the shortest vector problem will find a non-zero vector $v' \in \Lambda$ that satisfies $\|v'\|_{\Sigma_K} < \frac{1}{20n2^{2\varphi}}$, and thus such an iteration will not invoke CenterOfGravity. Therefore, Algorithm 2

runs CenterOfGravity at most $O(n\log(\gamma n 2^{\varphi}) + n\log(R)) = O(n(\varphi + \log(\gamma n R)))$ times. Since each run of CenterOfGravity makes one call to SO, the total number of calls to SO made by Algorithm 2 is thus $O(n(\varphi + \log(\gamma n R)))$. This finishes the proof of the lemma. □

Proof of Theorem 2.5.1. By the argument in the beginning of Section 2.2, we may assume without loss of generality that f has a unique minimizer $x^* \in S^n_\varphi$. The correctness of Algorithm 2 is given in Lemma 2.5.2, and its oracle complexity is upper bounded in Lemma 2.5.3. These finish the proof of the theorem. □

2.6 Efficient Implementation of the Meta Algorithm

In this section, we give an efficient implementation of Algorithm 2 from the previous section and prove Theorem 2.1.6 which we restate below for convenience.

Theorem 2.1.6 (Main result for rational polyhedra)**.** *Given a separation oracle* SO *for a convex function* f *defined on* $\mathbb{R}^n$*, and a* γ*-approximation algorithm* ApproxSVP *for the shortest vector problem which takes* T_{SVP} *arithmetic operations. If the set of minimizers* K^* *of* f *is a rational polyhedron contained in a box of radius* R *and has LCM vertex complexity at most* $\varphi \geq 0$*, then there is a randomized algorithm that with high probability finds a vertex of* K^* *using* $O(n(\varphi + \log(\gamma n R)))$ *calls to* SO *and* $\mathsf{poly}(n, \varphi, \log(\gamma R)) \cdot T_{\mathrm{SVP}}$ *arithmetic operations.*

2.6.1 The Efficient Implementation

By the argument in the beginning of Section 2.2, we may assume without loss of generality that f has a unique minimizer $x^* \in S^n_\varphi$. For simplicity, we present our algorithm under this assumption.

As mentioned in the last paragraph of Section 2.5.1, we can efficiently implement Line 8 of Algorithm 2 by using the approximate center of gravity method in Theorem 2.3.17. We now address the issue of efficiently implementing Line 13 of Algorithm 2 in the following.

To obtain an approximate centroid and covariance matrix of the polytope K after dimension reduction, our efficient algorithm maintains two polytopes $K_{\mathsf{SO}} \subseteq K_{\mathsf{free}}$. The polytope

K_{SO} plays the same role as K in Algorithm 2, and is the polytope formed by the separating hyperplanes from SO. And K_{free} is a simple polytope for which we always know an approximate centroid x_K and covariance matrix Σ_K. Our algorithm explicitly maintains the lists of constraints for the polytopes K_{SO} and K_{free} to efficiently perform computations on them. In particular, our algorithm can efficiently certify[7] that $K_{\mathsf{free}} = K_{\mathsf{SO}}$ when all the constraints for K_{SO} appear in the list of constraints for K_{free}, since it is always maintained that $K_{\mathsf{SO}} \subseteq K_{\mathsf{free}}$.

In the beginning of the algorithm, $K_{\mathsf{free}} = K_{\mathsf{SO}}$ and we run RandomWalkCG for both polytopes at the same time. When dimension reduction happens in Line 16-21, K_{SO} is updated to be $K_{\mathsf{SO}}^{\mathsf{new}} = K_{\mathsf{SO}} \cap P$ and we no longer have approximations to $\mathsf{cg}(K_{\mathsf{SO}}^{\mathsf{new}})$ and $\mathsf{Cov}(K_{\mathsf{SO}}^{\mathsf{new}})$. To bypass this difficulty, our strategy is to update K_{free} to be a simple polytope $K_{\mathsf{free}}^{\mathsf{new}}$ containing $K_{\mathsf{SO}}^{\mathsf{new}}$ for which we know $\mathsf{cg}(K_{\mathsf{free}}^{\mathsf{new}})$ and $\mathsf{Cov}(K_{\mathsf{free}}^{\mathsf{new}})$, and "learn" $\mathsf{cg}(K_{\mathsf{SO}}^{\mathsf{new}})$ and $\mathsf{Cov}(K_{\mathsf{SO}}^{\mathsf{new}})$ by shrinking $K_{\mathsf{free}}^{\mathsf{new}}$ via RandomWalkCG until it coincides with $K_{\mathsf{SO}}^{\mathsf{new}}$. Whenever $K_{\mathsf{free}}^{\mathsf{new}} = K_{\mathsf{SO}}^{\mathsf{new}}$ happens again (in the aforementioned sense that the constraints for $K_{\mathsf{SO}}^{\mathsf{new}}$ all appear in the list of constraints $K_{\mathsf{free}}^{\mathsf{new}}$), we have successfully learned an approximate centroid and covariance matrix of $K_{\mathsf{SO}}^{\mathsf{new}}$, and can continue to shrink $K_{\mathsf{SO}}^{\mathsf{new}}$ using RandomWalkCG as before.

Now we specify our choice of $K_{\mathsf{free}}^{\mathsf{new}}$ in the strategy above. Note that $K_{\mathsf{SO}}^{\mathsf{new}} \subseteq P \cap (x_K + 2n \cdot E(\Sigma_K^{-1}))$. Denoting the ellipsoid $P \cap (x_K + 2n \cdot E(\Sigma_K^{-1})) = E(w, A)$, we can simply choose $K_{\mathsf{free}}^{\mathsf{new}}$ to be the smallest hyperrectangle containing $E(w, A)$, i.e. $K_{\mathsf{free}}^{\mathsf{new}} = w + A^{-1/2} B_\infty$, for which it is easy to compute an exact centroid and covariance matrix.

Such choice of $K_{\mathsf{free}}^{\mathsf{new}}$ blows up the volume of the outer ellipsoid $P \cap (x_K + 2n \cdot E(\Sigma_K^{-1}))$ by a factor of $n^{O(n)}$, and thus shrinking $K_{\mathsf{free}}^{\mathsf{new}}$ seems to require much more SO calls. The crucial observation here is that when we shrink the volume of $K_{\mathsf{free}}^{\mathsf{new}}$, we do not need to make calls to SO since we already know the polytope $K_{\mathsf{SO}}^{\mathsf{new}} \subseteq K_{\mathsf{free}}^{\mathsf{new}}$. Instead, we simulate the separation oracle using the smaller polytope $K_{\mathsf{SO}}^{\mathsf{new}}$ via the procedure FreeCG (see Algorithm 4) until

[7] In general, our algorithm might not be able to efficiently verify that the geometric objects K_{free} being the same as K_{SO}. So whenever we say $K_{\mathsf{free}} = K_{\mathsf{SO}}$, we always mean it in the sense that it can be efficiently certified by checking that all constraints for K_{SO} appear in the list of constraints for K_{free}.

we have $K_{\mathsf{free}}^{\mathsf{new}} = K_{\mathsf{SO}}^{\mathsf{new}}$ again, at which point we regain approximations to $\mathsf{cg}(K_{\mathsf{SO}}^{\mathsf{new}})$ and $\mathsf{Cov}(K_{\mathsf{SO}}^{\mathsf{new}})$. If we are ever able to find a hyperplane P^{new} containing $K_{\mathsf{free}}^{\mathsf{new}} \cap S_\varphi^n$ even before reaching the point $K_{\mathsf{free}}^{\mathsf{new}} = K_{\mathsf{SO}}^{\mathsf{new}}$, we can further reduce the dimension. A formal description of the efficient implementation is given in Algorithm 3.

Algorithm 3

procedure MAIN(SO, R, φ)
  Affine subspace $W \leftarrow \mathbb{R}^n$, lattice $\Lambda \leftarrow \mathbb{Z}^n$
  Polytopes $(K_{\mathsf{free}}, K_{\mathsf{SO}}) \leftarrow (B_\infty(R), B_\infty(R))$ ▷ Maintain constraints explicitly for K_{free} and K_{SO}
  $x_K \leftarrow \mathsf{cg}(K_{\mathsf{free}})$ and $\Sigma_K \leftarrow \mathsf{Cov}(K_{\mathsf{free}})$ ▷ $x_K + E(\Sigma_K^{-1})/2 \subseteq K_{\mathsf{free}} \subseteq x_K + 2n \cdot E(\Sigma_K^{-1})$
  $\epsilon \leftarrow 0.01$, $\delta \leftarrow 1/\mathsf{poly}(n, \varphi, \log(\gamma R))$ ▷ Parameters in Theorem 2.3.17
  while $\dim(W) > 0$ **do**
    $v \leftarrow$ APPROXSVP(Λ, Σ_K) ▷ $v \in \Lambda \setminus \{0\}$
    if $\|v\|_{\Sigma_K} \geq \frac{1}{10n2^{2\varphi}}$ **then**
      if $K_{\mathsf{free}} = K_{\mathsf{SO}}$ **then** ▷ List of constraints for K_{free} include that of K_{SO}
        $(K', x_{K'}, \Sigma_{K'}) \leftarrow$ RANDOMWALKCG$(\mathsf{SO}, K_{\mathsf{free}}, x_K, \Sigma_K, \epsilon, \delta)$ as in Theorem 2.3.17
        $(K_{\mathsf{free}}, K_{\mathsf{SO}}) \leftarrow (K', K')$, $x_K \leftarrow x_{K'}$, $\Sigma_K \leftarrow \Sigma_{K'}$
      else
        $(K_{\mathsf{free}}, x_K, \Sigma_K) \leftarrow$ FREECG$(K_{\mathsf{free}}, K_{\mathsf{SO}}, x_K, \Sigma_K)$ ▷ No SO call in this step
      end if
    else
      Find $z \in \mathbb{Z}^n$ such that $v = \Pi_{W_0}(z)$ ▷ Subspace $W_0 = -x_K + W$
      Hyperplane $P \leftarrow \{y : v^\top y = (v - z)^\top x_K + \lceil z^\top x_K \rfloor_\varphi\}$
      $W \leftarrow W \cap P$, $K_{\mathsf{SO}} \leftarrow K_{\mathsf{SO}} \cap P$ ▷ Dimension reduction
      $K_{\mathsf{free}} \leftarrow w + A^{-1/2} B_\infty$ ▷ Ellipsoid $E(w, A) := P \cap (x_K + 2n \cdot E(\Sigma_K^{-1}))$
      $x_K \leftarrow \mathsf{cg}(K_{\mathsf{free}})$, $\Sigma_K \leftarrow \mathsf{Cov}(K_{\mathsf{free}})$
      Hyperplane $P_0 \leftarrow \{y : v^\top y = 0\}$, lattice $\Lambda \leftarrow \Pi_{P_0}(\Lambda)$ ▷ Lattice projection
    end if
  end while
  Return unique point $x^* \in K_{\mathsf{SO}}$
end procedure

Algorithm 4

procedure FREECG($K_{\mathsf{free}}, K_{\mathsf{SO}}, x_K, \Sigma_K$)
  if $x_K \notin K_{\mathsf{SO}}$ **then** ▷ Check the constraints for K_{SO}
    Find constraint $a^\top x \le b$ of K_{SO} violated by x_K
    $H \leftarrow \{x : a^\top x \le a^\top x_K\}$ ▷ x_K lies on the boundary of H
    $K'_{\mathsf{free}} \leftarrow K_{\mathsf{free}} \cap H$ ▷ Volume of K_{free} shrinks
    Obtain ϵ-approx. centroid $x_{K'}$ and cov. $\Sigma_{K'}$ of K'_{free} as in Theorem 2.3.17
  else
    Find any constraint $H = \{x : a^\top x \le b\}$ of K_{SO} that is not a constraint of K_{free} ▷ $x_K \in H$
    $K'_{\mathsf{free}} \leftarrow K_{\mathsf{free}} \cap H$ ▷ K_{free} learns one more constraint of K_{SO}
    Obtain ϵ-approx. centroid $x_{K'}$ and cov. $\Sigma_{K'}$ of K'_{free} as in Theorem 2.3.16 ▷ Validity by Lemma 2.3.13
  end if
  Return $K'_{\mathsf{free}}, x_{K'}, \Sigma_{K'}$
end procedure

2.6.2 Proof of Main Result

By the argument in the beginning of Section 2.2, we can assume wlog that f has a unique minimizer $x^* \in S^n_\varphi$. We first prove the correctness and oracle complexity of Algorithm 3. These proofs are very similar to the proofs of Lemma 2.5.2 and 2.5.3 from the previous section, so we only highlight the differences.

Lemma 2.6.1 (Correctness of MAIN)**.** *Assuming the conditions in Theorem 2.1.6 and that f has a unique minimizer $x^* \in S^n_\varphi$, Algorithm 3 finds x^*.*

Proof. As in the proof of Lemma 2.5.2, we only need to verify that $x^* \in K_{\mathsf{SO}}$ is preserved under dimension reduction in Line 16-21. Let's assume that $x^* \in K_{\mathsf{SO}}$ before dimension reduction. Since Theorem 2.3.17 guarantees $\|x_K - \mathsf{cg}(K_{\mathsf{free}})\|_{(\Sigma_K)^{-1}} \le \epsilon$ and $(1-\epsilon)\cdot \mathsf{Cov}(K_{\mathsf{free}}) \preceq \Sigma_K \preceq (1+\epsilon)\cdot \mathsf{Cov}(K_{\mathsf{free}})$ with $\epsilon = 0.01$, it follows from Theorem 2.3.12 that (2.7) still holds with K replaced by K_{free}:

$$x_K + E(\Sigma_K^{-1})/2 \subseteq K_{\mathsf{free}} \subseteq x_K + 2n \cdot E(\Sigma_K^{-1}).$$

Proceeding from here, the same argument as in the proof of Lemma 2.5.2 shows that $K_{\mathsf{free}} \cap S_\varphi^n \subseteq P$. Also note that Algorithm 3 always maintains $K_{\mathsf{SO}} \subseteq K_{\mathsf{free}}$. It follows that $K_{\mathsf{SO}} \cap S_\varphi^n \subseteq K_{\mathsf{free}} \cap S_\varphi^n \subseteq P$, i.e. all rational points in $K_{\mathsf{SO}} \cap S_\varphi^n$ are preserved during dimension reduction. This implies that $x^* \in K_{\mathsf{SO}} \cap P$ after dimension reduction and completes the proof of the lemma. □

Lemma 2.6.2 (Oracle complexity of MAIN)**.** *Assuming the conditions in Theorem 2.1.6 and that f has a unique minimizer $x^* \in S_\varphi^n$, Algorithm 3 makes at most $O(n(\varphi + \log(\gamma n R)))$ calls to the separation oracle SO with high probability.*

Proof. Note that Algorithm 3 always maintains $K_{\mathsf{SO}} \subseteq K_{\mathsf{free}}$, and SO is only called in Line 10 when $K_{\mathsf{SO}} = K_{\mathsf{free}}$. Since each run of RANDOMWALKCG in Line 10 succeeds with probability $\delta = 1/\mathsf{poly}(n, \varphi, \log(\gamma R))$ for a large enough polynomial by Theorem 2.3.17, union bound implies that with high probability, the first $O(n(\varphi + \log(\gamma n R)))$ run of RANDOMWALKCG in Line 10 all succeed. We condition on this event. Then applying exactly the same analysis as in the proof of Lemma 2.5.3 to the potential function

$$\Phi_{\mathsf{SO}} := \log(\mathsf{vol}(K_{\mathsf{SO}}) \cdot \det(\Lambda))$$

gives the oracle complexity bound in the lemma. □

Next, we show that Algorithm 3 makes at most $\mathsf{poly}(n, \varphi, \log(\gamma R))$ calls to FREECG with high probability. Since each call to FREECG can be implemented in $\mathsf{poly}(n, \varphi, \log(\gamma R))$ time by checking all the constraints of K_{SO}, this will imply the bound on the number of arithmetic operations in Theorem 2.1.6.

Lemma 2.6.3 (Number of FREECG calls)**.** *Assuming the conditions in Theorem 2.1.6 and that f has a unique minimizer $x^* \in S_\varphi^n$, Algorithm 3 makes at most $\mathsf{poly}(n, \varphi, \log(\gamma R))$ calls to* FREECG *with high probability.*

Proof. As in the proof above, we condition on the high probability event that the first $\mathsf{poly}(n, \varphi, \log(\gamma R))$ calls to RANDOMWALKCG as well as the sampling algorithm in Theo-

rem 2.3.16 all succeed. In the beginning of the algorithm, $K_{\mathsf{SO}} = B_\infty(R)$ and thus can be specified using $2n$ constaints. An additional constraint is placed on K_{SO} each time SO is called, and since the number of SO calls is at most $O(n(\varphi + \log(\gamma n R))$, the number of constraints Algorithm 3 maintains for the specification of K_{SO} can be at most $O(n(\varphi+\log(\gamma n R))$ throughout.

Now we upper bound the number of calls to FreeCG. In fact, we show that the total number of cutting plane steps for K_{free} in Line 10 and 13 of Algorithm 3 is at most $\mathsf{poly}(n, \varphi, \log(\gamma R))$. Our strategy is to consider the potential function

$$\Phi_{\mathsf{free}} := \log(\mathsf{vol}(K_{\mathsf{free}}) \cdot \det(\Lambda)),$$

and repeat the analysis as in the proof of Lemma 2.5.3. However, there are two main differences that we highlight below.

The first main difference is that when we reduce the dimension in Line 16-21 of Algorithm 3, we are not simply slicing K_{free} by the hyperplane P. Instead, we first replace K_{free} by its outer containing ellipsoid $x_K + 2n \cdot E(\Sigma_K^{-1})$, then further replace the sliced ellipsoid $E(w, A) = P \cap (x_K + 2n \cdot E(\Sigma_K^{-1}))$ by its outer containing hyperrectangle $K_{\mathsf{free}}^{\mathsf{new}} := w + A^{-1/2} B_\infty$. Since we have the sandwiching condition that

$$x_K + E(\Sigma_K^{-1})/2 \subseteq K_{\mathsf{free}} \subseteq x_K + 2n \cdot E(\Sigma_K^{-1}),$$

replacing K_{free} by $x_K + 2n \cdot E(\Sigma_K^{-1})$ increases its volume by at most $n^{O(n)}$. Also note that replacing an ellipsoid by its outer containing hyperrectangle increases its volume by at most $n^{O(n)}$. It then follows that these replacements contribute to at most a factor of $n^{O(n)}$ to $\mathsf{vol}(K_{\mathsf{free}})$ for each dimension reduction step. As there are at most n dimension reduction steps, the increase in Φ_{free} due to these replacements is at most $O(n^2 \log(n))$ additively.

The second main difference is that not every call to FreeCG decreases $\mathsf{vol}(K_{\mathsf{free}})$ by a constant factor. In particular, this is the case if $x_K \in K_{\mathsf{SO}}$ in Algorithm 4 and we add to

K_{free} one constraint of K_{SO} that is currently not a constraint of K_{free}. However, since we have shown above that K_{SO} has at most $O(n(\varphi + \log(\gamma nR)))$ constraints, this case can happen at most $O(n(\varphi+\log(\gamma nR)))$ in each dimension until all the constraints for K_{SO} appear in the list of constraints for K_{free}, in which case our algorithm can efficiently certify that $K_{\mathsf{free}} = K_{\mathsf{SO}}$. Whenever this happens, no additional call to FreeCG will happen until the dimension is further reduced.

Incorporating the above two differences into the analysis as in the proof of Lemma 2.5.3, we obtain that the total number of cutting plane steps in Line 10 and 13 applied to K_{free} is at most $O(n^2(\varphi + \log(\gamma nR)))$. This is also an upper bound on the number of calls to FreeCG, and thus proves the lemma. □

Proof of Theorem 2.1.6. By the argument in the beginning of Section 2.2, we may assume without loss of generality that f has a unique minimizer $x^* \in S^n_\varphi$. The correctness of Algorithm 3 is given in Lemma 2.6.1, and its oracle complexity is upper bounded in Lemma 2.6.2. We are thus left to upper bound the total number of arithmetic operations used by Algorithm 3.

By Lemma 2.6.3, Algorithm 3 makes at most $\mathsf{poly}(n, \varphi, \log(\gamma R))$ calls to FreeCG and each such step can be implemented using $\mathsf{poly}(n, \varphi, \log(\gamma R))$ arithmetic operations. Since ApproxSVP is called after each cutting plane step in Line 10 and 13, the total number of calls to ApproxSVP is at most $\mathsf{poly}(n, \varphi, \log(\gamma R))$. Note that the remaining part of the algorithm takes $\mathsf{poly}(n, \varphi, \log(\gamma R))$ arithmetic operations. This gives the upper bound on the number of arithmetic operations and finishes the proof of the theorem. □

2.7 Submodular Function Minimization

In this section, we do not seek to give a comprehensive introduction to submodular functions, but only provide the necessary definitions and properties that are needed for the proof of Theorem 2.1.7. We refer interested readers to the famous textbook by Schrijver [Sch03] or the extensive survey by McCormick [McC05] for more details on submodular functions.

2.7.1 Preliminaries

Throughout this section, we use $[n] = \{1, \cdots, n\}$ to denote the ground set and let $f : 2^{[n]} \to \mathbb{Z}$ be a set function defined on subsets of $[n]$. For a subset $S \subseteq [n]$ and an element $i \in [n]$, we define $S + i := S \cup \{i\}$. A set function f is *submodular* if it satisfies the following property of *diminishing marginal differences*:

Definition 2.7.1 (Submodularity)**.** *A function $f : 2^{[n]} \to \mathbb{Z}$ is submodular if $f(T + i) - f(T) \leq f(S + i) - f(S)$, for any subsets $S \subseteq T \subseteq [n]$ and $i \in [n] \setminus T$.*

Throughout this section, the set function f we work with is assumed to be submodular even when it is not stated explicitly. We may assume without loss of generality that $f(\emptyset) = 0$ by replacing $f(S)$ by $f(S) - f(\emptyset)$. We assume that f is accessed by an *evaluation oracle*, and use EO to denote the time to compute $f(S)$ for a subset S. Our algorithm for SFM is based on a standard convex relaxation of a submodular function, known as the Lovász extension [GLS88].

Definition 2.7.2 (Lovász extension)**.** *The Lovász extension $\hat{f} : [0, 1]^n \to \mathbb{R}$ of a submodular function f is defined as*

$$\hat{f}(x) = \mathbb{E}_{t \sim [0,1]}[f(\{i : x_i \geq t\})],$$

where $t \sim [0, 1]$ is drawn uniformly at random from $[0, 1]$.

The Lovász extension $\hat{f}$ of a submodular function f has many desirable properties. In particular, $\hat{f}$ is a convex relaxation of f and it can be evaluated efficiently.

Theorem 2.7.3 (Properties of Lovász extension)**.** *Let $f : 2^{[n]} \to \mathbb{Z}$ be a submodular function and $\hat{f}$ be its Lovász extension. Then,*

(a) $\hat{f}$ is convex and $\min_{x \in [0,1]^n} \hat{f}(x) = \min_{S \subseteq [n]} f(S)$;

(b) $f(S) = \hat{f}(I_S)$ for any subset $S \subseteq [n]$, where I_S is the indicator vector for S;

(c) Suppose $x \in [0,1]^n$ *satisfies* $x_1 \geq \cdots \geq x_n$*, then* $\hat{f}(x) = \sum_{i=1}^{n}(f([i]) - f([i-1]))x_i$*;*

(d) The set of minimizers of $\hat{f}$ *is the convex hull of the set of minimizers of* f*.*

Next we address the question of implementing the separation oracle (as in Definition 2.1.1) using the evaluation oracle of f.

Theorem 2.7.4 (Separation oracle for Lovász extension, Theorem 61 of [LSW15])**.** *Let* $f : 2^{[n]} \to \mathbb{Z}$ *be a submodular function and* $\hat{f}$ *be its Lovász extension, then a separation oracle for* $\hat{f}$ *can be implemented in time* $O(n \cdot \mathsf{EO} + n^2)$*.*

2.7.2 Proof of Theorem 2.1.7

Before presenting the proof, we restate Theorem 2.1.7 for convenience.

Theorem 2.1.7 (Submodular function minimization)**.** *Given an evaluation oracle* EO *for a submodular function* f *defined over subsets of an* n*-element ground set, there exist*

(a) a strongly polynomial algorithm that minimizes f *using* $O(n^3 \log\log(n)/\log(n))$ *calls to* EO*, and*

(b) an exponential time algorithm that minimizes f *using* $O(n^2 \log(n))$ *calls to* EO*.*

Proof. We apply Corollary 2.1.3 to the Lovász extension $\hat{f}$ of the submodular function f with $R = 1$. By part (a) and (d) of Theorem 2.7.3, $\hat{f}$ is a convex function that satisfies the assumption $(\star)$ in Corollary 2.1.3 Thus Corollary 2.1.3 gives a strongly polynomial algorithm for finding an integral minimizer of $\hat{f}$ that makes $O(n^2 \log\log(n)/\log(n))$ calls to a separation oracle of $\hat{f}$, and an exponential time algorithm that finds an integral minimizer of $\hat{f}$ using $O(n\log(n))$ separation oracle calls. This integral minimizer also gives a minimizer of f. Since a separation oracle for $\hat{f}$ can be implemented using $O(n)$ calls to EO by Theorem 2.7.4, the total number of calls to the evaluation oracle is thus $O(n^3 \log\log(n)/\log(n))$ for the strongly polynomial algorithm, and is $O(n^2\log(n))$ for the exponential time algorithm. This proves the theorem. □

Chapter 3

LOWER BOUNDS FOR SUBMODULAR FUNCTION MINIMIZATION

In this chapter, we present improved lower bounds for the query complexity and parallel complexity of Submodular Function Minimization. In particular, the parallel complexity lower bound in this chapter matches the upper bound given by the algorithm in Chapter 2 up to poly-logarithmic factors. This chapter is based on my joint work with Deeparnab Chakrabarty, Andrei Graur, and Aaron Sidford [CGJS22] which appeared in the *63rd IEEE Symposium on Foundations of Computer Science (FOCS 2022)*.

3.1 Introduction

A real-valued function $f : 2^V \to \mathbb{R}$ defined on subsets of an n-element ground set V is *submodular* if $f(X \cup \{e\}) - f(X) \geq f(Y \cup \{e\}) - f(Y)$ for any $X \subseteq Y \subseteq V$ and $e \in V \setminus Y$. Submodular functions are ubiquitous and include cut functions in (hyper-)graphs, set coverage functions, rank functions of matroids, utility functions in economics, and entropy functions in information theory, etc.

Given the expressive power of submodular functions, the optimization of these functions has been extensively studied. The problem of submodular function minimization (SFM), i.e. $\min_{S \subseteq V} f(S)$, given black-box access to an *evaluation oracle*, which returns the value $f(S)$ upon receiving a set $S \subseteq V$, encompasses many important problems in theoretical computer science, operations research, game theory, and more. Recently, SFM has found applications in computer vision, machine learning, and speech recognition [BVZ01, KKT08, KT10, LB11]. Correspondingly, SFM has been the subject of extensive research for decades and is foundational to the theory of combinatorial optimization.

Throughout the chapter, unless specified otherwise, we focus on the *strongly-polynomial* regime for the *query complexity* of SFM. We refer to an SFM algorithm as strongly-polynomial (in terms of query complexity) if the number of evaluation oracle queries it makes is at most a polynomial in n and does not depend on the range of the function. After decades of advances [GLS81, Cun85, GLS88, Sch00, FI00, IFF01, Iwa03, Vyg03, Orl09, IO09], the current state-of-the-art strongly-polynomial algorithms include an $O(n^2 \log n)$-query, $\exp(O(n))$-time algorithm [Jia22] and an $O(n^3 \log\log n / \log n)$-query, $\mathsf{poly}(n)$-time algorithm [Jia22], which improved (in query complexity) upon $\tilde{O}(n^3)$-query, $\tilde{O}(n^4)$-time algorithms of [LSW15, JLSW20, DVZ21].[1]

Despite the rich history of SFM research, obtaining *lower bounds* on the query complexity for SFM has been notoriously difficult. [Har08] described two different constructions of submodular functions whose minimization requires n-queries to an evaluation oracle; in fact, both can be minimized by querying all the n singletons. Later, [CLSW17] showed that one of the examples in [Har08] also needs $n/4$ gradient queries to the Lovász extension of the submodular function. This remained the best lower bound, until recently [GPRW20] proved a $2n$-query lower bound on SFM via a non-trivial construction of a submodular function (which can be minimized in $2n$ queries). For more discussions on difficulties in obtaining super-linear lower bounds, we refer the reader to Section 3.1.3.

More recently, there has been an interest in understanding the *parallel complexity* of SFM. Note that any SFM algorithm proceeds by making queries to an evaluation oracle in rounds, and the parallel complexity of SFM is the minimum number of rounds (also known as the depth) required by any *query-efficient* SFM algorithm that makes at most $\mathsf{poly}(n)$ evaluation oracle queries. All SFM algorithms described above proceed in $\Omega(n)$-rounds. The best known round-complexity is the algorithm due to [Jia22] which runs in $O(n \log n)$ rounds. On the lower bound side, [BS20] proved that any query-efficient SFM algorithm must proceed in $\Omega(\log n / \log\log n)$-rounds. This was improved in [CCK21] to an $\tilde{\Omega}(n^{1/3})$-lower bound on the

[1] Throughout, we use $\tilde{O}(\cdot)$ to hide polylogarithmic factors.

number of rounds for query-efficient SFM. The latter work also mentioned a bottleneck of $n^{1/3}$ to their approach and left open the question of whether a nearly-linear number of rounds are needed, or whether there is a query-efficient SFM algorithm proceeding in $n^{1-\delta}$ many rounds for some absolute constant $\delta > 0$.

3.1.1 Our Results.

In this chapter we provide improved lower bounds for both the query complexity for SFM, and the round complexity for query-efficient parallel SFM. We prove that any deterministic SFM algorithm requires $\Omega(n \log n)$ queries to an evaluation oracle, and that any parallel SFM algorithm making at most $\mathsf{poly}(n)$ queries must proceed in $\Omega(n/\log n)$ rounds.

Theorem 3.1.1 (Query complexity lower bound for deterministic algorithms)**.** *For any finite set V with n elements and deterministic SFM algorithm ALG, there exists a submodular function $F : 2^V \to \mathbb{R}$ such that ALG makes at least $\frac{n}{2}\log_2(\frac{n}{4})$ evaluation oracle queries to minimize F.*

Theorem 3.1.1 constitutes the first super-linear lower bound on the number of evaluation queries for SFM. The previous best lower bound was $2n$, due to [GPRW20].

Theorem 3.1.2 (Parallel lower bound for randomized algorithms)**.** *For any finite set V with n elements, constant $C \geq 2$, and (possibly randomized) parallel SFM algorithm ALG that makes at most $Q := n^C$ queries per round, there exists a submodular function $F : 2^V \to \mathbb{R}$ such that ALG takes at least $\frac{n}{2C\log_2 n}$ rounds to minimize F with high probability.*

Theorem 3.1.2 improves upon the previous best parallel lower bound of $\tilde{\Omega}(n^{1/3})$ due to [CCK21]. Furthermore, Theorem 3.1.2 is optimal up to logarithmic factors due to [Jia22], which yields an $O(n \log n)$-round, $O(\mathsf{poly}(n))$-queries algorithm.[2].

[2]This query bound is due to the fact that an algorithm in [Jia22] solves SFM with $O(n \log n)$ computations of the subgradients of the Lovász extension. Further, each computation of a subgradient can be implemented by making n queries to an evaluation oracle for the submodular function in parallel, i.e. a single round.

Both Theorem 3.1.1 and Theorem 3.1.2 are obtained by constructing a new family of submodular functions. This family of submodular functions and the analysis of their properties is our main technical contribution. At a high level, we glue together simple submodular functions, each of which is defined on a distinct part of a large partition of the ground set V and has a unique minimizer. The main novelty of our construction is an approach to assemble these functions into a layered structure in such a way that any SFM algorithm needs to effectively find the minimizer of one layer before obtaining any information about the functions in later layers. This forces any parallel algorithm to have depth equal to the number of parts, which implies our parallel lower bound. We also show that minimizing a single part needs a number of queries super-linear in the size of that part, implying the super-linear query complexity lower bound for deterministic algorithms. More insights into our construction and proofs are given in Section 3.1.2.

3.1.2 Our Techniques

Previous works on proving lower bounds for parallel SFM [BS20, CCK21] apply the following generic framework. At a high level, they design a family of hard submodular functions which are parameterized using a partition $(P_1, \ldots, P_\ell)$ of the ground set. The key property they show is that even after obtaining answers to polynomially many queries in round i, any algorithm (with high probability) doesn't possess any information about the elements in $P_{i+1}, \ldots, P_\ell$. Further, the construction also has the property that knowing which elements are in the final part P_ℓ is crucial in obtaining the minimizer. These properties prove an $\ell - 1$ lower bound on the number of rounds for parallel SFM.

This chapter also proceeds under the same generic framework, but departs crucially from prior work in the design of the family of hard submodular functions $\mathcal{F}$, which is the main technical innovation of this chapter. With this new construction, our query complexity lower bound follows by a careful adversarial choice of function $F \in \mathcal{F}$, and our parallel round complexity lower bound follows by choosing a random function uniformly at random from $\mathcal{F}$.

Recap of Previous Constructions. Before we dive into a high-level discussion of our construction, here we remind the reader of the construction ideas in [BS20] and [CCK21], and why they stop short of proving a nearly-linear lower bound on the number of rounds for parallel SFM. Both these works construct so-called *partition submodular functions* F where one is given a partition $(P_1, \ldots, P_\ell)$, and the value of $F(S)$ depends only on the *cardinality* of the sets $|S \cap P_1|, \ldots, |S \cap P_\ell|$. Note that when the algorithm has no information about $P_1, \ldots, P_\ell$, for instance in the first round of querying, then for any query set S, these cardinalities are roughly proportional to the cardinalities of each part. The main idea behind the constructions in [CCK21, BS20] is to come up with submodular functions where this "roughly proportional" property is used to hide any information about the parts $P_2, \ldots, P_\ell$. However, the fact that $|S \cap P_i|$'s can typically differ by a standard deviation necessarily requires each part P_i to be "sufficiently large" and this, in turn, puts a $o(n)$ bottleneck on the *number* of parts ℓ. As it stands, it is not clear how to obtain a better than $n^{1/3}$-lower bound on the round complexity of parallel SFM using partition submodular functions.

Interestingly, a similar approach as above has also been the main tool to prove lower bounds for parallel convex optimization [Nem94, BS18, BJL+19, DG19]. We defer to Section 3.1.3 for a more detailed discussion of this broader context.

Ideas Behind our Construction. Our construction deviates from the notion of partition submodular functions in that the function value $F(S)$ crucially depends on the *identity* of the set $S \cap P_i$ rather than the size, which helps us bypass the bottleneck in previous constructions and obtain nearly-linear lower bound on the number of rounds.

It is convenient to think of the family of functions we construct in a recursive fashion. Pick a subset $A \subseteq V$ of size $2r$, which corresponds to the first part P_1 in the partition described above, and denote $B := V \setminus A$ the remainder parts $P_2 \cup \cdots \cup P_\ell$. For notational convenience, we denote $S_A := S \cap A$ and $S_B := S \cap B$ for any set $S \subseteq V$. Let $R \subseteq A$ be a subset of size $|R| = r = |A|/2$, and consider the following function $F : 2^V \to \mathbb{R}$ defined as

$$F(S) := h_R(S) + \beta \cdot \mathbf{1}(S_A = R) \cdot g(S_B), \qquad \text{(Meta Definition)}$$

where $\mathbf{1}(\cdot)$ is the indicator function, and g is a submodular function which will recursively be the same as F defined over the smaller universe B. The parameter β is a small scalar, and should be thought of as $\Theta(\frac{1}{|V|})$. We aim to design the function $h_R(\cdot)$ to have the following two properties:

(P1) Any set $S \subseteq V$ is a minimizer of h_R if and only if $S_A = R$,

(P2) The function F defined in (Meta Definition) is submodular whenever g is submodular.

We now claim that obtaining such a function h_R suffices to prove an $\frac{n}{2C \log n}$-lower bound on the number of rounds required by any exact parallel SFM algorithm making $\leq n^C$ queries per round. In particular, the subsets $R \subseteq A \subseteq V$ with $|R| = |A|/2 = C \log n$, as well as the recursively defined function g, will be chosen uniformly at random.

To see this, first observe that when β is sufficiently small, if S_g^* is a (unique) minimizer of the function g, then the set $S^* := R \cup S_g^*$ is a (unique) minimizer of F. This crucially uses property (P1) which says that $R \cup S_B$ is a minimizer of h_R for any $S_B \subseteq B$. Next, consider the first round of queries $Q^1, \ldots, Q^T$. Since $R \subseteq A$ is chosen uniformly at random, and because $|R| = |A|/2 = C \log n$, the probability that one of these $Q_A^i = R$ is negligible if $T \leq n^C$. Therefore, all the answers to the queries in the first round are precisely $h_R(Q_i)$, revealing no information about the function g. On the other hand, the minimizer of F needs to minimize g. Therefore, if we pick g randomly from the same family of F but over the smaller universe B, we could apply the above argument recursively with $2C \log n$ fewer elements and one fewer round. In this way, we prove an $\frac{n}{2C \log n}$-lower bound on the number of rounds needed to exactly minimize the random submodular function F.

The big question left, of course, is whether one can construct a function h_R with the properties mentioned above. This is what we discuss next.

Obtaining Submodularity. Let us first discuss an idea that does not work and then fix it. One way to define h_R is to take a submodular function f_R defined *only* over elements of A, whose (unique) minimizer is the subset R, and then extend it as $h_R(S) := f_R(S_A)$. In

particular,

$$F(S) := f_R(S_A) + \beta \cdot \mathbf{1}(S_A = R) \cdot g(S_B). \qquad \text{(First Try)}$$

Note that it satisfies property (P1), i.e. S is a minimizer of h_R if and only if $S_A = R$. Unfortunately, the resulting function F may not be submodular even if both f_R and g are submodular. To see this, consider an element $e \in B$ and consider the marginal increase in F when e is added to a set S. Since f_R only depends on S_A and $e \in B$, in the marginal calculation of $F(S+e) - F(S)$, the f_R terms cancel out. In particular, we get that

$$F(S+e) - F(S) = \beta \cdot \mathbf{1}(S_A = R) \cdot (g(S_B + e) - g(S_B)) .$$

Suppose the parenthesized term is positive for some S_B (e.g. the maximal minimizer of g) and consider the sets $S := R \cup S_B$ and $S' := R' \cup S_B$, where R' is any strict subset of R. In this case $F(S+e) - F(S) > 0$ while $F(S'+e) - F(S') = 0$ and since $S' \subseteq S$, this violates submodularity.

To fix the above idea, we pad the function $f_R(S_A)$ with what we call a "*submodularizer function*" $\phi(S)$. Think of ϕ as taking two sets (S_A, S_B) as input; the first set is a subset of A the other is a subset of B. We define $h_R(S) := f_R(S_A) + \phi(S_A, S_B)$ and therefore,

$$F(S) := f_R(S_A) + \phi(S_A, S_B) + \beta \cdot \mathbf{1}(S_A = R) \cdot g(S_B). \qquad \text{(Layered Function)}$$

What properties do we need from ϕ? First, since (P1) requires that when $S_A = R$, the set S is a minimizer of $f + \phi$ irrespective of what S_B is, this suggests $\phi(R, S_B)$ is the same for any $S_B \subseteq B$. For simplicity, assume this is 0. That is, when $S_A = R$, the ϕ function doesn't have any effect. However, considering the reason our first attempt failed, when S'_A is a *strict* subset of R, then $\phi(S'_A, S_B)$ should be so defined such that adding an element $e \in B$ to S_B *strictly increases* the function value. This would make sure that $F(S'+e) - F(S') > 0$ for the violating example in the previous paragraph. Not only that, this strict increase should be *greater* than the increase in $F(S+e) - F(S)$, where $S = (R, S_B)$ is as in the previous

paragraph, and this increase is β times some marginal of g. To ensure that this occurs, we choose β to be "small enough"; it suffices to choose a constant factor less than the strict increase of the function ϕ. A similar argument also leads us to the conclusion that when S_A is a *strict superset* of R, then $\phi(S_A, S_B)$ should *strictly decrease* in value when an element is added to S_B. A definition of ϕ that works is the following:

$$\phi(S_A, S_B) := \begin{cases} +4\beta|S_B| & \text{if } S_A \text{ strict subset of } R \\ -4\beta|S_B| & \text{if } S_A \text{ strict superset of } R \\ 0 & \text{otherwise, and in particular if } S_A = R \end{cases} \qquad \text{(Submodularizer)}$$

Note we still have the parameter β unspecified, and we set it soon.

The above discussion only considered marginals of an element $e \in B$ to the function F. One also needs to be careful about the case when the element $e \in A$. This will put a restriction on what f_R and β are, and will form the last part of our informal description.

Consider an element $e \in A \backslash R$ and consider the function $\phi(R, S_B)$ for an arbitrary $S_B \subseteq B$. Note that, as defined, the value of $\phi(R, S_B) = 0$ and $\phi(R+e, S_B) = -4\beta|S_B|$. That is, adding e to $R \cup S_B$ can *decrease* the ϕ function value by $-4\beta|S_B|$. On the other hand, adding e to $(A - e) \cup S_B$ doesn't change the ϕ-value. Indeed, $\phi(A, S_B) = \phi(A - e, S_B) = -4\beta|S_B|$ since both A and $A - e$ are strict supersets of R (remember $e \notin R$). In short, the function ϕ is *not* submodular and this endangers the submodularity of the sum function $h_R = f_R + \phi$.

To fix this, we *make sure* that the function f_R has a "large gap" between $f_R(R + e)$ and $f_R(R)$. In particular, we ensure that $f_R(R + e) - f_R(R) = \Omega(1)$ while $\beta = O(1/n)$. In this way, although adding $e \in A \setminus R$ to (R, S_B) can decrease the ϕ value by $-4\beta|S_B|$, since $\beta = O(1/n)$ this decrease is smaller than the increase caused by $f_R(R + e) - f_R(R)$ when the constants are properly chosen. In particular, we define the function f_R on the universe

A as follows

$$f_R(S_A) := \begin{cases} 0 & \text{if } S_A = R \\ 1 & \text{if } S_A \text{ is a strict superset or a strict subset of } R \\ 2 & \text{otherwise} \end{cases} \tag{3.1}$$

It is not too hard to see that this function f_R is submodular; in fact, this function (or a scaled version if it) has been considered before in the submodular function literature [Har08, CLSW17]. This completes the informal description and motivation of our construction of hard functions; a formal presentation of our construction and the full proof of its properties can be found in Section 3.3 and Section 3.5.

Query Complexity Lower Bound. While discussed and motivated in terms of the number of parallel rounds for SFM, our construction can also prove an $\Omega(n \log n)$ lower bound on the *query complexity* of any *deterministic* SFM algorithm. Indeed, for this part, we consider the family where the size of $|A| = 2$, and R is a singleton among these two elements. Instead of selecting a random function from this family, we adversarially choose a worst-case function depending on the deterministic algorithm. Note that the function definition above doesn't require the size $|A|$ to be large; we made it large in the previous discussion since we were ruling out polynomial query parallel algorithms.

The main observation is the strong property that until the algorithm queries a set S with $S_A = R$, it obtains no information about the function g. Therefore, if we can prove a lower bound $L(n, r)$ on the number of oracle queries any algorithm needs to find such a set, with r being the size of R, then we can obtain an $\Omega(\frac{n}{r} \cdot L(n, r))$ lower bound on the exact SFM query complexity.

It is actually not too hard to prove $L(n, 2) \geq \lfloor \log_2 n \rfloor - 1$ for any *deterministic* algorithm. Note that R is a singleton element, and we overload notation and call that element R as well. First, note that for any query S, if $S_A \neq R$, then the value of $F(S)$ only reveals whether S contains "both" the elements of A, "none" of the elements of A, or the "other"

element in A that is not R; in the first case, the ϕ-function is negative, the second case it is positive and the last case it is 0. The lower bound can now be proved using an *adversary* argument against the deterministic algorithm, by choosing the function so that the oracle never answers "other." Since the algorithm is deterministic, the adversary can choose the set A depending on the queries. The adversary maintains an "active universe" U which initially contains all the elements. If the first query S contains $\leq |U|/2$ active elements, then the adversary puts both elements of A in $V \setminus S$, answers "none", and removes $U \cap S$ from U; if S contains $> |U|/2$ active elements then the adversary puts both elements in S, answers "both", and removes $U \setminus S$ from U. The algorithm can never reach the desired set until the number of active elements goes below 2. Since the number of active elements can at best be halved each time, this proves a $\log_2 n - 1$ lower bound on the number of queries. Together with our construction, we obtain an $\Omega(n \log n)$ lower bound on the query complexity of any deterministic SFM algorithm. This is the first super-linear lower bound for this question.

Limitations and Open Questions. We end this overview section by pointing out some limitations of our construction; we believe bypassing them would require new ideas. The first issue is the *range* of our submodular functions. Our current way of constructing the submodularizer ϕ in (Submodularizer) requires that the range of ϕ be distinctly smaller than the marginal increase in the f_R function. This is noted by the parameter β which is set to $\Theta(1/n)$. If there are $\ell = n/2r$ parts to the function, then due to the recursive nature of our construction, the smallest non-zero value our function takes is as small as $O(\frac{1}{n^\ell})$. When $\ell = \Theta(n/\log n)$, as is the case in our lower bound for parallel SFM, this is $2^{-\Theta(n)}$. Put differently, if we scale the function such that the range is integers, then our function's range takes exponentially large integer values. Therefore, our lower bounds are more properly interpreted in the *strongly polynomial* regime where the round/query-complexity needs to be independent of the range of the submodular function. In contrast, the submodular functions constructed in [CCK21] which proves an $\tilde{\Omega}(n^{1/3})$ lower bound on the number of rounds have range $\{-n, -n+1, \ldots, n-1, n\}$, and thus also constitute a lower bound in the *weakly polynomial* regime (its definition is deferred to Section 3.1.3). Interestingly, the lower bound

construction in [BS20] also has a large range; it remains an interesting open problem to prove a nearly-linear lower bound on the number of rounds for query-efficient parallel SFM for integer-valued submodular functions with $\mathsf{poly}(n)$-bounded range.

We prove an $\Omega(n \log n)$ lower bound for the query complexity of deterministic algorithms for SFM. Improving this to an n^{1+c}-lower bound for some constant $c > 0$ is an important open question. The collection of functions we construct can be minimized in $\tilde{O}(n)$ queries, and so one may need new ideas to obtain a truly super-linear lower bound. The main idea behind this algorithm is that in (Layered Function), an element of R can be recognized in $\mathsf{polylog}(n)$ queries using a binary-search style idea. Basically, given any set S the function value $F(S)$ gives the information whether S_A is a subset/superset of R (in which case it also gives the size $|S_A|$), or it tells if S_A is neither a subset or superset of R. With some work this leads to an $\tilde{O}(r)$ query algorithm to find R (here r is the size of R), and thus in $n/2r$ rounds with a total query complexity of $\tilde{O}(n)$ one minimizes F.

A final limitation is that we fall short of proving an $\Omega(n \log n)$ query lower bound for *randomized* SFM algorithms. Indeed, if one looks at the structure of our $\Omega(n \log n)$ proof, the "$\log n$" arises from $L(n, 2)$ which is a lower bound on the number of queries a deterministic algorithm needs to make to find a set S such that $S_A = R$. With randomization, this problem is trivially solved in $O(1)$ queries; a random set that contains each element with probability $1/2$ would do. One may wonder if $r = |R|$ was increased, whether a super-linear in r lower bound could be proved for $L(n, r)$. Unfortunately this is not possible; there is a randomized algorithm which finds a set S with $S_A = R$ in expected $O(r)$ queries. We leave proving a super-linear lower bound on the query complexity of randomized algorithms for SFM as an open question. The family we construct is a potential candidate for the lower bound, just that a new technique would be needed to show this.

3.1.3 Further Related Work

Other Regimes for SFM. Apart from the strongly-polynomial regime, there have also been multiple recent improvements to the complexity of SFM in other regimes that depend

on M, the range of the function, i.e. $\max_{S \subseteq V} |f(S)|$ when f is scaled to have an integer range. In particular, we refer to an algorithm as *weakly-polynomial* if the number of evaluation oracle queries it makes is polynomial in n and $\log M$, and *pseudo-polynomial* if the number of queries is a polynomial in n and M. State-of-the-art weakly-polynomial algorithms include $\tilde{O}(n^2 \log M)$-query, $O(n^3 \cdot \mathsf{poly}(n, M))$-time algorithms [LSW15, JLSW20], and state-of-the-art pseudo-polynomial algorithms include $\tilde{O}(n \cdot \mathsf{poly}(M))$-query, $\tilde{O}(n \cdot \mathsf{poly}(M))$-time algorithms [CLSW17, ALS20].

Query Lower Bounds and Cuts. As far as the query complexity of SFM is concerned, lower bounds have been stagnating at $\Omega(n)$. The first known lower bound, of n queries, is due to [Har08]. Motivated the problem of improving the lower bound, [RSW18] considered graph cut functions, which is a subclass of submodular functions, and the problem of computing a global minimum cut in a graph using cut queries. However, they instead showed an upper bound of $\tilde{O}(n)$ queries to find a (non-trivial) global minimum cut in an undirected, unweighted graph. [GPRW20] improve the lower bound for SFM to $2n$ using an adversarial input technique, and also introduce a novel concept, called the graph cut dimension, for proving lower bounds for the min-cut settings. The main insight is that the cut dimension of a graph, defined as the dimension of the span of all vectors representing minimum cuts (binary vectors in R^E), is a lower bound on the number of cut queries needed. However, [LLSZ21] has shown that the cut dimension of an unweighted graph is at most $2n - 3$, essentially eliminating the hope for a super-linear lower bound using this measure. Further, the recent work of [AEG+22] provides a randomized algorithm that makes $O(n)$ queries and computes the global minimum cut in an undirected, unweighted graph with probability $2/3$.

Parallel Convex Optimization. As far as parallel lower bounds are concerned, the general framework described in Section 3.1.2 and employed in [BS20, CCK21] is similar in spirit to the approach taken in [Nem94] to bound parallel non-smooth *convex* optimization. More precisely, [Nem94] considers the problem of minimizing a non-smooth convex function f (rescaled to be have range $[-1, +1]$) up to ε-additive error in an ℓ_∞-ball, where one has

access to first-order oracle and can make $\mathsf{poly}(n)$ queries to it in each round. [Nem94] shows that any query-efficient algorithm with parallel depth $\tilde{O}(n^c \log(1/\varepsilon))$ must have $c \geq 1/3$.

The proof relies on the idea of partitioning the universe V into $r = \tilde{\Omega}(n^{1/3} \log(1/\epsilon))$ parts, and considering functions f that are the maximum of functions f_i defined on these partitions.

[BJL+19] uses a similar framework to show that any query-efficient algorithm achieving parallel depth $\tilde{O}(n^c \log(1/\varepsilon))$ must have $c \geq 1/2$. [Nem94] hypothesises that such algorithms must have $c \geq 1$, but this is still open. The problem has also been studied [DBW12, BS18, DG19, BJL+19] when the dependence on $1/\varepsilon$ is allowed to be a polynomial, and we refer the interested reader to these works for more details.

Approximate SFM. Since the Lovász extension of a submodular function is a non-smooth convex function, the discussion in the above paragraph is related to understanding the parallel complexity of ε-*approximate* SFM. In this problem, we assume by scaling that the range of the function is in $[-1, +1]$ and the objective is to obtain an additive ε-approximation to the minimum value. The construction in [CCK21] shows that any query-efficient ε-approximate SFM algorithm with depth $\tilde{O}(n^c \log(1/\varepsilon))$ must have $c \geq 1/3$. Note the similarity with the lower bound in [Nem94] mentioned in the previous paragraph; this is not an accident since the bottlenecks due to standard deviation considerations are similar in both approaches. A reader may wonder if the constructions in this chapter also prove that any query-efficient ε-approximate SFM algorithm with depth $\tilde{O}(n^c \log(1/\varepsilon))$ must have $c \geq 1$. This is not the case; the functions we consider can be ε-approximated in $O(\log(1/\varepsilon))$-rounds. This stems from the limitation in our construction that the "scale" of the functions we consider across the layers decay geometrically, and thus one can get ε-close in $O(\log(1/\varepsilon))$-rounds.

The ε-approximate SFM question is also interesting when the dependence of the depth on $1/\varepsilon$ is allowed to be a polynomial. In this setting, one can leverage the parallel convex optimization works mentioned in the previous paragraph to obtain query-efficient ε-approximate SFM algorithms with depth being truly sub-linear in n. For instance, the algorithm in [BJL+19] implies a query-efficient ε-approximate SFM algorithm running in $\tilde{O}(n^{2/3}\varepsilon^{-2/3})$-rounds. On the other hand, the construction in [CCK21] shows that any query-efficient

ε-approximate SFM algorithm with depth $(1/\varepsilon)^c$ must have $c \geq 1$. Understanding the correct answer for query-efficient ε-approximate SFM, both when the dependence on ε is $\mathsf{poly}(1/\varepsilon)$ and when it is $\log(1/\varepsilon)$, is an interesting open question.

3.2 Preliminaries

Throughout, log denotes logarithm with base 2. For any two sets X and Y, we use $X \subseteq Y$ to denote that X is a subset of Y with possibly $X = Y$; we use $X \subsetneq Y$ to denote that X is a strict subset of Y, i.e. $X \subseteq Y$ and there exists at least one element $e \in Y$ such that $e \notin X$. Further, supersets, $\supseteq$, and strict supersets, $\supsetneq$, are defined analogously.

For any set X and element $e \notin X$, we let $X + e$ denote the set obtained by including e into X, i.e. $X \cup \{e\}$. Given two sets X and Y, we define $Y \setminus X = \{e \in Y : e \notin X\}$ to denote the set of elements in Y but not in X.

Definition 3.2.1 (Marginals)**.** *Let $f : 2^V \to \mathbb{R}$ for finite set V. For any $X \subsetneq V$ and $e \in V \setminus X$, we define $\partial_e f(X) := f(X + e) - f(X)$, the marginal of f at X when adding element e.*

Definition 3.2.2 (Submodular functions)**.** *A set function $f : 2^V \to \mathbb{R}$ for finite set V is* submodular *if $\partial_e f(Y) \leq \partial_e f(X)$, for any subsets $X \subseteq Y \subsetneq V$ and $e \in [n] \setminus Y$. An alternative definition is that for any two subsets $X, Y \subseteq V$, the following inequality holds*

$$f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y). \tag{3.2}$$

3.3 Our Construction

In this section, we describe our recursive construction of the family of non-negative functions $\mathcal{F}_r(V)$ on subsets of a given set of elements V, where $r \in \mathbb{Z}_+$ is an integer such that $2r$ divides $|V|$. We prove that any function $F \in \mathcal{F}_r(V)$ is submodular and its unique minimizer takes a special partition structure which is crucial to our proofs of lower bounds in Section 3.4.

We define the main building block behind our construction in Section 3.3.1, and use it to recursively construct the function family $\mathcal{F}_r(V)$ in Section 3.3.2.

3.3.1 Main Building Block

We start by describing the main building block for our construction, which relies on two components. The first component is a standard submodular function corresponding to the sum of the rank functions of two rank-1 matroids [Har08, CLSW17]. The second component is a "submodularizer" function ϕ. Despite not being submodular itself, this submodularizer function guarantees the submodularity of our main building block function.

Component I: Sum of Two Rank-1 Matroids. For any sets $R \subseteq A$, we define the function $f_{A,R} : 2^A \to \mathbb{R}$ as

$$f_{A,R}(S) := \begin{cases} 0 & \text{if } S = R, \\ 1 & \text{if } S \subsetneq R \text{ or } S \supsetneq R, \\ 2 & \text{otherwise.} \end{cases} \tag{3.3}$$

As noted in [Har08], the function $f_{A,R}$ above corresponds to the matroid intersection of two rank-1 matroids, and is therefore submodular.

Lemma 3.3.1 ([Har08])**.** *For any $R \subseteq A$, the function $f_{A,R} : 2^A \to \mathbb{R}$ defined above is submodular.*

In fact, the submodular function $f_{A,R}$ (appropriately scaled) has previously been used in [Har08] to prove an n lower bound on the number of evaluation oracle calls, and in [CLSW17] to show an $n/4$ lower bound on the number of sub-gradients of the Lovász extension for SFM.

Component II: The Submodularizer. Let $R \subseteq A \subseteq V$ be subsets of the ground set V, and denote $B := V \setminus A$. For any subset $S \subseteq V$, we denote $S_A := S \cap A$ and $S_B := S \cap B$.

Ideally, we would like to recursively define a function on V to be of the form $f_{A,R}(S_A) + \mathbf{1}(S_A = R) \cdot g(S_B)$, where $g : 2^B \to \mathbb{R}$ is a submodular function on B. However, as men-

tioned in Section 3.1.2, such a function may not be submodular even when both $f_{R,A}$ and g are submodular. For our recursive construction to go through, we define the following submodularizer function: $\phi_{V,A,R} : 2^V \to \mathbb{R}$ as

$$\phi_{V,A,R}(S) := \begin{cases} |S_B| & \text{if } S_A \subsetneq R, \\ -|S_B| & \text{if } S_A \supsetneq R, \\ 0 & \text{otherwise, and in particular when } S_A = R. \end{cases} \tag{3.4}$$

Note that the function $\phi_{V,A,R}$ defined above is not submodular, as witnessed by the following violation of the marginal property in Definition 3.2.2. To see this, let $X \subseteq Y \subseteq V$ be any two subsets such that $X_A = R$, $A \neq Y_A \supsetneq X_A$, and $X_B \neq \emptyset$. Note that Y_A is a strict superset of X_A. Pick an element $e \in A \setminus Y_A$. Then observe that $\partial_e \phi_{V,A,R}(X) = -|X_B| < 0$ since $\phi_{V,A,R}(X \cup e) = -|X_B|$ and $\phi_{V,A,R}(X) = 0$. On the other hand, both $\phi_{V,A,R}(Y \cup e) = \phi_{V,A,R}(Y) = -|Y_B|$ implying $\partial_e \phi_{V,A,R}(Y) = 0 > \partial_e \phi_{V,A,R}(X)$. This is a violation of submodularity. However, these are the only cases where submodularity is violated, and it turns out that this "almost submodularity" property helps to guarantee the submodularity of our main building block which we define next.

The main building block. Let $R \subseteq A \subseteq V$ be non-empty subsets of a finite set V and denote $B := V \setminus A$. Let $g : 2^B \to \mathbb{R}$ be a set function on B and $M \geq 0$ be a parameter such that $\max_{S \subseteq B} |g(S)| \leq M$. Our main building block is the function $F^{M,g}_{V,A,R} : 2^V \to \mathbb{R}$ defined as

$$F^{M,g}_{V,A,R}(S) := f_{A,R}(S \cap A) + \frac{1}{2|V|} \cdot \phi_{V,A,R}(S) + \frac{1}{4M|V|} \cdot \mathbf{1}(S_A = R) \cdot g(S \cap B). \tag{3.5}$$

The function $F^{M,g}_{V,A,R}$ will be used in Section 3.3.2 to construct a function family on V by choosing g from the function family recursively defined on B. To show the submodularity and structural properties of minimizers of this recursive constructed function family, we first prove the following properties of the function $F^{M,g}_{V,A,R}$.

Lemma 3.3.2 (Properties of main building block)**.** *Let V be a finite set of elements, $R \subseteq A \subseteq V$ be non-empty subsets of V, and denote $B := V \setminus A$. Let $g : 2^B \to \mathbb{R}$ be a submodular function taking values in $[0, M]$ that has a unique minimizer $S_g^* \subseteq B$. Then the function $F := F_{V,A,R}^{M,g}$ defined in (3.5) satisfies the following properties:*

1. *(Non-negativity and boundedness) For any subset $S \subseteq V$, we have $F(S) \in [0, 2]$,*
2. *(Unique Minimizer) F has a unique minimizer $R \cup S_g^*$,*
3. *(Submodularity) F is submodular.*

As mentioned in Section 3.1.2, the main insight behind the proof of Lemma 3.3.2 is that the scale of the function $\frac{1}{4M|V|} \cdot \mathbf{1}(S_A = R) \cdot g(S_B)$ is smaller than that of $\frac{1}{2|V|} \cdot \phi_{V,A,R}(S)$, and both are much smaller than that of $f_{A,R}$. As such, the minimizer S^* and the range of $F_{V,A,R}^{M,g}$ are dominantly determined by the function $f_{A,R}$, enforcing $S_A^* = R$ and thus $f_{A,R}(S_A^*) = \phi_{V,A,R}(S^*) = 0$. Moreover, most cases where submodularity fails to hold for the function $\frac{1}{4M|V|} \cdot \mathbf{1}(S_A = R) \cdot g(S_B)$ can be corrected by the submodularizer $\frac{1}{2|V|} \cdot \phi_{V,A,R}(S)$, and the very few cases where submodularity fails to hold for $\frac{1}{2|V|} \cdot \phi_{V,A,R}(S)$ can be fixed by the dominant submodular function $f_{A,R}$. We postpone a formal proof of Lemma 3.3.2 to Section 3.5.

3.3.2 The Function Family

Using our main building block described in Section 3.3.1, we now define the function family $\mathcal{F}_r(V)$ recursively for all finite sets V with $|V|$ divisible by $2r$.

The base case: when $|V| = 2r$. In this case, we let $\mathcal{F}_r(V) := \{f_{V,R} : R \subseteq V, |R| = r\}$.

Recursive definition. Suppose the function family $\mathcal{F}_r(V)$ has been defined for all $|V| = 2r(k-1)$ for integer $k \geq 2$, we now define the family $\mathcal{F}_r(V)$ for $|V| = 2rk$ as follows:

$$\mathcal{F}_r(V) := \{F_{V,A,R}^{2,g} : R \subseteq A \subseteq V, |R| = |A|/2 = r, g \in \mathcal{F}_r(V \setminus A)\},$$

where we recall from (3.5) that

$$F^{2,g}_{V,A,R} = f_{A,R}(S_A) + \frac{1}{2|V|} \cdot \phi_{V,A,R}(S) + \frac{1}{8|V|} \cdot \mathbf{1}(S_A = R) \cdot g(S_B). \tag{3.6}$$

This completes the recursive definition of the family of functions $\mathcal{F}_r(V)$, where $|V|$ is divisible by $2r$. When $|V|$ is not a multiple of $2r$, we may also naturally extend the definition above by making $|V| - 2r \cdot \lfloor \frac{|V|}{2r} \rfloor$ elements "dummy" in V. More precisely, we let $V' \subseteq V$ be an arbitrary subset with size $|V'| = 2r \cdot \lfloor \frac{|V|}{2r} \rfloor$, and define the function family to only depend on elements in V'.

Explicit Formula for Our Construction. We give more explicit expressions for functions in $\mathcal{F}_r(V)$ recursively defined above, assuming $|V|$ is divisible by $2r$. Let $\ell := |V|/2r$, and consider any partition $\mathcal{A}$ of the universe $V = A_1 \cup A_2 \cup \cdots \cup A_\ell$, where $|A_i| = 2r$ for all $i \in [\ell]$. Furthermore, we select subsets $R_i \subseteq A_i$ for each $i \in [\ell]$ with size $|R_i| = r$. Let $\mathcal{R}$ denote the collection of these R_i's. We denote $B_i := \cup_{j=i}^{\ell} A_j = V \setminus (\cup_{j=1}^{i-1} A_j)$ the remaining set of elements when $A_1, \cdots, A_{i-1}$ are removed from V. Given the partition $\mathcal{A}$ and the family of subsets $\mathcal{R}$, we define a function $F_{\mathcal{A},\mathcal{R}} : 2^V \to \mathbb{R}$ as follows. For any $S \subseteq V$, let k_S be the smallest index $k \in [\ell]$ such that $S_{A_k} := S \cap A_k \neq R_k$. If such an index k_S does not exist, that is $S \cap A_k = R_k$ for all $k \in [\ell]$, then we set $F_{\mathcal{A},\mathcal{R}}(S) := 0$. Otherwise, we define its value

$$F_{\mathcal{A},\mathcal{R}}(S) := \left(\prod_{j=0}^{k_S - 2} \frac{1}{8(|V| - 2jr)} \right) \cdot \left(f_{A_{k_S},R_{k_S}}(S_{A_{k_S}}) + \frac{1}{2|B_{k_S}|} \cdot \phi_{B_{k_S},A_{k_S},R_{k_S}}(S_{B_{k_S}}) \right) \tag{3.7}$$

where $f_{A_{k_S},R_{k_S}}$ and $\phi_{B_{k_S},A_{k_S},R_{k_S}}$ as defined in (3.3) and (3.4).

We now claim that the function family $\mathcal{F}_r(V)$ defined above coincides with the collection of all functions $F_{\mathcal{A},\mathcal{R}}$, for all partitions $V = A_1 \cup A_2 \cup \cdots \cup A_\ell$ with $|A_i| = 2r, \forall i \in [\ell]$ and subsets $R_i \subseteq A_i$ with $|R_i| = r, \forall i \in [\ell]$. To see why this is the case, note that in (3.6), the functions $f_{A_j,R_j}(S_{A_j}) = \phi_{B_j,A_j,R_j}(S_{B_j}) = 0$ for all $j \leq k_S - 1$, and the indicator $\mathbf{1}(S_{A_{k_S}} = R_{k_S}) = 0$. It follows that the functions $f_{A_{k_S},k_S}$ and $\phi_{B_{k_S},A_{k_S},R_{k_S}}$ are the only non-zero components when we expand out the recursive part g in (3.6).

The explicit expression (3.7) reveals important insights into why functions in $\mathcal{F}_r(V)$ take a large number of rounds to minimize. Roughly speaking, any query S would only reveal information about the subsets $R_j \subseteq A_j$ for $j \leq k_S$, but nothing about subsets $R_j \subseteq A_j$ for any $j \geq k_S + 1$. If in each round of queries, an algorithm advances k_S by at most 1, then obtaining full information about the function $F_{\{A_i\},\{R_i\}}$ requires at least $n/2r$ rounds of queries.

Properties of Our Construction. The following lemma collects properties of the function family $\mathcal{F}_r(V)$. In particular, any function $F \in \mathcal{F}_r(V)$ is submodular, and its unique minimizer admits a partition structure. These properties follow from the corresponding properties of our main building block proved in Lemma 3.3.2

Lemma 3.3.3 (Properties of our construction)**.** *Let V be a finite set of elements and $r \in \mathbb{Z}_+$ satisfies $2r$ divides $|V|$. Then any function $F \in \mathcal{F}_r(V)$ satisfies the following properties:*

1. *(Non-negativity and boundedness) For any subset $S \subseteq V$, we have $F(S) \in [0, 2]$,*

2. *(Unique Minimizer) F has a unique minimizer of the form $S^* = \cup_{i=1}^{\ell} R_i$, where $V = A_1 \cup \cdots \cup A_\ell$ forms a partition with $\ell = |V|/2r$ and $|A_i| = 2r, \forall i \in [\ell]$, and subsets $R_i \subseteq A_i$ have size $|R_i| = r, \forall i \in [\ell]$,*

3. *(Submodularity) F is submodular.*

Proof. We prove the lemma by induction based on the size of the ground set V.

The base case. The base case is when $|V| = 2r$ and the statement in this case follows because the function $f_{V,R}$ has range $\{0, 1, 2\}$, unique minimizer R and is submodular by Lemma 3.3.1.

The induction step. Suppose we have proven the three properties of the lemma when the size of the ground set is $2r(k-1)$ for some $k \geq 2$, we now prove the three properties for $|V| = 2rk$.

Note that any function $F \in \mathcal{F}_r(V)$ takes the form

$$F(S) = F^{2,g}_{V,A,R}(S) = f_{A,R}(S_A) + \frac{1}{2|V|} \cdot \phi_{V,A,R}(S) + \frac{1}{8|V|} \cdot \mathbf{1}(S_A = R) \cdot g(S_B).$$

for some subsets $R \subseteq A \subseteq V$ such that $|R| = |A|/2 = r$, and function $g \in \mathcal{F}_r(B)$ with $B = V \setminus A$. By induction hypothesis, g satisfies the three properties in the lemma. The three properties for function F then follows immediately from applying Lemma 3.3.2 with $M = 2$. □

3.4 Lower Bounds

In this section, we leverage our construction of the function family $\mathcal{F}_r(V)$ from Section 3.3 to prove lower bounds for SFM. In Section 3.4.1, we prove an $\Omega(n \log n)$ evaluation query complexity lower bound for any deterministic algorithm that minimizes functions in $\mathcal{F}_r(V)$, even when $r = 1$. Then, in Section 3.4.2, we show that any randomized parallel SFM algorithm that makes at most $Q = \mathsf{poly}(n)$ evaluation oracle queries per round, with high probability, takes at least $\Omega(n/\log n)$ rounds to minimize a uniformly random function $F \in \mathcal{F}_r(V)$ for $r = \Theta(\log n)$.

3.4.1 Query Complexity Lower Bound for Deterministic Algorithms

In this subsection, we prove the query complexity lower bound for deterministic SFM algorithms in Theorem 3.1.1, with the function F chosen adversarially from the function family $\mathcal{F}_1(V)$. More specifically, we prove the following theorem which immediately implies Theorem 3.1.1.

Theorem 3.4.1 (Query complexity lower bound for deterministic algorithms)**.** *Let V be a finite set with n elements. For any deterministic SFM algorithm* ALG*, there exists a submodular function $F \in \mathcal{F}_1(V)$ such that* ALG *makes at least $\frac{n}{2}\log_2(\frac{n}{4})$ evaluation oracle queries to minimize F.*

Let us fix a deterministic algorithm ALG. We prove that there exists a function $F \in \mathcal{F}_1(V)$ on which ALG must make at least $\frac{n}{2}\log\left(\frac{n}{4}\right)$ evaluation oracle queries. From (3.6), recall that any function $F \in \mathcal{F}_1(V)$ is specified by subsets $R \subseteq A \subseteq V$ where $|A| = 2$ and $|R| = 1$, and a function $g \in \mathcal{F}_1(B)$, where $B := V \setminus A$. As R contains only a single element and we abuse notation and call that element R as well. The function F is then given by $F(S) := f_{A,R}(S_A) + \frac{1}{2|V|} \cdot \phi_{V,A,R}(S) + \frac{1}{8|V|} \cdot \mathbf{1}(S_A = R) \cdot g(S_B)$. Recall S_A is the shorthand for $S \cap A$ and S_B is the shorthand for $S \cap B$. By Lemma 3.3.3, $F(S)$ has a unique minimizer S^* with $S^*_A = R$ and S^*_B is the unique minimizer of $g(S_B)$.

By construction, until ALG queries a set S with $S_A = R$, that is, $S \cap A$ is precisely the singleton R, it obtains no information about g. More precisely, the answers given to ALG are the same no matter which $g \in \mathcal{F}_1(B)$ is picked. The heart of the lower bound is the following lemma which asserts that an adversary can always choose an (A, R) pair such that the first $O(\log n)$-queries of ALG "miss R", that is, $S_i \cap A \neq R$.

Lemma 3.4.2. *Fix a deterministic algorithm* ALG *and let* $T := \lfloor \log n \rfloor - 1$. *There exist* $R \subseteq A \subseteq V$ *with* $|R| = 1$ *and* $|A| = 2$ *such that the first* T *(possibly adaptive) queries* $S^1, \cdots, S^T$ *made by* ALG *to the evaluation oracle* EO *satisfy* $S^i_A \neq R$ *for all* $i \in [T]$.

Before we prove the above lemma, let us first use it to prove Theorem 3.1.1.

Proof of Theorem 3.1.1. Fix a deterministic algorithm ALG. For any even integer $n \geq 2$, let $h(n)$ denote the smallest integer such that ALG makes at most $h(n)$ oracle calls to minimize any submodular function $F \in \mathcal{F}_1(V)$ with $|V| = n$, even when ALG is given the information that the submodular function is picked from this family. We claim that $h(n) \geq \frac{n}{2}\log(\frac{n}{4})$. Since by Lemma 3.3.3, any function $F \in \mathcal{F}_1(V)$ is submodular, this would imply Theorem 3.1.1. We prove the claim by induction; the base case of $n = 2$ holds vacuously.

Let $T = \lfloor \log n \rfloor - 1$. By Lemma 3.4.2, we can choose subsets $R \subseteq A \subseteq V$ such that $|R| = 1$, $|A| = 2$, and for the first T (possibly adpative) queries $S^1, \ldots, S^T$ of ALG, we have

$S_A^i \neq R$ hold for all $i \in [T]$. Now consider the function $F \in \mathcal{F}_1(V)$ defined as

$$F(S) := f_{A,R}(S_A) + \frac{1}{2|V|} \cdot \phi_{V,A,R}(S) + \frac{1}{8|V|} \cdot \mathbf{1}(S_A = R) \cdot g(S_B),$$

where (A, R) are these subsets, $B = V \setminus A$, and g, by induction, is the function in $\mathcal{F}_1(B)$ on which ALG takes $h(n-2)$ queries (since $|B| = |V| - 2$) to find the unique minimizer. By the choice of (A, R), since $S_A^i \neq R$, the evaluations of $F(S^i)$ are the same *for all* $g \in \mathcal{F}_1(B)$. In other words, in its first $T = \lfloor \log n \rfloor - 1$ queries, ALG does not obtain any information about the function g.

After T queries, suppose we provide ALG with (A, R). By Lemma 3.3.3, ALG now needs to minimize g. Since the answers received by ALG are consistent with any $g \in \mathcal{F}_1(B)$, by induction, ALG takes at least $h(n-2)$ queries to minimize g. Therefore, we get the recursive inequality $h(n) \geq h(n-2) + \lfloor \log n \rfloor - 1$. This implies $h(n) \geq \frac{n}{2} \log(\frac{n}{4})$. proving the theorem statement. □

Now we are left to prove Lemma 3.4.2.

Proof of Lemma 3.4.2. The proof is via an adversary argument where the EO is an adversary trying to foil the deterministic algorithm ALG. In particular, EO can choose to not commit to the sets (A, R) in the definition of the function $F \in \mathcal{F}_1$ at the beginning. Instead, at every query S^i, the adversary oracle EO gives an answer consistent with a function $F(S) = f_{A,R}(S_A) + \frac{1}{2|V|} \cdot \phi_{V,A,R}(S) + \mathbf{1}(S_A = R) g(S_B)$ for some (A, R) such that $S_A^i \neq R$ and such that all previous query answers are also consistent with S. We now show that this is possible for the first T queries.

It is in fact convenient to consider the following modified evaluation oracle EO′. When queried with a set $S \subseteq V$, EO′ returns the following information: (1) whether $S_A = R$, or $S_A \subsetneq R$, or $R \subsetneq S_A$, or if S_A is neither a subset nor a superset of R, and (2) the size of $|S_A|$. Note that unless $S_A = R$, the information returned by EO′ is enough for the algorithm to compute $F(S)$. Indeed, when $S_A \neq R$, the function $F(S) = f_{A,R}(S_A) + \frac{1}{2|V|} \cdot \phi_{V,A,R}(S)$ so the

information in (1) and (2), together with $|S|$ determine the value of $F(S)$. In short, we can use EO' to simulate EO till a query S with $S_A = R$ is made. We now show how to construct the adversary EO' such that in the first T queries, it can give answers such that $S^i_A \neq R$ for all $i \in [T]$ and there exists an $R \subseteq A \subseteq V$ consistent with all answers given so far.

The adversary EO' maintains an active set U^1 of elements which is initialized to V. Consider the first query S^1 made by ALG. If $|U^1 \cap S^1| \geq |U^1|/2$, then EO' does the following: (a) it sets $U^2 \leftarrow U^1 \cap S^1$, and (b) answers $S^1_A = A$, that is, $R \subsetneq S^1_A$ and $|S^1_A| = 2$. If $|U^1 \cap S^1| < |U^1|/2$, then EO' does the following: (a) it sets $U^2 \leftarrow U^1 \setminus S^1$, and (b) answers $S^1_A = \emptyset$, that is, $R \supsetneq S^1_A$ and $|S^1_A| = 0$. In short, the adversary EO' commits that $A \subseteq U^2$, and for *any* such A and any $R \subseteq A$, the answer given above would be consistent.

More generally, at the beginning of round i, the adversary EO' has an active set U^i with ≥ 4 elements. Upon query S^i, if $|U^i \cap S^i| \geq |U^i|/2$, then EO' answers $R \subsetneq S^i_A$ and $|S^i_A| = 2$, and modifies $U^{i+1} \leftarrow U^i \cap S^i$, otherwise, EO' answers $R \supsetneq S^i_A$ and $|S^i_A| = 0$, and modifies $U^{i+1} \leftarrow U^i \setminus S^i$.

Since the size of U^i can at most halve, at the end of $T = \lfloor \log_2(n) \rfloor - 1$ rounds, the adversary EO' ends up with a set U^{T+1} with ≥ 2 elements. At this point, EO' can choose any subset $R \subseteq A \subseteq U^{T+1}$ with $|A| = 2$ and $|R| = 1$, and (a) all answers given above are consistent, and (b) $S^i_A \neq R$ for all $i \in [T]$. This completes the proof of the lemma. □

Remark 3.4.3. *We note that Lemma 3.4.2 is false if* ALG *is allowed to be randomized. Indeed, if* $|A| = 2$ *and* $R \subseteq A$ *has* $|R| = 1$*, then any query* S *which picks every element with probability* $1/2$ *will satisfy* $S_A = R$ *with probability* $1/4$*. Therefore, the proof idea breaks down for randomized algorithms. On the other hand, we do not know of a randomized algorithm for minimize functions in* $\mathcal{F}_1(V)$ *that makes* $O(n)$ *queries and succeeds with constant probability.*

3.4.2 Parallel Lower Bound for Randomized Algorithms

In this subsection, we prove the $\Omega(n/C \log n)$-lower bound on the number of rounds for (possibly randomized) parallel SFM algorithms in Theorem 3.1.2. By Yao's minimax principle,

Theorem 3.1.2 is implied by the following theorem where the function F is chosen uniformly at random from the family $\mathcal{F}_r(V)$ with $r = C \log n$.

Theorem 3.4.4 (Parallel lower bound for randomized algorithms)**.** *Let $C \geq 2$ be any constant. Let V be a finite set with n elements, and $r \geq C \log n$ be an integer such that $2r$ divides n. Then any parallel algorithm that makes at most $Q := n^C$ queries per round, and runs for $< (n/2r)$ rounds, fails to minimize a uniformly random submodular function $F \in \mathcal{F}_r(V)$, with high probability.*

Proof. By the recursive construction of the function family $\mathcal{F}_r(V)$ in Section 3.3.2, we may view a random submodular function F drawn from the uniform distribution over $\mathcal{F}_r(V)$ being obtained as follows. We first select a uniformly random subset $A_1 \subseteq V$ of size $|A_1| = 2r$ and a uniformly random subset $R_1 \subseteq A_1$ with size $|R_1| = r$. Denoting $B := V \setminus A_1$, we then draw a uniformly random function $g \in \mathcal{F}_r(B)$, and let $F(S) := f_{A_1,R_1}(S_{A_1}) + \frac{1}{2|V|} \cdot \phi_{V,A_1,R_1}(S) + \frac{1}{8|V|} \cdot \mathbf{1}(S_{A_1} = R_1) \cdot g(S_B)$. Coupled with $F(S)$ in terms of the randomness of the subsets A_1 and R_1, we also let $F'(S) := f_{A_1,R_1}(S_{A_1}) + \frac{1}{2|V|} \cdot \phi_{V,A_1,R_1}(S)$.

Since we have specified a distribution over submodular functions, it suffices to prove that any deterministic algorithm which runs in $< \frac{n}{2r}$ rounds and makes $\leq n^C$ queries per round, fails to find the minimizer of F with high probability. In the remainder we prove this statement.

Consider the set of queries $S_1^1, \cdots, S_1^Q$ made by a deterministic algorithm ALG in the first round. We start by showing that with high probability, $S_1^i \cap A_1 \neq R_1$ for all $i \in [Q]$. This is because for any S_1^i and any fixed outcome of A_1, since R_1 is a uniformly random subset of A_1 with size r, there are $\binom{2r}{r} \geq \frac{2^{2r}}{2r+1} \geq \frac{n^{2C}}{2C \log n + 1}$ possible choices of R. Therefore, for any query S_1^i and any fixed outcome of A_1, the probability that $S_1^i \cap A_1 = R_1$ is at most $\frac{2C \log n + 1}{n^{2C}}$. It then follows by a union bound over all S_1^i that with probability at least $1 - \frac{2C \log n + 1}{n^C}$, the event $\mathcal{E}_1 := \{S_1^i \cap A_1 \neq R_1, \forall i \in [Q]\}$ holds.

Now conditioning on the event $\mathcal{E}_1$, the output of the evaluation oracle when queried with S_1^i would be $F(S_1^i) = F'(S_1^i)$, for all $i \in [Q]$. Note, however, that the function F' does not

depend on the randomness of $g \in \mathcal{F}_r(B)$. Thus, even when given the information of R and A after the first round of queries, ALG does not obtain any information about the uniformly random function $g \in \mathcal{F}_r(B)$. Therefore, we can apply the argument in the previous paragraph to the set of queries $S_2^1, \ldots, S_2^Q$ in the second round of the algorithm. In particular, with probability at least $1 - 1/n^C$, the event $\mathcal{E}_2 := \{S_2^i \cap A_2 \neq R_2, \forall i \in [Q]\}$ holds.

More generally, if the algorithm makes $k < n/2r$ rounds of queries, then with probability $\geq 1 - \frac{k(2C\log n+1)}{n^C} > 1 - \frac{1}{n^{C-1}}$ all the events $\mathcal{E}_i$ occur. This implies that the answers obtained by the algorithm are consistent with any function in $\mathcal{F}_r(V)$ where the sets $A_1, \ldots, A_k$ and $R_1, \ldots, R_k$ are fixed, but the sets $A_{k+1}, \ldots, A_{n/2r}$ and $R_{k+1}, \ldots, R_{n/2r}$ are completely random. Since the unique minimizer of F is the set $(R_1 \cup R_2 \cup \cdots \cup R_{n/2r})$, no matter which set the deterministic algorithm returns, it will err with probability at least $1 - \frac{1}{n^{C-1}}$. This completes the proof of the theorem. □

3.5 Proof of Properties of Main Building Block

In this section, we give the proof for Lemma 3.3.2 which we restate below for convenience.

Lemma 3.3.2 (Properties of main building block)**.** *Let V be a finite set of elements, $R \subseteq A \subseteq V$ be non-empty subsets of V, and denote $B := V \setminus A$. Let $g : 2^B \to \mathbb{R}$ be a submodular function taking values in $[0, M]$ that has a unique minimizer $S_g^* \subseteq B$. Then the function $F := F_{V,A,R}^{M,g}$ defined in (3.5) satisfies the following properties:*

1. *(Non-negativity and boundedness) For any subset $S \subseteq V$, we have $F(S) \in [0, 2]$,*
2. *(Unique Minimizer) F has a unique minimizer $R \cup S_g^*$,*
3. *(Submodularity) F is submodular.*

Proof. We prove the three properties in the lemma statement separately below.

Property 1: Non-negativity and boundedness. For any subset $S \subseteq V$, we consider three different cases depending on the relation between S_A and R.

Case 1: $S_A = R$**.** In this case, $f_{A,R}(S_A) = 0$ and $\phi_{V,A,R}(S) = 0$, so we have

$$F(S) = 0 + 0 + \frac{1}{4M|V|} \cdot g(S_B)$$

Since $g(S_B) \in [0, M]$ in this case we get $F(S) \in [0, \frac{1}{4|V|}] \in [0, 1/4]$.

Case 2: $S_A \subsetneq R$ **or** $S_A \supsetneq R$**.** In this case, $f_{A,R}(S_A) = 1$ and $|\phi_{V,A,R}(S)| = |S_B| \leq |V|$. Furthermore, $\mathbf{1}(S_A = R) = 0$. Thus,

$$F(S) = 1 + \frac{1}{2|V|} \cdot \phi_{V,A,R}(S)$$

So, in this case, $F(S) \in [0.5, 1.5]$.

Case 3: S_A **is neither a subset nor a superset of** R**.** In this case, $f_{A,R}(S_A) = 2$ and $\phi_{V,A,R}(S) = 0$, and therefore $F(S) = 2 \in [0, 2]$.

This completes the proof of Property 1.

Property 2: Unique minimizer. An inspection of the cases in the above argument regarding Property 1 shows that for any subset S with $S_A \neq R$, we have $F(S) \geq 0.5$, while when $S_A = R$, we have $F(S) \leq 0.25$. Therefore, the minimizer S of F must have $S_A = R$. Furthermore, when $S_A = R$ then $F(S) = \frac{1}{4M|V|} \cdot g(S_B)$ and the function is minimized when $S_B = S_g^*$. This proves the second property in the lemma statement.

Property 3: Submodularity. This is the most interesting part of the proof. Let $X, Y \subseteq V$ be two arbitrary subsets of the ground set. Our goal is to prove

$$F(X) + F(Y) \geq F(X \cup Y) + F(X \cap Y). \tag{3.8}$$

In the following, we prove (3.8) by a case analysis. For convenience, define the collection of subsets of A that are either subsets or supersets of R as $\mathcal{H}_{A,R} := \{S \subseteq A : S \subseteq R \text{ or } S \supseteq R\}$. Note that R lies in this family as well. We consider three different cases depending on whether or not X_A and Y_A lie in the set family $\mathcal{H}_{A,R}$. For notational simplicity, the subscripts in the notations $f_{A,R}$, $\phi_{V,A,R}$ and $\mathcal{H}_{A,R}$ will be dropped throughout the rest of this proof since the

sets V, A, R have been fixed and there is no ambiguity.

(Case 1): $X_A, Y_A \notin \mathcal{H}$**.** In this case, we have $\phi(X) = \phi(Y) = 0$, $f(X_A) = f(Y_A) = 2$, and $\mathbf{1}(X_A = R) = \mathbf{1}(Y_A = R) = 0$. Thus the LHS of (3.8) is simply $F(X) + F(Y) = f(X_A) + f(Y_A) = 4$.

Now, note that $(X \cup Y)_A := (X \cup Y) \cap A = X_A \cup Y_A$ and $(X \cap Y)_A := (X \cap Y) \cap A = X_A \cap Y_A$. Therefore, if $X_A, Y_A \notin \mathcal{H}$, then neither $(X \cup Y)_A$ nor $(X \cap Y)_A$ can be R. If the former, then both $X_A, Y_A \subseteq R$ implying both are in $\mathcal{H}$. If the latter, then both $X_A, Y_A \supseteq R$ implying both are in $\mathcal{H}$. Therefore, the RHS of (3.8) doesn't have any "g-terms", and is

$$\mathsf{RHS} = f(X_A \cap Y_A) + f(X_A \cup Y_A) + \frac{1}{2|V|} \cdot (\phi(X \cap Y) + \phi(X \cup Y)).$$

Note that if we also have $X_A \cap Y_A, X_A \cup Y_A \notin \mathcal{H}$, then the contribution of ϕ to the RHS would be 0, and $\mathsf{LHS} \geq \mathsf{RHS}$ follows from the submodularity of f in Lemma 3.3.1. So we only need to consider the scenarios where $X_A \cap Y_A \in \mathcal{H}$ or $X_A \cup Y_A \in \mathcal{H}$ (or both). In any of these scenarios, we have $f(X_A \cap Y_A) + f(X_A \cup Y_A) \leq 3$, since $f(S_A) = 1$ for $S_A \in \mathcal{H}$. Now since $|\phi(S)| \leq |V|$ for any subset $S \subseteq V$, we have $\frac{1}{2|V|} \cdot (\phi(X \cap Y) + \phi(X \cup Y)) \leq 1$. Thus, $\mathsf{RHS} \leq 4$, and (3.8) immediately follows.

(Case 2): $X_A, Y_A \in \mathcal{H}$**.** In this case, we need to consider multiple further subcases depending on whether X_A or Y_A coincide with R.

Case 2.1: $X_A = Y_A = R$**.** In this subcase, $F(S) = f(S_A) + \frac{1}{4M|V|} \cdot g(S_B)$ for all $S \in \{X, Y, X \cap Y, X \cup Y\}$, so (3.8) follows from the submodularity of f and g.

Case 2.2: $R \subsetneq X_A, Y_A$**.** In this subcase, we have $R \subsetneq X_A \cup Y_A$ and $R \subseteq X_A \cap Y_A$. If it happens that $X_A \cap Y_A = R$, then we have

$$\mathsf{LHS} = f(X_A) + f(Y_A) + \frac{1}{2|V|} \cdot (\phi(X) + \phi(Y)),$$
$$\mathsf{RHS} = f(R) + f(X_A \cup Y_A) + \frac{1}{4M|V|} \cdot g(X_B \cap Y_B) + \frac{1}{2|V|} \cdot \phi(X \cup Y).$$

Notice that $f(X_A) = f(Y_A) = f(X_A \cup Y_A) = 1$ but $f(R) = 0$. It follows that

$$\begin{aligned}\mathsf{LHS} - \mathsf{RHS} &= 1 - \frac{1}{2|V|} \cdot (|X_B| + |Y_B|) + \frac{1}{2|V|} \cdot |X_B \cup Y_B| - \frac{1}{4M|V|} \cdot g(X_B \cap Y_B) \\ &= 1 - \frac{1}{2|V|} \cdot |X_B \cap Y_B| - \frac{1}{4M|V|} \cdot g(X_B \cap Y_B) > 0,\end{aligned}$$

where the last inequality follows because the range of g is within $[0, M]$ by lemma assumption. If, on the other hand, that $R \subsetneq X_A \cap Y_A$, then the RHS of (3.8) becomes

$$\mathsf{RHS} = f(X_A \cap Y_A) + f(X_A \cup Y_A) + \frac{1}{2|V|} \cdot (\phi(X \cap Y) + \phi(X \cup Y)).$$

By a simple counting we have

$$\phi(X) + \phi(Y) = -(|X_B| + |Y_B|) = -(|X_B \cap Y_B| + |X_B \cup Y_B|) = \phi(X \cap Y) + \phi(X \cup Y).$$

and in this case, $\mathsf{LHS} - \mathsf{RHS} = 0$.

Case 2.3: $X_A, Y_A \subsetneq R$**.** The analysis in this subcase is almost identical to **Case 2.2**.

Case 2.4: $X_A \subsetneq R \subsetneq Y_A$ **or** $Y_A \subsetneq R \subsetneq X_A$**.** We assume it is the former by symmetry between X and Y. Then we have $X_A \cap Y_A = X_A \subsetneq R$ and $X_A \cup Y_A = Y_A \supsetneq R$. From the definition of F, it follows that

$$\begin{aligned}\mathsf{LHS} - \mathsf{RHS} = (f(X_A) + f(Y_A) - f(X_A \cap Y_A) - f(X_A \cup Y_A)) + \\ \frac{1}{2|V|} \cdot (\phi(X) + \phi(Y) - \phi(X \cap Y) + -\phi(X \cup Y))\end{aligned}$$

The first term is ≥ 0 because of the submodularity of f. Furthermore, in this case

$$\phi(X) + \phi(Y) = \frac{1}{2|V|} \cdot (|X_B| - |Y_B|)$$
$$\phi(X \cap Y) + \phi(X \cup Y) = \frac{1}{2|V|} \cdot (|X_B \cap Y_B| - |X_B \cup Y_B|) \leq \frac{1}{2|V|} \cdot (|X_B| - |Y_B|)$$

and thus the second term is also ≥ 0. This proves (3.8) in this case.

Case 2.5: $X_A \subsetneq R = Y_A$ **or** $Y_A \subsetneq R = X_A$**.** We assume wlog that it is the former. Note that $X_A \cap Y_A = X_A$ and $X_A \cup Y_A = Y_A = R$. Therefore,

$$\mathsf{LHS} = f(X_A) + f(Y_A) + \frac{1}{2|V|} \cdot |X_B| + \frac{1}{4M|V|} \cdot g(Y_B),$$
$$\mathsf{RHS} = f(X_A) + f(Y_A) + \frac{1}{2|V|} \cdot |X_B \cap Y_B| + \frac{1}{4M|V|} \cdot g(X_B \cup Y_B).$$

In the above, if $X_B = X_B \cap Y_B$ then it must be that $X_B \subseteq Y_B$. It follows that $Y_B = X_B \cup Y_B$ and we obtain equality in (3.8). On the other hand, if $X_B \neq X_B \cap Y_B$, then $|X_B| \geq |X_B \cap Y_B| + 1$, and so we have

$$\mathsf{LHS} - \mathsf{RHS} \geq \frac{1}{2|V|} + \frac{1}{4M|V|} \cdot (g(Y_B) - g(X_B \cup Y_B)) \geq \frac{1}{4|V|} > 0,$$

where we used the lemma assumption that the range of g is within $[0, M]$. This again proves (3.8).

Case 2.6: $X_A = R \subsetneq Y_A$ **or** $Y_A = R \subsetneq X_A$**.** Assume wlog that it is the former. Then we have

$$\mathsf{LHS} = f(X_A) + f(Y_A) + \frac{1}{4M|V|} \cdot g(X_B) - \frac{1}{2|V|}|Y_B|,$$
$$\mathsf{RHS} = f(X_A) + f(Y_A) + \frac{1}{4M|V|} \cdot g(X_B \cap Y_B) - \frac{1}{2|V|} \cdot |X_B \cup Y_B|.$$

The analysis from here is almost identical to that in **Case 2.5**.

(Case 3): $X_A \in \mathcal{H}, Y_A \notin \mathcal{H}$ **or** $Y_A \in \mathcal{H}, X_A \notin \mathcal{H}$**.** We assume wlog that it is the former. Note that $f(Y_A) = 2$. This case is further divide into three subcases below depending on the relation between X_A and R.

Case 3.1: $X_A = R$. And so, $f(X_A) = 0$. In this subcase, note that $X_A \cap Y_A \subsetneq R$ since R isn't be a subset of Y_A, and $R \subsetneq X_A \cup Y_A$. So, $f(X_A \cap Y_A) = f(X_A \cup Y_A) = 1$. Then,

$$\begin{aligned}
\mathsf{LHS} &= f(X_A) + f(Y_A) + \frac{1}{4M|V|} \cdot g(X_B) = 2 + \frac{1}{4M|V|} \cdot g(X_B) \geq 2, \\
\mathsf{RHS} &= f(X_A \cap Y_A) + f(X_A \cup Y_A) + \frac{1}{2|V|} \cdot (|X_B \cap Y_B| - |X_B \cup Y_B|) \leq 2.
\end{aligned}$$

where we used the non-negativity of g in the argument about LHS. In this case, we have established (3.8).

Case 3.2: $X_A \subsetneq R$. And so, $f(X_A) = 1$. In this case also, we have $X_A \cap Y_A \subsetneq R$. Also note that $X_A \cup Y_A \neq R$ since Y_A is not a subset of R. Therefore,

$$\begin{aligned}
\mathsf{LHS} &= f(X_A) + f(Y_A) + \frac{1}{2|V|} \cdot |X_B| = 3 + \frac{1}{2|V|} \cdot |X_B|, \\
\mathsf{RHS} &= f(X_A \cap Y_A) + f(X_A \cup Y_A) + \frac{1}{2|V|} \cdot |X_B \cap Y_B| + \frac{1}{2|V|} \cdot \phi(X \cup Y) \\
&= 1 + f(X_A \cup Y_A) + \frac{1}{2|V|} \cdot |X_B \cap Y_B| + \frac{1}{2|V|} \cdot \phi(X \cup Y).
\end{aligned}$$

If $X_A \cup Y_A \in \mathcal{H}$, then $f(X_A \cup Y_A) = 1$ and $\phi(X \cup Y) \leq |X_B \cup Y_B| \leq |V|$. Thus,

$$\mathsf{RHS} \leq 2 + \frac{1}{2|V|}|X_B \cup Y_B| + \frac{1}{2|V|}|X_B \cap Y_B| \leq 2.5 + \frac{1}{2|V|}|X_B| < \mathsf{LHS}$$

Thus, (3.8) holds.

If $X_A \cup Y_A \notin \mathcal{H}$, then $f(X_A \cup Y_A) = 2$ and $\phi(X \cup Y) = 0$, and so $\mathsf{RHS} = 3 + \frac{1}{2|V|} \cdot |X_B \cap Y_B| \leq \mathsf{LHS}$ and thus (3.8) holds in this case as well.

Case 3.3: $R \subsetneq X_A$. In this case, we have $R \subsetneq X_A \cup Y_A$ and then

$$\begin{aligned}
\mathsf{LHS} &= f(X_A) + f(Y_A) - \frac{1}{2|V|} \cdot |X_B| = 3 - \frac{1}{2|V|} \cdot |X_B|, \\
\mathsf{RHS} &= f(X_A \cap Y_A) + f(X_A \cup Y_A) + \frac{1}{2|V|} \cdot \phi(X \cap Y) - \frac{1}{2|V|} \cdot |X_B \cup Y_B| \\
&= 1 + f(X_A \cap Y_A) + \frac{1}{2|V|} \cdot \phi(X \cap Y) - \frac{1}{2|V|} \cdot |X_B \cup Y_B|.
\end{aligned}$$

From here one can proceed similarly as in **Case 3.2** to prove (3.8).

Combining all the cases above, we established (3.8) which implies the submodularity of the function F. This completes the proof of the entire lemma. □

Part II

ROUNDING VIA DISCREPANCY THEORY

Chapter 4

MATRIX DISCREPANCY I: PARTIAL COLORING BOUNDS VIA MIRROR DESCENT

In this chapter and the next two chapters, we study matrix discrepancy, a topic that has attracted significant attention in the last decade. The main problem that we will study is the matrix Spencer conjecture (Conjecture 1.3.1). In this chapter, we present our first non-trivial progress towards resolving this conjecture, which improves upon the naïve bound obtained by a random coloring. This chapter is based on a joint paper with Daniel Dadush and Victor Reis [DJR22] that appeared in the *54th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2022)*.

4.1 Introduction

Discrepancy minimization has been a well-studied area of research both in mathematics and computer science [Cha00, Mat99]. We start with a classical setting: given vectors $a_1, \dots, a_n \in \mathbb{R}^m$ each satisfying $\|a_i\|_\infty \le 1$, the goal is to find a coloring $x \in \{\pm 1\}^n$ that minimizes the discrepancy, defined as $\|\sum_{i=1}^n x_i a_i\|_\infty$. A seminal result of Spencer [Spe85] improves upon the $O(\sqrt{n \log m})$ bound of a random coloring via Chernoff and union bound:

Theorem 4.1.1 (Spencer [Spe85])**.** *Let $m \ge n$. Given vectors $a_1, \dots, a_n \in \mathbb{R}^m$ with $\|a_i\|_\infty \le 1$, there exists $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i a_i\|_\infty \le O(\sqrt{n \log(2m/n)})$.*

In particular, when $m = n$, Theorem 4.1.1 states that the discrepancy is at most $O(\sqrt{n})$, as opposed to the $O(\sqrt{n \log n})$ bound for a random coloring. Spencer's theorem is known to be tight up to constants for all $m \ge n$ [Cha00, Mat99].

The Partial Coloring Method. All known proofs of Spencer's theorem are essentially based on the *partial coloring* method, one of the most important and widely applied tech-

niques in discrepancy theory. The method states that to obtain the type of discrepancy bound in Theorem 4.1.1, it suffices to prove the same bound for a partial coloring $x \in [-1,1]^n$ with at least $\Omega(n)$ coordinates in $\{\pm 1\}$. This process is then iterated over the set of coordinates $\{i : |x_i| < 1\}$ to obtain a full coloring. For Spencer-type problems, the discrepancy of the full coloring is at most a constant factor off from the discrepancy of the partial coloring (see Corollary 4.3.2).

The partial coloring method was developed in the early 80s by Beck and refined by Spencer using the entropy method [Bec81, Spe85]. A convex geometry view of partial coloring was developed independently by Gluskin [Glu89]. While these original arguments used the pigeonhole principle and were non-algorithmic, a breakthrough result of Bansal [Ban10], followed by a rich line of work [LM15a, Rot17, LRR17, ES18, RR20a], gave various algorithmic versions. These recent developments also led to new results in approximation algorithms and differential privacy [Rot13, NTZ13, BCKL14a, BN17].

Matrix Spencer Setting. A natural generalization of Spencer's setting to matrices is the following. Given matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$, each satisfying $\|A_i\|_{\mathsf{op}} \leq 1$, the goal is to find a coloring $x \in \{\pm 1\}^n$ that minimizes $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}}$. In particular, Spencer's setting corresponds to the case where all matrices A_i are diagonal.

In the matrix Spencer setting, the non-commutative Khintchine inequality of Lust-Piquard and Pisier [LPP91, Pis03] shows that a random coloring $x \in \{\pm 1\}^n$ has expected discrepancy $\mathbb{E}[\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}}] \leq O(\sqrt{n \log r})$, where each matrix A_i has rank at most $r \leq m$. It is conjectured that the discrepancy bound in Theorem 4.1.1 can be generalized as follows:

Conjecture 1.3.1 (Matrix Spencer Conjecture, [Zou12, Mek14])**.** *Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with each $\|A_i\|_{\mathsf{op}} \leq 1$, there exist signs $x \in \{\pm 1\}^n$ such that $\left\|\sum_{i=1}^n x_i A_i\right\|_{\mathsf{op}} \leq O(\sqrt{n} \cdot \max\{1, \sqrt{\log(m/n)}\})$. In particular, the matrix discrepancy is $O(\sqrt{n})$ for $m = n$.*

In particular, when $\sqrt{n} \leq m \leq n$, the conjectured discrepancy bound is $O(\sqrt{n})$. Despite significant effort, Conjecture 1.3.1 has remained largely open, with partial progress

for block diagonal matrices [LRR17] and rank-1 matrices [MSS15, KLS20]. A solution to Conjecture 1.3.1 will thus likely lead to new techniques and insights in discrepancy theory beyond what is currently known for vector discrepancy.

We note that in Spencer's setting (Theorem 4.1.1) we may assume without loss of generality that $m \geq n$ by the iterated rounding technique [BF81, Bár08, LRS11]. For matrix Spencer, however, the interesting regime starts at $m \geq \sqrt{n}$ (iterated rounding only works when $m^2 < n$). Conjecture 1.3.1 remains open even when $m = n^{1/2+\varepsilon}$ for any constant $\varepsilon > 0$.

Matrix Discrepancy for Schatten Norms. More generally, let[1] $2 \leq p \leq q \leq \infty$, we consider the following matrix discrepancy setting for Schatten norms. Given matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$, each satisfying $\|A_i\|_{S_p} \leq 1$, where $\|\cdot\|_{S_p}$ denotes the Schatten-p norm. The goal is to find a coloring $x \in \{\pm 1\}^n$ to minimize $\|\sum_{i=1}^n x_i A_i\|_{S_q}$, the $S_p \to S_q$ discrepancy. In particular, the matrix Spencer setting corresponds to the case where $p = q = \infty$.

The diagonal case of $S_p \to S_q$ discrepancy, i.e. $\ell_p \to \ell_q$ discrepancy for vectors, is well studied (see [DNTTJ18, RR20a] and the references therein). In fact, the well-known Komlós conjecture asserts that the $\ell_2 \to \ell_\infty$ discrepancy can be upper bounded by a universal constant. For general $\ell_p \to \ell_q$ discrepancy, Reis and Rothvoss [RR20a] proves an optimal partial coloring bound of $O(\sqrt{\min(p, \log(m/n))} \cdot n^{1/2-1/p+1/q})$, assuming $m \geq n$ and $2 \leq p \leq q \leq \infty$. It is a natural question whether these bounds generalize to $S_p \to S_q$ discrepancy.

The Challenge in Using Partial Coloring Method for Matrix Discrepancy. Central to the partial coloring method is to show that the discrepancy body $D := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\| \leq t\}$, i.e. the set of fractional colorings with discrepancy at most t under norm $\|\cdot\|$, is "large" in some sense. A natural notion of largeness, due to Gluskin [Glu89], is that the body D has Gaussian measure at least $2^{-O(n)}$. This measure of largeness has been adopted (sometimes implicitly) in essentially all work on partial coloring [Ban10, LM15a, Rot17, LRR17, ES18, RR20a].

For the setting in Theorem 4.1.1, the discrepancy body D is a polytope defined by the

[1] We make the assumption that $p \leq q$ to avoid a polynomial dependence on m in the discrepancy bound. If $q < p$, then even a single matrix (i.e. $n = 1$) can have discrepancy $m^{1/q-1/p}$.

intersection of strips of the form $|\langle r_i, x\rangle| \le t$, where $r_i \in \mathbb{R}^n$ are the rows of the $m \times n$ matix whose columns are $a_1, \ldots, a_n$. Therefore, Šidák's lemma [Šid67] can be readily used to give a Gaussian measure lower bound of the form $\gamma_n(D) \ge \prod_{i=1}^m \gamma_n(\{x \in \mathbb{R}^n : |\langle r_i, x\rangle| \le t\})$.

In the setting of matrix discrepancy, however, the discrepancy body D has an infinite number of facets. This prevents the use of Gaussian correlation inequalities to lower bound $\gamma_n(D)$. To get around this barrier and use the partial coloring method for matrix discrepancy, one needs a different approach for proving Gaussian measure lower bounds.

4.1.1 Our Results

We lower bound the Gaussian measure of the discrepancy body D via covering numbers for its polar D° with respect to the ℓ_∞-ball (see Section 4.3.1). We then prove the desired covering number estimates using mirror descent, the powerful convex optimization primitive of Nemirovski and Yudin [NY83] (see Sections 4.3.2 to 4.3.4). Our method yields the following applications.

Matrix Spencer for Low-Rank Matrices. Our first result is the following improvement over the $O(\sqrt{n \log r})$ bound for random coloring in the matrix Spencer setting.

Theorem 4.1.2 (Matrix Spencer for Low-Rank Matrices)**.** *Let $m \ge \sqrt{n}$. Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{\mathsf{op}} \le 1$ and $\mathsf{rank}(A_i) \le r$ for all $i \in [n]$, one can efficiently find a coloring $x \in \{\pm 1\}^n$ such that*

$$\left\| \sum_{i=1}^n x_i A_i \right\|_{\mathsf{op}} \le O(\sqrt{n \cdot \max(1, \log(r \cdot \min(1, m/n)))}).$$

When the input matrices have rank $r \le O(n/m)$, the discrepancy bound in Theorem 4.1.2 is $O(\sqrt{n})$ and this proves Conjecture 1.3.1 for low rank matrices in the regime where $m \le n$.

Matrix Spencer for Block Diagonal Matrices. Our second application is the following improved matrix Spencer bound for block diagonal matrices.

Theorem 4.1.3 (Matrix Spencer for Block Diagonal Matrices)**.** *Let $m \geq \sqrt{n}$ and $h \leq m$. Given block diagonal symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{\mathsf{op}} \leq 1$ and block size $h \times h$, one can efficiently find a coloring $x \in \{\pm 1\}^n$ with*

$$\left\| \sum_{i=1}^{n} x_i A_i \right\|_{\mathsf{op}} \leq O(\sqrt{n \cdot \max(1, \log(hm/n))}).$$

In particular, Theorem 4.1.3 proves Conjecture 1.3.1 whenever $h \leq O(n/m)$. This bound was previously proved in [LRR17] under the assumption $h \leq \sqrt{n}$, which we remove here.

We also obtain the following reduction of Conjecture 1.3.1 to the construction of a better quantum relative entropy net for the spectraplex $\mathcal{S}_m := \{X \in \mathbb{R}^{m \times m} : X \succeq 0, \mathsf{tr}(X) = 1\}$.

Corollary 4.1.4 (Better Entropy Net Implies Matrix Spencer)**.** *Let $m \geq \sqrt{n}$. If we can find $T \subseteq \mathcal{S}_m$ with $|T| \leq 2^{O(n)}$ such that for each $X \in \mathcal{S}_m$ there exists $Y \in T$ with $S(X\|Y) \leq O(\max(1, \log(m/n)))$, where $S(X\|Y)$ is the quantum relative entropy between X and Y, then the matrix Spencer conjecture is true.*

In particular, in the proof of Theorem 4.1.3, we construct a $O(\max(1, \log(hm/n)))$-relative entropy net for the set of block diagonal matrices on $\mathcal{S}_m$ with block size $h \times h$ (see Section 4.3.4). Our construction of such relative entropy nets might be of independent interest.

Matrix Discrepancy for Schatten Norms. Theorem 4.1.2 is a special case of the following general matrix discrepancy bound for Schatten norms.

Theorem 4.1.5 (Matrix Discrepancy for Schatten Norms)**.** *Let $m \geq \sqrt{n}$ and $2 \leq p \leq q \leq \infty$. Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \leq 1$ and $\mathrm{rank}(A_i) \leq r$ for all $i \in [n]$, one can efficiently find $x \in [-1, 1]^n$ so that $|\{i : |x_i| = 1\}| \geq n/2$ and*

$$\left\| \sum_{i=1}^{n} x_i A_i \right\|_{S_q} \leq O(\sqrt{n \cdot \min(p, \max(1, \log(rk)))} \cdot k^{1/p - 1/q}),$$

where we denote $k := \min(1, m/n)$. Moreover, we can find a full coloring $x \in \{\pm 1\}^n$ at the expense of a factor of $(1/2 + 1/q - 1/p)^{-1}$.

Our partial coloring result in Theorem 4.1.5 is tight when either $m = \Theta(\sqrt{n})$ (for which we give an alternative proof using Banaszczyk's result [Ban98] in Section 4.7), or when $r = 1$ and $m \geq n$. We provide matching lower bounds for both cases in Sections 4.6.1 and 4.6.2. In particular, our lower bound examples imply a tight $\Omega(\sqrt{n})$ lower bound for rank-1 matrix Spencer when $m = n$.

Corollary 4.1.6 (Rank-1 Matrix Spencer Lower Bound)**.** *There exist rank-1 symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{n \times n}$ with $\|A_i\|_{\mathsf{op}} \leq 1$ such that any $x \in \{\pm 1\}^n$ has $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} \geq \Omega(\sqrt{n})$.*

Another immediate consequence of our lower bounds is an optimal $\Omega(\sqrt{\min(m,n)})$ lower bound for $S_2 \to S_\infty$ discrepancy. This is in stark contrast to the well-known Komlós conjecture for vectors, which asserts that the $\ell_2 \to \ell_\infty$ discrepancy is $O(1)$. Corollary 4.1.7 states that such a conjecture is far from being true for matrices.

Corollary 4.1.7 (Lower Bound for Matrix Komlós)**.** *For any m and n, there exist symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_F \leq 1$ such that any $x \in \{\pm 1\}^n$ has $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} \geq \Omega(\sqrt{\min(m,n)})$.*

Finally, we propose the following generalization of Conjecture 1.3.1:

Conjecture 4.1.8 ($S_p \to S_q$ Matrix Discrepancy)**.** *Let $m \geq \sqrt{n}$ and $2 \leq p \leq q \leq \infty$. Given matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \leq 1$, there exists $x \in \{\pm 1\}^n$ such that*

$$\left\| \sum_{i=1}^n x_i A_i \right\|_{S_q} \leq O\left(\sqrt{n \cdot \min(p, \max(1, \log(m/n)))}\right) \cdot \min(1, m/n)^{1/p - 1/q}).$$

When $m = n$, the right hand side is $O(\sqrt{n})$, and for diagonal matrices the conjecture is known to be true for any $2 \leq p \leq q$. When $p = q$, the conjecture is also known to be true for diagonal matrices for all m and n [RR20a].

4.1.2 Overview of Our Approach

We give a brief overview of our partial coloring framework in this subsection, and leave a more detailed discussion to Section 4.3.

Partial Coloring via Covering Numbers. Let $K := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\| \leq 1\}$ be the unit discrepancy body[2] and t be the target discrepancy bound. A recent refinement by Reis and Rothvoss [RR20a] of Gluskin's convex geometry approach [Glu89] shows that whenever $\gamma_n(tK) \geq 2^{-O(n)}$ for any constant in the exponent, one can efficiently find a partial coloring $x \in O(tK) \cap [-1,1]^n$ with at least $n/2$ coordinates in $\{-1,1\}$ (see Theorem 4.3.1). For settings where the target discrepancy bound is $n^{\Omega(1)}$, we may iterate the partial coloring to find a full coloring with the same discrepancy bound up to constants (see Corollary 4.3.2).

Our new approach for proving a Gaussian measure lower bound $\gamma_n(tK) \geq 2^{-O(n)}$ is via the covering numbers (Definition 4.2.2) of K or K° with respect to the Euclidean ball B_2^n or the ℓ_∞ ball B_∞^n. Informally, for convex bodies $A, B \subseteq \mathbb{R}^n$, we use $\mathcal{N}(A,B)$ to denote the minimum number of translates of B to cover A. In particular, since $\gamma_n(\sqrt{n}B_2^n)$ has constant Gaussian measure, as long as $\mathcal{N}(\sqrt{n}B_2^n, tK) \leq 2^{O(n)}$, we get $\gamma_n(tK) \geq 2^{-O(n)}$. Using the duality of covering numbers and connections with volume, we obtain several equivalent conditions for $\gamma_n(tK) \geq 2^{-O(n)}$ in terms of covering (Lemma 4.3.3). The condition that we will work with is $\mathcal{N}(K^\circ, \frac{t}{n}B_\infty^n) \leq 2^{O(n)}$, where $K^\circ = \{(\langle A_1, U\rangle, \ldots, \langle A_n, U\rangle) : \|U\|_* \leq 1\}$ is the polar discrepancy body.

Covering via Mirror Descent. We prove the covering number bound $\mathcal{N}(K^\circ, \frac{t}{n}B_\infty^n) \leq 2^{O(n)}$ using mirror descent, a powerful convex optimization primitive of Nemirovski and Yudin [NY83] (see Section 4.3.2 for an overview). In particular, denote the linear map $\mathcal{A}(U) := (\langle A_1, U\rangle, \ldots, \langle A_n, U\rangle)$. We shall assume that each $\|A_i\| \leq 1$. This is true for the matrix Spencer setting with $\|\cdot\|$ being the operator norm. In the more general setting of matrix discrepancy for Schatten norms, we have $\|A_i\|_{S_p} \leq 1$ while the norm for measuring

[2] To avoid confusion when talking about discrepancy bodies, K denotes the unit discrepancy body, and D denotes a scaling of K by the target discrepancy bound.

discrepancy is $\|\cdot\|_{S_q}$. One can get around this issue by leveraging known covering number estimates between Schatten classes (Theorem 4.2.6).

For any matrix $\|U\|_* \le 1$, consider minimizing the function $f_U(X) := \|\mathcal{A}(X-U)\|_\infty$ over the dual unit ball $B_* := \{U : \|U\|_* \le 1\}$. The function has minimum value $f_U(U) = 0$ and since it has subgradients in $\{\pm A_1, \dots, \pm A_n\}$ with $\|A_i\| \le 1$, the function $f_U(X)$ is 1-Lipschitz with respect to the dual norm $\|\cdot\|_*$. So as long as there exists a 1-strongly convex mirror map Φ on B_*, we can minimize $f_U(X)$ by starting from some matrix $U_0 = U_0(U) \in B_*$ and running mirror descent for n steps. Denoting by U_s the matrix in the s-th step, standard guarantees for mirror descent (Theorem 4.3.5) yield

$$\min_{s\in[n]} f_U(U_s) = \min_{s\in[n]} f_U(U_s) - f_U(U) \le \sqrt{\frac{2D_\Phi(U,U_0)}{n}}, \tag{4.1}$$

where $D_\Phi(U,U_0) = \Phi(U) - \Phi(U_0) - \langle \nabla\Phi(U_0), U - U_0\rangle$ is the Bregman divergence. We let T be the set of all matrices encountered when running mirror descent for all possible $U \in B_*$, i.e. $T := \{U_s : s \in [n], U \in B_*\}$, and $T_0 := \{U_0 : U \in B_*\}$ be the set of all starting matrices. The net $\mathcal{A}(T)$ will be our covering for K°.

To see that this indeed gives a good covering, we denote $D_\Phi^{\max} := \sup_{U\in B_*} D_\Phi(U\|U_0)$. By the definition of the function f_U, we have from (4.1) that

$$\min_{s\in[n]} \|\mathcal{A}(U) - \mathcal{A}(U_s)\|_\infty \le \sqrt{\frac{2D_\Phi(U,U_0)}{n}} \le \sqrt{\frac{2D_\Phi^{\max}}{n}},$$

and so the dual body admits the covering $K^\circ \subseteq \mathcal{A}(T) + \sqrt{2D_\Phi^{\max}/n}\cdot B_\infty^n$. Thus as long as our target discrepancy bound $t \le \sqrt{2nD_\Phi^{\max}}$, we have $\mathcal{N}(K^\circ, \frac{t}{n}B_\infty^n) \le |T|$, which we need to show to be at most $2^{O(n)}$.

The key observation we make here is that for our choices of the mirror maps in Sections 4.4 and 4.5, U_s only depends[3] on the sum of the subgradients, but not on their order. Since there

[3]In general, mirror descent projects back onto the feasible set according to the Bregman divergence in each iteration, and therefore might not satisfy this property.

are only $2n$ choices of subgradients $\{\pm A_i\}_{i\in[n]}$ and we run mirror descent for n steps, a counting argument reveals that there are at most $2^{O(n)}$ possible sums of gradients (Lemma 4.3.6). So long as the starting matrices satisfy $|T_0| \le 2^{O(n)}$, we have $|T| \le |T_0| \cdot 2^{O(n)} \le 2^{O(n)}$.

A View of Mirror Descent as Refining the Net. In the diagonal case, i.e. Spencer's setting, we can directly build the net T by repeatedly sampling the ith diagonal coordinate $e_i e_i^\top$ proportional to its weight in the target matrix. Since the set of diagonal matrices on the Schatten-1 ball has only $2m$ vertices $\{\pm e_i e_i^\top\}_{i\in[m]}$, the approximate Carathéodory theorem (see [Ver18], Theorem 0.0.2) implies that the image of the net $\mathcal{A}(T)$ already gives a good covering for K°, and mirror descent is not necessary in this case.

However, this argument fails beyond diagonal matrices, as the number of vertices becomes infinite. In these more general cases, we use mirror descent to boost a coarse net T_0 to a finer net T which has a better covering guarantee in the image space, at the expense of increasing the size of the net by a factor of $2^{O(n)}$.

Relative Entropy Nets for the Spectraplex. For our application in Section 4.5 to low-rank matrices, it suffices to take $T_0 = \{0\}$. For the application in Section 4.4 to block diagonal matrix Spencer, we run mirror descent on the spectraplex $\mathcal{S}_m := \{X \in \mathbb{R}^{m\times m} : X \succeq 0, \mathsf{tr}(X) = 1\}$ and carefully construct a set $|T_0| \le 2^{O(n)}$ with small $D_\Phi^{\max}$. Since $D_\Phi(X\|Y)$ is the quantum relative entropy between X and Y in the spectraplex setup, we refer to such T_0 as a (quantum) relative entropy net (Definition 4.3.7).

We use an operator norm net for the Schatten-1 ball from [HPV17] to construct a relative entropy net with error $O(\log(m^2/n))$ for the spectraplex $\mathcal{S}_m$ (Lemma 4.3.8). When restricted to block diagonal matrices with block size $h \times h$, we use a hybrid of this argument and the earlier approximate Caratheodory argument to find a refined relative entropy net with error $O(\log(hm/n))$ (Theorem 4.3.9). Taking T_0 to be this net in our mirror descent framework gives Theorem 4.1.3. This also allows us to reduce the matrix Spencer conjecture to the existence of a better relative entropy net with error $O(\log(m/n))$ for the spectraplex (Corollary 4.1.4).

4.1.3 Further Related Work

Banaszczyk's Approach. While the partial coloring method has been extensively applied in discrepancy and obtains the optimal bound for many problems, for several applications where the target discrepancy bound is $n^{o(1)}$ (e.g. the Komlós problem or Tusnady's problem), partial coloring is potentially sub-optimal by a logarithmic factor. In breakthrough work, Banaszczyk [Ban98] obtained an improvement over the partial coloring method for these applications using deep techniques from convex geometry. While Banaszczyk's original proof is non-constructive, a fascinating recent line of work has obtained algorithmic versions of Banaszczyk's result [DGLN16, BDG16, BG17, LRR17, BDGL18].

Matrix Spencer Conjecture and Non-commutative Random Matrix Theory. The typical value of $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}}$ for a random coloring has attracted significant attention in random matrix theory. For commutative matrices, the bound

$$\mathbb{E}\Big[\Big\|\sum_{i=1}^n x_i A_i\Big\|_{\mathsf{op}}\Big] \leq O(\sqrt{n \log m})$$

by matrix Khintchine [LPP91, Pis03] or matrix Chernoff bound [AW02] is in general tight. It is also known to be tight for Toeplitz matrices [Mec07]. For matrices with certain non-commutative structures (e.g. random Gaussian matrices), improved bounds of $O(\sqrt{n})$ are known (see [Ver18, BBvH21]). In the context of Conjecture 1.3.1, these results imply that a random coloring already achieves the conjectured bound when the input matrices have certain non-commutative structures. On the other hand, by Theorem 4.1.1, Conjecture 1.3.1 is known when all the matrices commute.

4.2 Preliminaries

Norms and Convex Bodies. A convex body is a compact convex set with non-empty interior. We say a convex set K is symmetric if $x \in K$ implies $-x \in K$. We use $\|\cdot\|_p$ to denote the ℓ_p-norm and $\|\cdot\|_{S_p}$ to denote the Schatten-p norm. In particular, the operator norm

$\|\cdot\|_{\mathsf{op}} = \|\cdot\|_{S_\infty}$ and the Frobenius norm $\|\cdot\|_F = \|\cdot\|_{S_2}$. We use B_p^n to denote the unit ℓ_p-ball in $\mathbb{R}^n$ and $B_{S_p}^n := \{A \in \mathbb{R}^{n\times n} : \|A\|_{S_p} \le 1\}$ to denote the unit Schatten-p ball in $\mathbb{R}^{n\times n}$, with $B_{\mathrm{op}}^n := B_{S_\infty}^n$. Let $\mathbb{R}_+^n$ denote the set of non-negative vectors in $\mathbb{R}^n$ and denote the simplex $\Delta_n := \{x \in \mathbb{R}_+^n : \|x\|_1 = 1\}$. Let $\mathbb{S}_+^n$ (resp. $\mathbb{S}_{++}^n$) denote the set of positive semidefinite (resp. positive definite) $n \times n$ matrices, and define the spectraplex $\mathcal{S}_n := \{X \in \mathbb{S}_+^n : \mathsf{tr}(X) = 1\}$. For a norm $\|\cdot\|$ in $\mathbb{R}^n$, we define the dual norm as $\|x\|_* := \sup\{\langle y, x\rangle : y \in \mathbb{R}^n, \|y\| \le 1\}$. Dual norms are similarly defined for matrix norms.

Convex Functions. A convex function $f : \mathcal{X} \to \mathbb{R}$ is said to be L-Lipschitz with respect to a norm $\|\cdot\|$ if $\|g\|_* \le L$ for all subgradients $g \in \partial f(x)$. We say that f is α-strongly convex with respect to a norm $\|\cdot\|$ if $f(y) \ge f(x) + g^\top(y - x) + \frac{\alpha}{2}\|x - y\|^2$, for all $x, y \in \mathcal{X}$ and all subgradients $g \in \partial f(x)$.

Polar. Given a convex set $K \subseteq \mathbb{R}^n$ with $0 \in K$, we define the polar of K to be $K^\circ := \{y \in \mathbb{R}^n : \sup_{x\in K}\langle x, y\rangle \le 1\}$. It is immediate from the definition that for any constant $t > 0$, $(tK)^\circ = \frac{1}{t}K^\circ$. When K is closed, the polarity theorem states that $(K^\circ)^\circ = K$.

Lemma 4.2.1 (Polar of Discrepancy Set)**.** *Given matrices $A_1, \ldots, A_n \in \mathbb{R}^{m\times m}$ and a norm $\|\cdot\|$ in $\mathbb{R}^{m\times m}$, we define the unit discrepancy set as $K := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\| \le 1\}$. Then $K' := \{(\langle A_1, U\rangle, \ldots, \langle A_n, U\rangle) : \|U\|_* \le 1\}$ is the polar body $K' = K^\circ$.*

Proof. By the definition of polar body, we may write

$$\begin{aligned}(K')^\circ &= \Big\{x \in \mathbb{R}^n : \sum_{i=1}^n x_i\langle A_i, U\rangle \le 1,\ \forall\ U \text{ s.t. } \|U\|_* \le 1\Big\}\\ &= \Big\{x \in \mathbb{R}^n : \Big\langle \sum_{i=1}^n x_i A_i, U\Big\rangle \le 1,\ \forall\ U \text{ s.t. } \|U\|_* \le 1\Big\}\\ &= K,\end{aligned}$$

by the definition of dual norm. It then follows from the polarity theorem that $K' = K^\circ$. □

Covering Numbers. We start with the definition of covering numbers.

Definition 4.2.2 (Covering Numbers)**.** *For two convex bodies $K, T \subseteq \mathbb{R}^n$, we define the covering number $\mathcal{N}(K,T)$ as the minimum number N such that there exist centers $x_1, \ldots, x_N \in \mathbb{R}^n$ with $K \subseteq \cup_{i=1}^N (x_i + T)$, i.e. K can be covered by N translates of T.*

We need the following few standard facts about covering numbers (see [AAGM15]).

Lemma 4.2.3 (Volume Bounds for Covering Numbers)**.** *Given convex bodies $K, T \subseteq \mathbb{R}^n$. If T is symmetric, we have $\frac{\mathsf{vol}_n(K)}{\mathsf{vol}_n(T)} \leq \mathcal{N}(K,T) \leq 2^n \cdot \frac{\mathsf{vol}_n(K+\frac{T}{2})}{\mathsf{vol}_n(T)}$.*

Lemma 4.2.4 (Symmetrization)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body, then $\mathcal{N}(K-K, K) \leq 2^{O(n)}$.*

Theorem 4.2.5 (Duality of Covering Numbers, [KM87])**.** *Given symmetric convex bodies $K, T \subseteq \mathbb{R}^n$, we have*

$$2^{-\Theta(n)} \cdot \mathcal{N}(T^\circ, K^\circ) \leq \mathcal{N}(K,T) \leq 2^{\Theta(n)} \cdot \mathcal{N}(T^\circ, K^\circ).$$

We will also need the following upper bound on the covering numbers of Schatten balls[4].

Theorem 4.2.6 ([HPV17], Theorem 1.1)**.** *Let $m, n \in \mathbb{N}$ and $1 \leq p \leq q \leq \infty$. Then we have*

$$\mathcal{N}\Big(B_{S_p}^m, \min\Big(1, \frac{m}{n}\Big)^{1/p-1/q} B_{S_q}^m\Big) \leq 2^{O(n)}.$$

Gaussian Measure. We use $\gamma_n(\cdot)$ to denote the standard Gaussian measure on $\mathbb{R}^n$. Gaussian measure is log-concave, i.e. $\gamma_n(\lambda A + (1-\lambda)B) \geq \gamma_n(A)^\lambda \gamma_n(B)^{1-\lambda}$ for any compact subsets $A, B \subseteq \mathbb{R}^n$. In particular, by taking $A = -x + K$ and $B = x + K$ for any $x \in \mathbb{R}^n$ and symmetric convex body K, and $\lambda = 1/2$, we have the following lemma.

Lemma 4.2.7 (Translation Decreases Gaussian Measure)**.** *Given any symmetric convex body $K \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$, we have $\gamma_n(K) \geq \gamma_n(x + K)$.*

We also use the following powerful Gaussian correlation inequality.

[4]We note that [HPV17] claims the bound only up to a constant depending on p and q, but their argument readily gives a universal constant in the regime of $p, q \geq 1$.

Theorem 4.2.8 (Gaussian Correlation Inequality, [Roy14])**.** *Given any symmetric convex sets $K, T \subseteq \mathbb{R}^n$, we have $\gamma_n(K \cap T) \geq \gamma_n(K) \cdot \gamma_n(T)$.*

4.3 Our Framework for Partial Coloring

4.3.1 Partial Coloring via Covering Numbers

Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$, a norm $\|\cdot\|$ on $\mathbb{R}^{m \times m}$ for measuring the discrepancy, and a target discrepancy bound t, let $D := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\| \leq t\}$ be the associated discrepancy body. The following partial coloring lemma from [RR20a] states that one can efficiently find a partial coloring with discrepancy $O(t)$ as long as $\gamma_n(D) \geq 2^{-O(n)}$.

Theorem 4.3.1 ([RR20a], special case of Theorem 6)**.** *For any constant $\alpha > 0$, there is a constant $c := c(\alpha) > 0$ and a randomized polynomial time algorithm that for a symmetric convex set $D \subseteq \mathbb{R}^n$ with $\gamma_n(D) \geq 2^{-\alpha n}$ and a shift $y \in (-1, 1)^n$, finds $x \in (c \cdot D) \cap [-1, 1]^n$ so that $x + y \in [-1, 1]^n$ and $|\{i \in [n] : |(x + y)_i| = 1\}| \geq n/2$.*

We have the following corollary for full colorings. Here $K_S := K \cap \{x \in \mathbb{R}^n : x_i = 0, \forall i \notin S\}$.

Corollary 4.3.2. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex set. Given a function $f : [n] \to \mathbb{R}_{>0}$ with $\gamma_S(f(|S|) \cdot K_S) \geq 2^{-O(|S|)}$ for every $S \subseteq [n]$, there exists a randomized polynomial time algorithm to find a full coloring $x \in \{\pm 1\}^n$ so that $x \in \lambda K$, where $\lambda \leq O(\sum_{i=0}^{\lfloor \log n \rfloor} f(n/2^i))$. In particular, when $f(n) \leq O(n^\beta)$ for some $\beta \leq 1$, we have $\lambda \leq O(\frac{1}{\beta} n^\beta)$.*

Proof. Indeed, repeated iterations of Theorem 4.3.1 with $y_0 := 0$ and subsequent shifts y_{i+1} being the coordinates not reaching $\{-1, 1\}$ find $x := x_0 + \cdots + x_T \in \{\pm 1\}^n$ for $T := \lfloor \log n \rfloor$ with $x_t \in O(f(n/2^t)) \cdot K$. When $f(n) \leq O(n^\beta)$, the summation is upper bounded by

$$\sum_{i=0}^{\infty} (n/2^i)^\beta = (1 - 2^{-\beta})^{-1} \cdot n^\beta \leq O(\frac{1}{\beta} \cdot n^\beta),$$

and this proves the statement. □

We show that a $2^{-O(n)}$ Gaussian measure lower bound is equivalent to a $2^{O(n)}$ upper bound for certain covering numbers.

Lemma 4.3.3. *The following conditions are equivalent for a symmetric convex body $D \subseteq \mathbb{R}^n$:*

1. $\gamma_n(D) \geq 2^{-O(n)}$,
2. $\mathcal{N}(\sqrt{n}B_2^n, D) \leq 2^{O(n)}$,
3. $\mathcal{N}(nB_1^n, D) \leq 2^{O(n)}$,
4. $\mathcal{N}(D^\circ, \frac{1}{\sqrt{n}}B_2^n) \leq 2^{O(n)}$,
5. $\mathcal{N}(D^\circ, \frac{1}{n}B_\infty^n) \leq 2^{O(n)}$.

Proof. We start by proving that condition (1) implies (2). Suppose $\gamma_n(D) \geq 2^{-O(n)}$, then Theorem 4.2.8 implies $\gamma_n(D') \geq 2^{-O(n)}$, where we define $D' := D \cap \sqrt{n}B_2^n$. We thus also have $\mathsf{vol}_n(D') \geq \gamma_n(D') \geq 2^{-O(n)}$. Then by Lemma 4.2.3, we have

$$\mathcal{N}(\sqrt{n}B_2^n, D) \leq \mathcal{N}(\sqrt{n}B_2^n, D') \leq 2^n \cdot \frac{\mathsf{vol}_n(\sqrt{n}B_2^n + D')}{\mathsf{vol}_n(D')} \leq 2^n \cdot \frac{\mathsf{vol}_n(2\sqrt{n}B_2^n)}{\mathsf{vol}_n(D')} \leq 2^{O(n)}.$$

We next show that condition (2) implies (1). Since $\gamma_n(\sqrt{n}B_2^n) = \Omega(1)$, we have $\gamma_n(x + D) \geq 2^{-O(n)}$ for some $x \in \mathbb{R}^n$. Lemma 4.2.7 then gives $\gamma_n(D) \geq \gamma_n(x + D) \geq 2^{-O(n)}$.

The implication (3) $\Rightarrow$ (2) immediately follows from $\sqrt{n}B_2^n \subseteq nB_1^n$. To prove the reverse implication (2) $\Rightarrow$ (3), we use Lemma 4.2.3 to obtain

$$\mathcal{N}(\sqrt{n}B_1^n, B_2^n) \leq 2^n \cdot \frac{\mathsf{vol}_n(\sqrt{n}B_1^n + B_2^n)}{\mathsf{vol}_n(B_2^n)} \leq 2^{O(n)} \cdot \frac{\mathsf{vol}_n(\sqrt{n}B_1^n)}{\mathsf{vol}_n(B_2^n)} \leq 2^{O(n)}.$$

It thus follows that $\mathcal{N}(nB_1^n, D) \leq \mathcal{N}(nB_1^n, \sqrt{n}B_2^n) \cdot \mathcal{N}(\sqrt{n}B_2^n, D) \leq 2^{O(n)}$.

The last two equivalences follow from the duality of covering numbers in Theorem 4.2.5.

□

For our mirror descent framework, we use the following corollary:

Corollary 4.3.4. *Given matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$, let $K_{q+}^\circ := \{(\langle A_1, U\rangle, \dots, \langle A_n, U\rangle) : U \in B_{S_q}^m, U \succeq 0\}$. If we have $\mathcal{N}(K_{q+}^\circ, \frac{t}{n}B_\infty^n) \le 2^{O(n)}$, then we can efficiently find a partial coloring $x \in [-1,1]^n$ with $|\{i : |x_i| = 1\}| \ge n/2$ and $\|\sum_{i=1}^n x_i A_i\|_{S_q} \lesssim t$.*

Proof. Recall that $D := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\|_{S_q} \le t\}$ denotes the discrepancy body. Since $tD^\circ \subseteq K_{q+}^\circ - K_{q+}^\circ$, by Lemma 4.2.4 we have $\mathcal{N}(D^\circ, \frac{1}{n}B_\infty^n) = 2^{O(n)}$. The equivalence (1) $\Leftrightarrow$ (5) in Lemma 4.3.3 implies $\gamma_n(D) \ge 2^{-O(n)}$, and Theorem 4.3.1 gives the corollary. □

4.3.2 Mirror Descent: An Overview

The mirror descent method was introduced by Nemirovski and Yudin [NY83]. Here, we follow the presentation in [Bub15]. Let $\mathcal{D}$ be an open subset of $\mathbb{R}^m$ and $\mathcal{X}$ a subset of its closure. We fix a convex function $f : \mathcal{X} \to \mathbb{R}$ assumed to be L-Lipschitz with respect to a norm $\|\cdot\|$, and a differentiable function $\Phi : \mathcal{D} \to \mathbb{R}$ that is ρ-strongly convex with respect to $\|\cdot\|$ and has a surjective gradient $\nabla\Phi : \mathcal{D} \to \mathbb{R}^m$. The mirror descent algorithm, given a starting point $x_0 \in \mathcal{X} \cap \mathcal{D}$, consists of the iterations

$$\nabla\Phi(y_{t+1}) := \nabla\Phi(x_t) - \eta g_t,$$
$$x_{t+1} := \operatorname{argmin}_{x\in\mathcal{X}\cap\mathcal{D}} D_\Phi(x, y_{t+1}),$$

where $g_t \in \partial f(x_t)$ and $D_\Phi(x,y) := \Phi(x) - \Phi(y) - \nabla\Phi(y)^\top(x-y)$ is the Bregman divergence. Note that $y_t \in \mathcal{D}$ and $x_t \in \mathcal{X}\cap\mathcal{D}$ for all $t \ge 0$. We use the following convergence guarantee:

Theorem 4.3.5 ([Bub15], Theorem 4.2)**.** *Let f be L-Lipschitz and Φ be ρ-strongly convex with respect to $\|\cdot\|$, and $D_\Phi^{\max} \ge D_\Phi(x^*, x_0)$ be any upper bound. Then the mirror descent algorithm with $\eta := \frac{1}{L}\sqrt{\frac{2\rho D_\Phi^{\max}}{T}}$ satisfies*

$$\min_{s\in[T]} f(x_s) - f(x^*) \le L\sqrt{\frac{2D_\Phi^{\max}}{\rho T}}.$$

The Spectraplex Setup. Here we take $\mathcal{X} := \mathcal{S}_m = \{X \in \mathbb{S}^m_+ : \mathsf{tr}(X) = 1\}$. The mirror map is $\Phi(X) = \mathsf{tr}(X \log X)$, defined on $\mathcal{D} = \mathbb{S}^m_{++}$, which is $\frac{1}{2}$-strongly convex with respect to the Schatten-1 norm by the quantum Pinsker inequality [Car16]. Then the convergence bound in Theorem 4.3.5 becomes $2L\sqrt{\frac{S(X^*\|X_0)}{T}}$, where $S(X\|Y) := \mathsf{tr}(X(\log X - \log Y))$ is the quantum relative entropy between matrices $X, Y \in \mathcal{S}_m$. The projection step corresponds to a trace normalization, so given a starting point $X_0 \in \mathcal{S}_m \cap \mathbb{S}^m_{++}$, we may write in closed form

$$X_{t+1} = \frac{\exp(\log X_0 - \eta \sum_{i=0}^t g_i)}{\mathsf{tr}(\exp(\log X_0 - \eta \sum_{i=0}^t g_i))}, \tag{4.2}$$

for subgradients $g_i \in \partial f(X_i)$.

The Schatten Norm Setup. Here we take $\mathcal{X} = \mathcal{D} = \mathbb{R}^{m\times m}$, so that $X_t = Y_t$ for all t. The mirror map is $\Phi(X) := \frac{1}{2(p-1)}\|X\|^2_{S_p}$, which is known to be 1-strongly convex for all $p \in (1, 2]$ [BCL94]. Thus given a starting point $X_0 \in \mathbb{R}^{m\times m}$, we may write in closed form

$$X_{t+1} = \nabla\Phi^{-1}\Big(\nabla\Phi(X_0) - \eta \sum_{i=0}^t g_i\Big), \tag{4.3}$$

for subgradients $g_i \in \partial f(X_i)$.

4.3.3 Covering via Mirror Descent

Given symmetric matrices $A_1, \ldots, A_n$ with $\|A_i\| \le 1$ for all $i \in [n]$, where the dual norm $\|\cdot\|_*$ is either the Schatten-1 norm or the Schatten-p norm for some $p \in (1, 2]$, we apply mirror descent on functions of the form $f_U(X) := \max_{i\in[n]} |\langle A_i, X - U\rangle|$ to cover the polar discrepancy body

$$K^\circ := \{\mathcal{A}(U) : \|U\|_* \le 1\}, \text{ where } \mathcal{A}(U) := (\langle A_1, U\rangle, \ldots, \langle A_n, U\rangle).$$

Note that $f_U(X) = \|\mathcal{A}(X) - \mathcal{A}(U)\|_\infty$ and that f is 1-Lipschitz with respect to $\|\cdot\|_*$. The key property of such functions is that we may always choose subgradients from the set of $2n$ matrices $\{\pm A_i : i \in [n]\}$, which allows us to upper bound the number of different matrices encountered during the mirror descent process.

Lemma 4.3.6. *Let $\|\cdot\|_*$ be either $\|\cdot\|_{S_1}$ as in the Spectraplex Setup, or $\|\cdot\|_{S_p}$ with $p \in (1, 2]$ as in the Schatten Norm Setup, and $\mathcal{X}, \mathcal{D}$ be defined accordingly. Let $T_0 \subseteq \mathcal{X} \cap \mathcal{D}$ be a set with size $|T_0| \le 2^{O(n)}$ and $K^\circ \supseteq K' = \mathcal{A}(T')$ the convex body to be covered, where $T' \subseteq \mathcal{X} \cap \mathcal{D}$. If for every $U \in T'$ there exists a starting point $U_0 := U_0(U) \in T_0$ with $D_\Phi(U, U_0) \le D_\Phi^{\max}$, then we can bound*

$$\mathcal{N}\Big(K', \sqrt{\frac{D_\Phi^{\max}}{n}} B_\infty^n\Big) \le 2^{O(n)}.$$

Proof. The key observation is that in either setup of mirror descent, the point X_t in (4.2) or (4.3) depends only on the starting point U_0 and on the sum of gradients $g_0, \ldots, g_{t-1}$, but not on their order. Moreover, we can always choose from the set of $2n$ gradients $\{\pm A_i : i \in [n]\}$ at each step. Thus applying mirror descent to the function f_U for all possible U with the same starting point U_0, the total number $N(U_0)$ of points visited in $T := n$ iterations satisfies

$$N(U_0) \le \sum_{t=0}^{n} \binom{t + 2n - 1}{2n - 1} \le (n + 1) \cdot \binom{3n}{n} \le 2^{O(n)}.$$

Since $|T_0| \le 2^{O(n)}$, we obtain a set of $2^{O(n)}$ points $\mathcal{U}$ such that for every $Y = \mathcal{A}(U) \in K'$, there exists some $\tilde{U} \in \mathcal{U}$ so that $\|\mathcal{A}(\tilde{U}) - \mathcal{A}(U)\|_\infty = f_U(\tilde{U}) = f_U(\tilde{U}) - f_U(U) \le O(\sqrt{D_\Phi^{\max}/n})$. □

In the Schatten Norm Setup, we shall pick $K' = K^\circ$ and $T_0 = \{0\}$, i.e. U_0 is always 0. For the Spectraplex Setup, we carefully choose a set of starting points $|T_0| \le 2^{O(n)}$ which has small $D_\Phi^{\max}$ with respect to $K' = \{\mathcal{A}(U) : U \in \mathcal{S}_m\}$. Since $D_\Phi(X\|Y)$ is the quantum relative entropy between X and Y in the Spectraplex Setup, we shall refer to the set of starting points T_0 as a (quantum) relative entropy net for $\mathcal{S}_m$.

Definition 4.3.7 (Quantum Relative Entropy Net)**.** *Given subsets $T, \mathcal{M} \subseteq \mathcal{S}_m$, T is a relative entropy net of $\mathcal{M}$ with error ε if for any $X \in \mathcal{M}$, we can find $Y \in T$ such that $S(X\|Y) \leq \varepsilon$.*

4.3.4 Initialization for Spectraplex Setup: Relative Entropy Net

We start with the following lemma which constructs a relative entropy net on $\mathcal{S}_m$ from an operator norm net.

Lemma 4.3.8 (Relative Entropy Net from Operator Norm Net)**.** *Let $X, Y \in \mathcal{S}_m$ satisfies $\|X - Y\|_{\mathsf{op}} \leq \varepsilon$ for some $\varepsilon \geq 1/m$. Then $S(X\|Y') \leq \log(2m\varepsilon)$, where $Y' := \frac{1}{2}(Y + \frac{I_m}{m}) \in \mathcal{S}_m$.*

Proof. Recall that $\log(\cdot)$ is operator monotone and note that $X \preceq Y + \varepsilon I_m$. We then have

$$\begin{aligned} S(X\|Y') &= \mathsf{tr}(X \cdot (\log X - \log Y')) \\ &\leq \mathsf{tr}(X \cdot (\log(Y + \varepsilon I_m) - \log Y')) \\ &\leq \mathsf{tr}(X) \cdot \|\log(Y + \varepsilon I_m) - \log Y'\|_{\mathsf{op}} \\ &\leq \log\left(2 \cdot \left\|\frac{Y + \varepsilon I_m}{Y + \frac{I_m}{m}}\right\|_{\mathsf{op}}\right) \leq \log(2m\varepsilon), \end{aligned}$$

where the first inequality follows from the operator monotonicity of $\log(\cdot)$, the second follows from matrix Hölder, and the last follows because $\varepsilon \geq 1/m$ and $\|Y\|_{\mathsf{op}} \leq 1$. □

Using the lemma above, we give the following construction for relative entropy nets on $\mathcal{S}_m$.

Theorem 4.3.9 (Entropy Net for Spectraplex)**.** *Given positive integers h, m and n such that m/h is an integer, let $\mathcal{S}_m^h \subseteq \mathcal{S}_m$ be the set of $m \times m$ block diagonal matrices on the spectraplex with block size $h \times h$. Then we can find a relative entropy net T for $\mathcal{S}_m^h$ with error at most $\max(1, \log(2hm/n))$ and size $|T| \leq 2^{O(n)}$.*

Proof. By merging blocks as needed, we may assume $hm \geq n$. By Lemma 4.3.8, it suffices to find an operator norm net T' with size $|T'| \leq 2^{O(n)}$ and distance $\varepsilon = \frac{\max\{h, \log(m/hn)\}}{n}$. Let $\ell := m/h$ be the number of blocks, $X_1, \ldots, X_\ell \in \mathbb{R}^{h \times h}$ denote the blocks of matrix $X \in \mathcal{S}_m^h$, and $N := 2/\varepsilon = 2n/\max\{h, \log(\ell/n)\}$ (we assume that N is an integer). Let $Z := \{z \in \mathbb{Z}_{\geq 0}^\ell : \sum_{i=1}^\ell z_i = N\}$, and for each $z \in Z$, we define

$$T_z := \{X \in \mathcal{S}_m^h : \mathsf{tr}(X_i) = z_i/N, \forall i \in [\ell]\}.$$

It follows from a standard rounding argument that for any matrix $X \in \mathcal{S}_m^h$, one can find a matrix $Y \in \cup_{z \in Z} T_z$ with $\|X - Y\|_{\mathsf{op}} \leq 1/N = \varepsilon/2$.

We first show that $|Z| \leq 2^{O(n)}$. When $\ell \leq 2n$, we have

$$|Z| \leq \binom{N+\ell}{\ell} \leq \binom{N+2n}{2n} \leq \binom{\frac{2n}{h}+2n}{2n} \leq 2^{O(n)}.$$

When $\ell \geq 2n \geq N$, we can bound

$$|Z| \leq \binom{N+\ell}{N} \leq \binom{2\ell}{N} \leq \binom{2\ell}{\frac{2n}{\log(\ell/n)}} \leq \left(\frac{e\ell \log(\ell/n)}{n}\right)^{\frac{2n}{\log(\ell/n)}} \leq 2^{O(n)}.$$

It therefore suffices to construct an $\varepsilon/2$-operator norm net for each T_z.

Fix an arbitrary $z \in Z$. Note that the ith block of the matrices in T_z comes from $\frac{z_i}{N} \cdot \mathcal{S}_h$. Pick $n_i := z_i h$, we have from Theorem 4.2.6 that

$$\mathcal{N}\Big(\frac{z_i}{N}\mathcal{S}_h, \frac{z_i}{N} \cdot \frac{h}{n_i} B_{\mathsf{op}}^h\Big) = \mathcal{N}\Big(\mathcal{S}_h, \frac{h}{n_i} B_{\mathsf{op}}^h\Big) \leq 2^{O(n_i)}.$$

We denote this net as $\widetilde{T}_{z,i}$. It follows from the above that for any $X_i \in \frac{z_i}{N}\mathcal{S}_h$, there exists $Y_i \in \widetilde{T}_{z,i}$ with $\|X_i - Y_i\|_{\mathsf{op}} \leq \frac{z_i}{N} \cdot \frac{h}{n_i} = \varepsilon/2$. Define $\widetilde{T}_z := \{\mathsf{diag}(Y_1, \ldots, Y_\ell) : Y_i \in \widetilde{T}_{z,i}\ \forall i \in [\ell]\}$. Then for any $X \in T_z$, there exists $Y \in \widetilde{T}_z$ such that $\|X - Y\|_{\mathsf{op}} \leq \varepsilon/2$, and thus $\widetilde{T}_z$ is indeed

an $\varepsilon/2$-operator norm net for T_z. Furthermore, the size of $\widetilde{T}_z$ can be upper bounded as

$$|\widetilde{T}_z| \le \prod_{i\in[\ell]} 2^{O(n_i)} = 2^{O(\sum_{i=1}^n z_i h)} = 2^{O(Nh)} \le 2^{O(n)},$$

since $N \le 2n/h$. This proves that $\widetilde{T} := \cup_{z\in Z}\widetilde{T}_z$ is an ε-operator norm net for $\mathcal{S}_m^h$ and has size at most $|\widetilde{T}| \le 2^{O(n)}$, where we recall that $\varepsilon = \frac{\max\{h,\log(m/hn)\}}{n}$. Finally, invoking Lemma 4.3.8, $\widetilde{T}$ can be transformed into a relative entropy net T with size $|T| \le 2^{O(n)}$ and error at most $\log(2m\varepsilon) \le \log(2hm/n)$. This finishes the proof of the theorem. □

4.4 Applications of the Spectraplex Setup

In this section, we prove our matrix Spencer bound for block diagonal matrices in Theorem 4.1.3, which we restate below.

Theorem 4.1.3 (Matrix Spencer for Block Diagonal Matrices)**.** *Let $m \ge \sqrt{n}$ and $h \le m$. Given block diagonal symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$ with $\|A_i\|_{\mathsf{op}} \le 1$ and block size $h \times h$, one can efficiently find a coloring $x \in \{\pm 1\}^n$ with*

$$\Big\|\sum_{i=1}^n x_i A_i\Big\|_{\mathsf{op}} \le O(\sqrt{n \cdot \max(1, \log(hm/n))}).$$

Proof of Theorem 4.1.3. By Theorem 4.3.9, we can find a relative entropy net T_0 of $\mathcal{S}_m^h$ with error $D_\Phi^{\max} := \max(1, \log(2hm/n))$ and size $|T_0| \le 2^{O(n)}$. Then using Lemma 4.3.6 with the Spectraplex Setup for $K' := \mathcal{A}(\mathcal{S}_m^h)$ and T_0 being the relative entropy net, we obtain

$$\mathcal{N}\Big(K', \frac{t}{n}B_\infty^n\Big) \le 2^{O(n)},$$

where $t = \sqrt{n\max(1, \log(2hm/n))}$. Let $\mathbb{S}_m^h$ be the set of $m \times m$ symmetric block diagonal matrices with block size $h \times h$. Define convex body $K'' := \mathcal{A}(B_{S_1}^m \cap \mathbb{S}_m^h \cap \mathbb{S}_+^m)$. We first prove that $\mathcal{N}(K'', \frac{t}{n}B_\infty^n) \le 2^{O(n)}$. Since $\mathcal{N}(K', \frac{t}{n}B_\infty^n) \le 2^{O(n)}$ by Theorem 4.3.9, we also have $\mathcal{N}(\frac{j}{n^2}K', \frac{t}{n}B_\infty^n) \le 2^{O(n)}$ for each integer $j \in [n^2]$. We let H_j be the set of centers

for the minimum covering of $\frac{j}{n^2}K'$ by translates of $\frac{t}{n}B_\infty^n$ and define $H = \cup_{j\in[n^2]}H_j$. Since $|H_j| \le 2^{O(n)}$, it follows that $|H| \le 2^{O(n)}$. For each $X \in B_{S_1}^m$ that satisfies $X \succeq 0$, we let $\frac{j}{n^2}$ be the multiple of $\frac{1}{n^2}$ that is closest to $\mathsf{tr}(X)$, and set $X' := \frac{j}{n^2\mathsf{tr}(X)} \cdot X$. Then we have

$$\|\mathcal{A}(X') - \mathcal{A}(X)\|_\infty \le \frac{1}{n^2} \cdot \|\mathcal{A}(X)\|_\infty \le \frac{t}{n}.$$

As $\mathsf{tr}(X') = \frac{j}{n^2}$, we can also find $Y \in H_j$ with $\|\mathcal{A}(X') - Y\|_\infty \le \frac{t}{n}$. Therefore, $\|\mathcal{A}(X) - Y\|_\infty \le \frac{2t}{n}$, and it follows that $K'' \subseteq H + \frac{2t}{n}B_\infty^n$. This implies $\mathcal{N}(K'', \frac{t}{n}B_\infty^n) \le 2^{O(n)}$.

Next note that the dual discrepancy body is $K^\circ := \mathcal{A}(B_{S_1}^m) = \mathcal{A}(B_{S_1}^m \cap \mathbb{S}_m^h)$ since each $A_i \in \mathbb{S}_m^h$. We have $K^\circ = K'' - K''$, so using Lemma 4.2.4 we get $\mathcal{N}(K^\circ, K'') \le 2^{O(n)}$. Thus

$$\mathcal{N}\Big(K^\circ, \frac{t}{n}B_\infty^n\Big) \le \mathcal{N}(K^\circ, K'') \cdot \mathcal{N}\Big(K'', \frac{t}{n}B_\infty^n\Big) \le 2^{O(n)},$$

and $\gamma_n(tK) \ge 2^{-O(n)}$ by using Lemma 4.3.3. Corollary 4.3.2 then gives a full coloring $x \in \{\pm 1\}^n$ with discrepancy $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} \le O(t)$. This finishes the proof of the theorem. □

The analysis above also shows that if we can improve the bound in Theorem 4.3.9 to $O(\log(m/n))$ for any block size h, then the matrix Spencer conjecture is true.

Corollary 4.1.4 (Better Entropy Net Implies Matrix Spencer)**.** *Let $m \ge \sqrt{n}$. If we can find $T \subseteq \mathcal{S}_m$ with $|T| \le 2^{O(n)}$ such that for each $X \in \mathcal{S}_m$ there exists $Y \in T$ with $S(X\|Y) \le O(\max(1, \log(m/n)))$, where $S(X\|Y)$ is the quantum relative entropy between X and Y, then the matrix Spencer conjecture is true.*

4.5 Matrix Discrepancy for Schatten Norms

In this section, we prove the following generalization of Theorem 4.1.2 for arbitrary Schatten norms by using a different regularizer for mirror descent.

Theorem 4.1.5 (Matrix Discrepancy for Schatten Norms)**.** *Let $m \ge \sqrt{n}$ and $2 \le p \le q \le \infty$. Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m\times m}$ with $\|A_i\|_{S_p} \le 1$ and $\mathrm{rank}(A_i) \le r$ for*

all $i \in [n]$, one can efficiently find $x \in [-1,1]^n$ so that $|\{i : |x_i| = 1\}| \geq n/2$ and

$$\Big\| \sum_{i=1}^n x_i A_i \Big\|_{S_q} \leq O(\sqrt{n \cdot \min(p, \max(1, \log(rk)))} \cdot k^{1/p-1/q}),$$

where we denote $k := \min(1, m/n)$. Moreover, we can find a full coloring $x \in \{\pm 1\}^n$ at the expense of a factor of $(1/2 + 1/q - 1/p)^{-1}$.

We first use mirror descent to prove the following covering lemma.

Lemma 4.5.1. *Let $m \geq \sqrt{n}$, $2 \leq p \leq q < \infty$, $k := \min(1, m/n)$, $t := \sqrt{(p-1)n} \cdot k^{1/p-1/q}$ and $q^* := q/(q-1)$. Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \leq 1$, we have*

$$\mathcal{N}\Big(\mathcal{A}(B^m_{S_{q^*}}), \frac{t}{n} B^n_\infty\Big) \leq 2^{O(n)}.$$

Proof. Denote $p^* := p/(p-1)$. Theorem 4.2.6 implies $\mathcal{N}(\mathcal{A}(B^m_{S_{q^*}}), k^{1/q^*-1/p^*}\mathcal{A}(B^m_{S_{p^*}})) \leq 2^{O(n)}$, so it suffices to show

$$\mathcal{N}\Big(\mathcal{A}(B^m_{S_{p^*}}), \sqrt{\frac{p-1}{n}} B^n_\infty\Big) \leq 2^{O(n)}.$$

This is a direct consequence of Lemma 4.3.6 with norm $\|\cdot\|_{S_{p^*}}$, as the Bregman divergence is $D_\Phi(U, 0) = \Phi(U) \leq \frac{1}{2(p^*-1)} = \frac{p-1}{2}$ for $\|U\|_{S_{p^*}} \leq 1$. □

Lemma 4.5.1 together with Lemma 4.3.3 immediately gives the following weaker measure bound, which we then bootstrap to prove the stronger bound in Theorem 4.1.5.

Corollary 4.5.2. *Let $m \geq \sqrt{n}$, $2 \leq p \leq q < \infty$ and $k := \min(1, m/n)$. Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \leq 1$, define the convex body*

$$K := \Big\{ x \in \mathbb{R}^n : \Big\| \sum_{i=1}^n x_i A_i \Big\|_{S_q} \leq 1 \Big\}.$$

Then $\gamma_n(\sqrt{(p-1)n} \cdot k^{1/p-1/q} \cdot K) \geq 2^{-O(n)}$.

Proof of Theorem 4.1.5. Let $p_0 := \max(2, \log(2rk))$. For $p \leq p_0$ the result follows directly from Corollary 4.5.2, so we may assume $p \geq p_0$. Also note that we may assume $rk \geq 1$ since we can increase smaller values of r without changing the bound on the right side. Remark that $\|A_i\|_{S_{p_0}} \leq r^{1/p_0-1/p}\|A_i\|_{S_p} \leq r^{1/p_0-1/p}$ since the matrices have rank at most r. Corollary 4.5.2 then implies that the convex body

$$\sqrt{p_0 n} \cdot k^{1/p_0-1/q} \cdot r^{1/p_0-1/p} \cdot K$$

has Gaussian measure $2^{-O(n)}$. Since $\sqrt{p_0 n} \cdot k^{1/p_0-1/q} \cdot r^{1/p_0-1/p} \leq O(\sqrt{p_0 n} \cdot k^{1/p-1/q})$ by the choice of p_0, it follows that

$$\gamma_n(\sqrt{n \max(1, \log(rk))} \cdot k^{1/p-1/q} \cdot K) \geq 2^{-O(n)},$$

so that Theorem 4.3.1 and Corollary 4.3.2 yield the partial coloring and full coloring, respectively. The factor $(1/2 + 1/p - 1/q)^{-1}$ comes from the contribution of the exponent of n in the geometric sum, analogous to the second part of Corollary 4.3.2. □

4.6 Lower Bound Examples for Matrix Discrepancy

In this section, we give a few examples to illustrate the tightness of our results in Theorem 4.1.5 for various regimes of the dimension m and rank r of the input matrices.

4.6.1 Low Dimension Regime of $m = \Theta(\sqrt{n})$

In the regime of $m = \Theta(\sqrt{n})$, we have $k = \min(1, m/n) = \Theta(1/\sqrt{n})$ and $r \leq O(\sqrt{n})$ and our partial coloring bound in Theorem 4.1.5 is thus $O(n^{1/2+1/2q-1/2p})$. This bound is tight up to constants due to the following example[5].

[5]Thanks to Aleksandar Nikolov for suggesting this construction.

Lemma 4.6.1 (Example: $m = \sqrt{n}$)**.** *Let $m = \sqrt{n}$ be a power of 2, and $2 \le p \le q \le \infty$. There exist matrices*[6] *$A_1, \dots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \le 1$ such that $\|\sum_{i=1}^n x_i A_i\|_{S_q} \ge \Omega(n^{1/2+1/2q-1/2p})$ for any partial coloring $x \in \{\pm 1\}^n$ with $|\{i : |x_i| = 1\}| \ge n/2$.*

Proof. The idea is to construct an orthogonal basis on $\mathbb{R}^{m \times m}$ with $\|A_i\|_F^2 = m$. Let $H \in \mathbb{R}^{m \times m}$ be the Walsh-Hadamard matrix, and $D_1, \dots, D_m$ be diagonal matrices with $(D_i)_{j,j} := H_{i,j}$. Let $P_1, \dots, P_m$ be disjoint permutation matrices, i.e. each P_i permutes the standard orthonormal basis $\{e_1, \dots, e_m\}$ and each pair P_i, P_j have disjoint non-zero entries. For instance, we may take $(P_i)_{j,k} := 1$ if $j - k \equiv i \mod m$ and 0 otherwise. We then define the n matrices $A_{i+mj} := D_i P_j$ for $i, j \in [m]$. Note that these matrices form an orthogonal basis of $\mathbb{R}^{m \times m}$, so for any partial coloring $x \in \{\pm 1\}^n$ with $|\{i : |x_i| = 1\}| \ge n/2$, we have

$$\left\| \sum_{i=1}^n x_i A_i \right\|_F^2 = \mathsf{tr}\left(\left(\sum_{i=1}^n x_i A_i \right)^2 \right) = m \cdot \sum_{i=1}^n x_i^2 \ge mn/2.$$

By Hölder's inequality, this implies that

$$\left\| \sum_{i=1}^n x_i A_i \right\|_{S_q} \ge m^{1/q-1/2} \cdot \left\| \sum_{i=1}^n x_i A_i \right\|_F \ge \Omega(n^{1/2+1/2q}).$$

Also note that each matrix A_i has all singular values equal to 1, and therefore $\|A_i\|_{S_p} = m^{1/p} = n^{1/2p}$. Scaling the matrices A_i down by a factor of $n^{1/2p}$ proves the lemma. □

4.6.2 Rank-1 Matrices and $m \ge n$

In the regime of $r = 1$ and $m \ge n$, we may assume wlog that $p = 2$. Then the discrepancy bound in Theorem 4.1.2 is $O(\sqrt{n})$. This bound is again tight up to a constant factor.

Lemma 4.6.2 (Example: $r = 1$ and $m = n$)**.** *Let $2 \le q \le \infty$. There exist symmetric rank-1 matrices $A_1, \dots, A_n \in \mathbb{R}^{n \times n}$ with $\|A_i\|_F \le 1$ such that any partial coloring $x \in [-1, 1]^n$ with*

[6] These matrices can easily be made symmetric in $\mathbb{R}^{2m \times 2m}$.

$|\{i : |x_i| = 1\}| \geq n/2$ *satisfies*

$$\left\| \sum_{i=1}^{n} x_i A_i \right\|_{S_q} \geq \left\| \sum_{i=1}^{n} x_i A_i \right\|_{\mathsf{op}} \geq \Omega(\sqrt{n}).$$

Proof. For each $i \in [n-1]$, we define the rank-1 matrices $A_i := \frac{1}{2}(e_i + e_n)(e_i + e_n)^\top$ for $i \in [n]$, where $e_i \in \mathbb{R}^n$ is the unit vector with a single 1 in the ith coordinate and 0 elsewhere, and $A_n = 0$. Note that each $\|A_i\|_F = 1$ by definition. For any partial coloring $x \in [-1, 1]^n$ with $|\{i : |x_i| = 1\}| \geq n/2$, we have

$$\sum_{i=1}^{n} x_i A_i = \frac{1}{2} \cdot \begin{pmatrix} x_1 & 0 & \cdots & 0 & x_1 \\ 0 & x_2 & \cdots & 0 & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & x_{n-1} & x_{n-1} \\ x_1 & x_2 & \cdots & x_{n-1} & \sum_{i=1}^{n-1} x_i \end{pmatrix}.$$

It then follows that

$$\left\| \sum_{i=1}^{n} x_i A_i \right\|_{\mathsf{op}} \geq \left\| \sum_{i=1}^{n} x_i A_i e_n \right\|_2 \geq \Omega(\sqrt{n}).$$

This completes the proof of the lemma. □

As an immediate corollary of Lemma 4.6.2, we obtain an $\Omega(\sqrt{n})$ lower bound for matrix Spencer when $m = n$ and all matrices are rank-1.

Corollary 4.1.6 (Rank-1 Matrix Spencer Lower Bound)**.** *There exist rank-1 symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{n \times n}$ with $\|A_i\|_{\mathsf{op}} \leq 1$ such that any $x \in \{\pm 1\}^n$ has $\| \sum_{i=1}^{n} x_i A_i \|_{\mathsf{op}} \geq \Omega(\sqrt{n})$.*

Another immediate consequence of Lemma 4.6.2 is a lower bound of $\Omega(\sqrt{\min(m, n)})$ for Schatten-2 to operator norm discrepancy, which is the generalization of the Komlós problem

to matrices. This shows that the Komlós conjecture, which states that the ℓ_2 to ℓ_∞ vector discrepancy is upper bounded by a universal constant, cannot be true for matrices.

Corollary 4.1.7 (Lower Bound for Matrix Komlós)**.** *For any m and n, there exist symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$ with $\|A_i\|_F \le 1$ such that any $x \in \{\pm 1\}^n$ has $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} \ge \Omega(\sqrt{\min(m,n)})$.*

4.7 An Application of Banaszczyk's Theorem

We give an alternative simpler proof of the $O(m^{1+1/q-1/p})$ bound for S_p to S_q matrix discrepancy when $m = O(\sqrt{n})$ using the following theorem of Banaszczyk [Ban98].

Theorem 4.7.1 (Banaszczyk [Ban98])**.** *Let $K \subseteq \mathbb{R}^m$ be a convex body with $\gamma_m(K) \ge 1/2$. Then for any vectors $v_1, \dots, v_n \in \mathbb{R}^m$ with $\|v_i\|_2 \le 1$, there exists $x \in \{\pm 1\}^n$ such that $\sum_{i=1}^n x_i v_i \in 5K$.*

Applying Theorem 4.7.1 to a suitable scaling of the operator norm ball immediately gives the following matrix discrepancy bound.

Corollary 4.7.2. *Let $2 \le p \le q \le \infty$. Given matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$ with $\|A_i\|_{S_p} \le 1$, there exists $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i A_i\|_{S_q} \le O(m^{1+1/q-1/p})$.*

Proof. Note that $\|A_i\|_{S_p} \le 1$ implies $\|A_i\|_{S_2} \le m^{1/2-1/p}$. It is well-known that $\gamma_m(4m^{1/2} \cdot B_{\mathsf{op}}^m) \ge 1/2$ (see Theorem 7.3.1 of [Ver18]). Thus, Theorem 4.7.1 yields some $x \in \{\pm 1\}^n$ such that $\sum_{i=1}^n x_i A_i \in O(m^{1-1/p}) \cdot B_{\mathsf{op}}^m$. It follows that $\|\sum_{i=1}^n x_i A_i\|_{S_q} \le O(m^{1+1/q-1/p})$. □

Corollary 4.7.3 (Matrix Komlós)**.** *Given matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$ with $\|A_i\|_F \le 1$, there exists $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i A_i\|_{S_q} \le O(\sqrt{\min(m,n)})$, matching the lower bound in Corollary 4.1.7.*

Proof. It suffices to take the best between a random coloring, which has discrepancy $O(\sqrt{n})$, and that of Corollary 4.7.2. □

Chapter 5

MATRIX DISCREPANCY II: THE INVERSE POLYNOMIAL BARRIER METHOD

In this chapter, we continue our study of the matrix Spencer conjecture (Conjecture 1.3.1) and present an elementary approach for this problem based on the inverse polynomial barrier potential function. In particular, we show that this approach gives a stronger result than the ones in Chapter 4 and an independent and concurrent work [HRS22]. This chapter is based on an unpublished joint work with Nikhil Bansal and Raghu Meka [BJM22a].

But before we present our new approach, for the convenience of the readers and for the current chapter to be self-contained, let us briefly recall the motivations and definitions that have already been presented in the preceding chapter.

5.1 Introduction

Let us start with the classical discrepancy setting where given vectors $a_1, \ldots, a_n \in \mathbb{R}^m$ satisfying $\|a_i\|_\infty \leq 1$ for all $i \in [n]$, and the goal is to find signs $x_1, \ldots, x_n \in \{\pm 1\}$ to minimize the discrepancy $\|\sum_{i=1}^n x_i a_i\|_\infty$. In a celebrated result, Spencer showed that the $O(\sqrt{n \log m})$ bound obtained by a random coloring is not tight and showed the following bound, which is also the best possible in general.

Theorem 4.1.1 (Spencer [Spe85]). *Let $m \geq n$. Given vectors $a_1, \ldots, a_n \in \mathbb{R}^m$ with $\|a_i\|_\infty \leq 1$, there exists $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i a_i\|_\infty \leq O(\sqrt{n \log(2m/n)})$.*

In particular for $m = O(n)$, this gives an $O(\sqrt{n})$ bound, in contrast to the $O(\sqrt{n \log n})$ bound for random coloring obtained by applying Chernoff and union bounds.

To prove this result, Spencer developed the powerful partial-coloring method via the entropy method, building on previous work of Beck [Bec81]. Another approach to prove

Theorem 4.1.1 based on convex geometry was developed independently by Gluskin [Glu89]. While these original arguments used the pigeonhole principle and were non-algorithmic, in recent years there has been a rich line of work [Ban10, BS13, LM15a, Rot17, LRR17, ES18, RR20a] on their algorithmic versions.

Matrix Spencer Setting. A natural generalization of Spencer's problem to matrices is the following. Let $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ be symmetric matrices with maximum singular value, or operator norm, $\|A_i\|_{\mathsf{op}} \leq 1$. Find a coloring $x \in \{\pm 1\}^n$ that minimizes $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}}$. In particular, Spencer's result corresponds to the case when all the $A_i = \mathsf{diag}(a_i)$ are diagonal.

As in the vector case, for a random coloring $x \in \{\pm 1\}^n$, the matrix Chernoff bound of Ahlswede and Winter [AW02], generalizing the scalar Chernoff type bounds, gives that

$$\mathbb{E}\Big[\big\|\sum_i x_i A_i\big\|_{\mathsf{op}}\Big] = O\Big(\sqrt{\log m} \cdot \Big\|\sum_i A_i^2\Big\|_{\mathsf{op}}^{1/2}\Big).$$

This implies a bound of $O(\sqrt{n \log m})$ on the matrix discrepancy. Matrix concentration bounds are powerful and widely used tools in mathematics and computer science, and it is natural to ask when one can beat them. In particular, the following natural analogue of Spencer's result for matrices has received considerable attention recently. $\big\|\sum_{i=1}^n x_i A_i\big\|_{\mathsf{op}} \leq O(\sqrt{n} \cdot \max\{1, \sqrt{\log(m/n)}\})$.

Conjecture 1.3.1 (Matrix Spencer Conjecture, [Zou12, Mek14])**.** *Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with each $\|A_i\|_{\mathsf{op}} \leq 1$, there exist signs $x \in \{\pm 1\}^n$ such that $\big\|\sum_{i=1}^n x_i A_i\big\|_{\mathsf{op}} \leq O(\sqrt{n} \cdot \max\{1, \sqrt{\log(m/n)}\})$. In particular, the matrix discrepancy is $O(\sqrt{n})$ for $m = n$.*

While this conjecture is still open, there has been interesting progress on important special cases.

Low-Rank Matrices. Recently, Hopkins, Raghavendra and Shetty [HRS22] established an ingenious connection between matrix discrepancy and quantum communication complex-

ity, and used sophisticated methods from quantum information theory and sketching to show that Conjecture 1.3.1 holds for matrices of rank $O(\sqrt{n})$. More generally, they show the following.

Theorem 5.1.1. *(Moderate-Rank Matrix Spencer [HRS22]) Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ each satisfying $\|A_i\|_{\mathsf{op}} \leq 1$ and Frobenius norm $\|A_i\|_F \leq n^{1/4}$, there exist signs $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} \leq O(\sqrt{n})$. Moreover, these signs can be computed efficiently.*

It also known that the discrepancy must be $\Omega(\sqrt{n})$, even for rank-1 matrices [DJR22].

This result significantly improves upon the previous $O(\sqrt{n \log r})$ bound based on the non-commutative Khinchine inequalities [LP86, LPP91, Pis03]. A more refined $O(\sqrt{\log r} \cdot \|\sum_i A_i^2\|_{\mathsf{op}}^{1/2})$ bound[1] was given by Kyng, Luh and Song [KLS20] for $r = 1$, based on the breakthrough work of Marcus, Spielman and Srivastava on the Kadison-Singer problem [MSS15], and extended to general r by Song and Zhang [SZ20]. Making these results algorithmic is an outstanding open question. Recently, Dadush, Jiang and Reis [DJR22] gave a $O(\sqrt{n \log(rm/n)})$ bound based on a convex geometric approach. This improves upon the $O(\sqrt{n \log r})$ bound from non-commutative Khinchine inequalities for $m \ll n$, but is weaker than the bound in Theorem 5.1.1.

Notice, however, that the rank condition on the matrices A_i can be quite restrictive, e.g., already in Spencer's classical setting, the diagonal matrices can have rank $\Omega(n)$. In fact, for diagonal matrices with rank $\sqrt{n}$, a substantially better $O(n^{1/4}\sqrt{\log n})$ bound follows from Banaszczyk's result on the Komlós problem [Ban98].

Block-Diagonal Matrices. A natural generalization of the diagonal setting in Spencer's result is the setting of block-diagonal matrices. Let us say that a symmetric matrix $A \in \mathbb{R}^{m \times m}$ is h-block diagonal if it can be written as $A = \mathsf{diag}(B_1, \ldots, B_{m/h})$, where each B_j is a symmetric $h \times h$ matrix. For such matrices, Levy, Ramadas and Rothvoss [LRR17] and

[1] However this can be as large as $\sqrt{n \log r}$ in general.

Dadush, Jiang and Reis [DJR22] showed the following result[2].

Theorem 5.1.2 (Matrix Spencer for Block-Diagonal Matrices [LRR17, DJR22])**.** *Given h-block diagonal symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ each satisfying $\|A_i\|_{\mathsf{op}} \leq 1$, one can efficiently find a coloring $x \in \{\pm 1\}^n$ with*

$$\Big\| \sum_i x_i A_i \Big\|_{\mathsf{op}} \leq O(\sqrt{n} \cdot \max\{1, \sqrt{\log(hm/n)}\}).$$

This generalizes Theorem 4.1.1, which corresponds to $h = 1$.

Notice that Theorems 5.1.1 and 5.1.2 handle very different types of matrices: Theorem 5.1.2 only uses the block-diagonal structure and works for arbitrary rank. On the other hand, Theorem 5.1.1 only uses the rank condition, and does not seem to give anything better if matrices have additional properties such as block-structure or if m is small.

Their proofs also use very different techniques — convex duality, quantum communication lower bounds and sketching in [HRS22], Gaussian volume and covering number bounds for geometric bodies in [DJR22], and a variant of the multiplicative weight approach in [LRR17]. This makes it unclear how to combine these ideas and obtain results for more general classes of matrices.

5.1.1 Our Results

An Elementary Approach for All State-of-the-Art Results. In this work we show that all the known state-of-the-art results for the Matrix Spencer problem, and more, can be obtained in an elementary and unified way using a barrier-based potential function approach. This unified approach also allows us to prove the Matrix Spencer Conjecture for a wider class of matrices that simultaneously generalizes the results in both Theorems 5.1.1 and 5.1.2.

We develop this barrier approach for the matrix discrepancy setting in Section 5.3, and give a general condition that expresses the matrix discrepancy in terms of simple parameters

[2]The result of [LRR17] requires the mild restriction that $h \leq \sqrt{n}$.

of the matrices. In Section 5.4.1, we show how the results for block-diagonal matrices in [LRR17, DJR22] follow directly from these conditions. Next, in Section 5.4.2 we use this framework to obtain the bounds of [HRS22]. We first show the following version of Theorem 5.1.1.

Theorem 5.1.3 (Matrix Spencer for Moderate-Rank Matrices, Theorem 1.3 in [HRS22])**.** *Given symmetric matrices $A_1, \cdots, A_n \in \mathbb{R}^{m\times m}$ with $\|A_i\|_{\mathsf{op}} \le 1$ and $\sum_{i=1}^n \|A_i\|_F^2 \le nf$, then one can efficiently find a coloring $x \in \{\pm 1\}^n$ with*

$$\|\sum_i x_i A_i\|_{\mathsf{op}} \le O\left(\sqrt{n}\cdot \max\{1, \sqrt{\log(f^2/n)}\}\right).$$

In its most general form, [HRS22] showed the following bound for partial coloring. Note that Theorem 5.1.4 is scale-invariant and does not require the assumption $\|A_i\|_{\mathsf{op}} \le 1$.

Theorem 5.1.4 (Theorem 3.1 in [HRS22])**.** *Given symmetric matrices $A_1, \cdots, A_n \in \mathbb{R}^{m\times m}$, one can efficiently find a partial coloring $x \in [-1,1]^n$ with $|\{i : x_i \in \{\pm 1\}\}| = \Omega(n)$ such that*

$$\|\sum_i x_i A_i\|_{\mathsf{op}} \le O\left(\left\|\sum_i A_i^2\right\|_{\mathsf{op}}^{1/2} \cdot \max\left\{1, \sqrt{\log\left(\frac{\sum_i \mathsf{tr}(A_i^2)}{\sqrt{n}\|\sum_i A_i^2\|_{\mathsf{op}}}\right)}\right\}\right).$$

In Section 5.6 we show how this follows from the proof of Theorem 5.1.3.

We also remark that our barrier-based algorithm is deterministic, while the previous algorithms in [DJR22] and [HRS22] are both randomized.

A More General Matrix Discrepancy Bound. Next, in Section 5.5, we describe a more general setting that combines both the moderate-rank and block-diagonal settings as special cases, and show the following more general matrix discrepancy bound.

Theorem 5.1.5 (General Matrix Discrepancy Bound)**.** *Given block-diagonal symmetric matrices $A_i = \mathsf{diag}(D_i^1, \cdots, D_i^\ell)$ with diagonal blocks $D_i^j \in \mathbb{R}^{h_j\times h_j}$. If $\|D_i^j\|_{\mathsf{op}} \le 1$ and $\sum_{i=1}^n \|D_i^j\|_F^2 \le gn$ for all $i \in [n]$ and $j \in [\ell]$, then one can efficiently find a coloring*

$x \in \{\pm 1\}^n$ *with*

$$\left\| \sum_i x_i A_i \right\|_{\mathsf{op}} \leq O\left(\sqrt{n} \cdot \max\{1, \sqrt{\log(g^2\ell/n)}\}\right).$$

For the moderate-rank case, setting $\ell = 1$ recovers Theorem 5.1.3 (and therefore Theorem 5.1.1). For the block-diagonal case, writing the bound in Theorem 5.1.2 as $O(\sqrt{n} \cdot \max\{1, \sqrt{\log(h^2\ell/n)}\})$ (since $m = \ell h$), we see that the bound of $O(\sqrt{n} \cdot \max\{1, \sqrt{\log(g^2\ell/n)}\})$ in Theorem 5.1.5 can only be better as $\|D\|_F^2 \leq h$ for any $h \times h$ block D with $\|D\|_{\mathsf{op}} \leq 1$, and hence $g \leq h$.

Note that the bound in Theorem 5.1.5 can be strictly stronger than both bounds in Theorem 5.1.3 and Theorem 5.1.2 simultaneously. For instance, consider $m = n$ and each matrix $A_i = \mathsf{diag}(D_i^1, \cdots, D_i^\ell)$ is h-block diagonal with $h = \sqrt{n}$ (i.e. there are $\ell = \sqrt{n}$ diagonal blocks), and each $\|D_i^j\|_F^2 = \sqrt{h} = n^{1/4}$. For this parameter setting, $f = \sqrt{h}\ell = n^{3/4}$ in Theorem 5.1.3 and thus the bound in Theorem 5.1.3 is $O(\sqrt{n \log n})$; $hm = n^{3/2}$ so the bound in Theorem 5.1.2 is also $O(\sqrt{n \log n})$. However, notice that $g = \sqrt{h} = n^{1/4}$ and $\ell = \sqrt{n}$, and therefore Theorem 5.1.5 gives a bound of $O(\sqrt{n})$, matching that of Conjecture 1.3.1 for this parameter setting.

5.1.2 Technical Overview

The algorithm proceeds by maintaining a fractional coloring $x(t) \in [-1, 1]^n$ over time. Initially, $x(0)$ is the all-zero coloring and eventually all variables reach ± 1. Once a variable reaches ± 1 it is no longer *alive* and is not updated any more. Using standard ideas, it suffices to bound the discrepancy for a partial coloring, where the number of alive variables reduces from n to $n/2$[3].

For a fractional coloring $x \in [-1, 1]^n$ at time t, let $D(t) = \sum_{i=1}^n x_i A_i$ denote the current discrepancy. To ensure a discrepancy bound of $\|D(t)\|_{\mathsf{op}} \leq \sqrt{b(t)}$, we adopt the standard

[3]As our discrepancy bounds scale as $\approx n^{1/2}$, the total discrepancy over $\log n$ partial coloring phases only is only $O(1)$ factor worse than in the first phase.

barrier function approach. We define the $m \times m$ slack matrix

$$S(t) = b(t) \cdot I_m - D(t)^2,$$

for square discrepancy and ensure that $S(t) \succ 0$ for all time step t as the coloring evolves. To do this, we consider the potential function

$$\Phi(t) = \mathsf{tr}(S(t)^{-p})$$

for a suitable parameter $p \geq 1$, and ensure that the potential can only decrease over time as the coloring is updated. This suffices as $S(0) = b(t) \cdot I_m$ initially, and thus $\Phi(t) \leq \Phi(0) = m \cdot b(0)^{-p}$ for all t which ensures that the eigenvalues of $D(t)^2$ never get too close to $b(t)$.

Updating the Coloring. Fix some time t. We now describe how to update the coloring when t advances by Δt so that the potential does not increase. Let $\Delta x(t)$ denote the update of $x(t)$, and $\Delta D(t) = \sum_i \Delta x_i(t) A_i$ be the corresponding discrepancy update. We will also increase $b(t)$ by $\Delta b(t) = c(t)$. To guarantee progress, $\Delta x(t)$ is chosen orthogonal to $x(t)$ so that $\|x(t)\|_2$ monotonically increases. For simplicity, we drop the dependence on t below when the context is clear.

Now Φ decreases due to the increase of b and could possibly increase due to the update ΔD, so our goal is to choose ΔD (via Δx) such that the increase is less than the decrease. Let $0 < s_1 \leq \cdots \leq s_m$ be the eigenvalues of S with corresponding eigenvectors $q_1, \ldots, q_m$. It turns out that the small s_j contribute relatively more to $\Delta\Phi$ upon the update ΔD. Roughly speaking, the contribution to $\Delta\Phi$ due to s_j is proportional to $-c(t) + \|(\Delta D)q_j\|^2 \cdot p/s_j$.

Blocking small s_i. So the first main idea is to simply *block* the small s_j by choosing Δx so that $(\Delta D)q_j = (\sum_i \Delta x_i A_i)q_j = \mathbf{0}$. Since $q_j \in \mathbb{R}^m$, blocking each such s_j gives m linear constraints (and h if the matrices are block-diagonal) in the variables Δx_i. As there are roughly n such variables for partial coloring, we can block the smallest $J = n/2m$

eigenvalues s_i, while keeping $n/2$ degrees of freedom for choosing Δx. Then, averaging over a random choice of $\Delta x \perp x$ respecting the above constraints lead to a condition on $c(t)$ that is sufficient for potential decrease (see (5.6) and (5.9) for details). Setting the parameters $c(t)$ and p appropriately, this gives a partial coloring discrepancy bound of $O(\sqrt{n \log(m/|J|)})$ in a generic way.

It is useful to compare this with random coloring, which incurs discrepancy $O(\sqrt{n \log m})$. So this ability to block the smallest J directions crucially gives us the improvement.

Recovering the State-of-the-Art Bounds. In the block-diagonal setting, as we can block $J = n/2h$ smallest s_j, we have $m/J = 2hm/n$ and the generic framework above gives the bound $O(\sqrt{n \log(hm/n)})$, which proves Theorem 5.1.2.

Theorem 5.1.1 requires more care. Let us suppose for the discussion here that $m = n$ and the matrices A_i have rank $f = \sqrt{n}$, and our goal is to prove the $O(\sqrt{n})$ bound. As $m = n$, a priori the generic framework above only lets us block $J = n/m = O(1)$ constraints, which is not at all useful. To get around this, we will show how to use the rank condition to effectively reduce $m \approx f$, in which case the generic framework will give the bound $O(\sqrt{n \log(f^2/n)}) \approx O(\sqrt{n})$.

Let us denote $M = \sum_i A_i^2$ and note that $\|M\|_{\mathsf{op}} \leq n$ as $\|A_i\|_{\mathsf{op}} \leq 1$. Now, if we had that $\|M\|_{\mathsf{op}} \leq n/\log m$, then a random coloring would already work, as by matrix Chernoff bounds $\|\sum_i x_i A_i\|_{\mathsf{op}} \approx \sqrt{\log m} \cdot \|M\|_{\mathsf{op}} \leq \sqrt{n}$ with high probability. So suppose that $\|M\|_{\mathsf{op}} > n/\log m$. But since $\mathsf{tr}(M) \leq nf$, at most $m_0 = f \log m = \tilde{O}(\sqrt{n})$ eigenvalues of M can be larger than $n/\log m$. Intuitively, this suggests that only these m_0 (heavy) directions should require careful handling, which may allow us to pretend that A_i essentially behave like $m_0 \times m_0$ matrices. If so, applying the generic framework above to obtain $O(\sqrt{n \log(m_0^2/n)})) \approx O(\sqrt{n})$ discrepancy.

The intuition above is over-simplified and not quite correct, but it can be made precise by decomposing the matrices A_i into different classes $k \geq 0$. We refer to Section 5.4.2 for details, but roughly we decompose each A_i into certain L-shaped pieces $L_{k,i}$ for classes

$k = 1, \ldots, O(\log n)$, corresponding to eigenvalues of M of size $n/10^k$, i.e. $\|\sum_i (L_{k,i})^2\|_{\mathsf{op}} \leq n/10^k$. We then apply the generic framework to all these classes simultaneously with different parameters $b_k(0)$, $c_k(t)$ and p_k and k different potential functions $\Phi_k(t)$, so that the resulting partial coloring satisfies $\|\sum_i x_i L_{k,i}\|_{\mathsf{op}} \leq O(2^{-k/2}\sqrt{n})$ for each class k simultaneously, The $O(\sqrt{n})$ bound then follows as $\|\sum_i x_i A_i\|_{\mathsf{op}} \leq \sum_k \|\sum_i x_i L_{k,i}\|_{\mathsf{op}}$ by the triangle inequality.

A More General Matrix Discrepancy Bound. The two main ideas of blocking small s_i and dividing into different weight classes allow us to cleanly interpolate between the block-diagonal setting in Theorem 5.1.2 and the moderate-rank setting in Theorem 5.1.1. In particular, if the matrices $A_i = \mathsf{diag}(D_i^1, \cdots, D_i^\ell)$ are both block-diagonal and satisfy the moderate-rank condition $\sum_{i=1}^n \|D_i^j\|_F^2 \leq gn$ for all $i \in [n]$ and $j \in [\ell]$, then one can essentially use the same multi-class potential analysis but can block the small s_i's with fewer constraints due to the block-diagonal structure. This leads to the general matrix discrepancy bound in Theorem 5.1.5 that is stronger than both Theorem 5.1.1 and Theorem 5.1.2. We leave the details to Section 5.5.

5.1.3 Further Related Works

Discrepancy Theory. Discrepancy theory is widely studied and has applications to many other areas in mathematics and computer science. We refer readers to the excellent books [Cha00, Mat99, CST+14] for a more comprehensive account of the rich history of discrepancy theory. Recent developments of discrepancy theory lead to numerous applications in approximation algorithms, differential privacy, fair allocation, experimental design and more [MN12, Rot13, NTZ13, BCKL14b, BN17, JKS19, HSSZ19, BJSS20, BRS22].

Matrix Discrepancy and Non-Commutativity Random Matrix Theory. Many natural problems can be viewed as questions about matrix discrepancy, e.g. graph sparsification [BSS12, RR20b], the Kadison-Singer problem [MSS15] and its generalization [KLS20], and the design of quantum random access codes [ANTSV02, HRS22].

Matrix discrepancy is also closely related to non-commutative random matrix theory, where the typical value of $\|\sum_i x_i A_i\|_{\mathsf{op}}$ for a random coloring x has received significant attention. The bound of $\mathbb{E}[\|\sum_i x_i A_i\|_{\mathsf{op}}] \leq O(\sqrt{n \log m})$ by matrix Chernoff [AW02] or matrix Khintchine [LP86, LPP91, Pis03] that is generally tight for commutative matrices, can be often improved in the non-commutative case (e.g. [Ver18, BBvH21] and the references therein). We refer readers to the book [Tao12, Vu14] for a more comprehensive account of random matrix theory.

Barrier Potential Function Approach. The inverse polynomial barrier potential function was first used in the seminal work of Batson, Spielman and Srivastava on graph sparsification [BSS12]. Similar potential functions have also been used in the context of discrepancy [BS13, BS20]. Recently, Bansal, Laddha and Vempala [BLV22] showed how various state-of-the-art results in vector discrepancy can be obtained using this method. We remark that the way the barrier potential function is typically used in discrepancy theory is different from that in [BSS12]. In discrepancy, one typically obtains a fractional update on the coloring, while in [BSS12] a rank-1 update is incurred from adding a single vector.

5.2 Preliminaries

We recall some basic facts about matrices and describe the notation. For a square matrix $A \in \mathbb{R}^{m \times m}$ with entries a_{ij}, its trace $\mathsf{tr}(A) = \sum_i a_{ii}$ and Frobenius norm $\|A\|_F = \sqrt{\mathsf{tr}(A^T A)} = (\sum_{ij} a_{ij}^2)^{1/2}$. If A is symmetric with eigenvalues $\lambda_1, \ldots, \lambda_n$, then we have $\mathsf{tr}(A) = \sum_i \lambda_i$, $\|A\|_F = (\sum_i \lambda_i^2)^{1/2}$ and its operator norm $\|A\|_{\mathsf{op}} = \max_{\|x\|_2=1} \|Ax\|_2 = \max_i |\lambda_i|$. For a rectangular matrix R its operator norm $\|R\|_{\mathsf{op}} = \max_{\|x\|_2=1} \|Rx\|_2 = \sqrt{\|R^T R\|_{\mathsf{op}}}$.

By the spectral theorem, any symmetric matrix A can be written as $A = QDQ^T$ where D is a diagonal matrix with entries $D_{ii} = \lambda_i$ and Q is an orthogonal matrix where the columns of Q are the corresponding unit eigenvectors of A. A symmetric matrix A is positive semidefinite (PSD) if all its eigenvalues $\lambda_i \geq 0$, and denote this as $A \succeq 0$ and $A \succ 0$ if the λ are strictly positive. For symmetric matrices $A, B \in \mathbb{R}^{m \times m}$, we have the partial ordering

$A \preceq B$ if $B - A \succeq 0$.

A standard computation together with the cyclic property of trace $\mathsf{tr}(ABC) = \mathsf{tr}(CAB)$, gives the following.

Lemma 5.2.1 (Directional Derivatives of $\mathsf{tr}(X^{-p})$)**.** *Let $X \in \mathbb{R}^{m\times m}$ be positive definite and $p > 1$, then the first and second order directional derivatives of the function $\Phi(X) = \mathsf{tr}(X^{-p})$ are given by*

$$D\Phi(X)[H] = -p \cdot \mathsf{tr}(X^{-(p+1)}H)$$
$$D^2\Phi(X)[H_1, H_2] = p \cdot \sum_{k=1}^{p+1} \mathsf{tr}(X^{-k}H_1X^{-(p+2-k)}H_2).$$

We need the following generalized Lieb-Thirring inequality. The proof below is from [Eld13].

Lemma 5.2.2 (Generalized Lieb-Thirring)**.** *Given a symmetric matrix B, a PSD matrix A and $\alpha \in [0, 1]$, we have $\mathsf{tr}(A^\alpha BA^{1-\alpha}B) \le \mathsf{tr}(AB^2)$.*

Proof. Working in the eigenbasis of A, wlog we can assume that $A = \mathsf{diag}(a_1, \cdots, a_n)$ is diagonal. Then by the AM-GM inequality,

$$\begin{aligned}\mathsf{tr}(A^\alpha BA^{1-\alpha}B) &= \sum_{i,j} a_i^\alpha a_j^{1-\alpha} B_{i,j}^2 \le \sum_{i,j} (\alpha a_i + (1-\alpha)a_j)B_{i,j}^2 \\ &= \alpha \sum_{i,j} a_i B_{i,j}^2 + (1-\alpha)\sum_{i,j} a_j B_{i,j}^2 = \mathsf{tr}(AB^2).\end{aligned}$$

□

5.3 Barrier Potential for Matrix Discrepancy: A Meta Analysis

To handle both the moderate-rank setting of Theorem 5.1.1 and the block-diagonal setting in Theorem 5.1.2, we need to work with the more general matrix discrepancy setting for rectangular matrices. In particular, given rectangular matrices $R_1, \cdots, R_n \in \mathbb{R}^{m\times m'}$, the goal is to find a coloring $x \in \{\pm 1\}^n$ to minimize the discrepancy $\|\sum_{i=1}^n x_i R_i\|_{\mathsf{op}}$.

As is standard for Spencer-type result, it suffices to find a partial coloring $x \in [-1,1]^n$ with $|\{i : x_i \in \{\pm 1\}\}| = \Omega(n)$ that has small discrepancy.

5.3.1 The Barrier Potential Function for Squared Discrepancy

The algorithm starts with an initial coloring $x(0)$ and updates $x(t)$ at each time t based on a barrier potential function Φ that controls the squared discrepancy $\|\mathcal{R}(x(t))^\top \mathcal{R}(x(t))\|_{\mathsf{op}}$, where $\mathcal{R}(x) := \sum_{i=1}^n x_i R_i$. We will assume for simplicity that $x(0)$ is the all-zero coloring, as this will not affect any of the arguments below.

Let $b(t) > 0$ be a target upper bound for the squared discrepancy at time t. Let the slack matrix $S(t) \in \mathbb{S}^{m\times m}$ be defined as

$$S(t) := b(t) \cdot I_m - \mathcal{R}(x(t))^\top \mathcal{R}(x(t)), \tag{5.1}$$

and the potential function

$$\Phi(t) = \mathsf{tr}(S(t)^{-p}),$$

where the parameter $p \geq 1$ will be specified later. Note that initially $\Phi(0) = m \cdot b(0)^{-p}$. The algorithm will inductively maintain the conditions $S(t) \succ 0$ and $\Phi(t) \leq \Phi(0)$, at each time t.

Suppose these conditions hold at time t. We update $b(t)$ and $x(t)$ as follows. Advancing time by Δt, we update $b(t)$ by $\Delta b(t) := b(t+\Delta t) - b(t) = c(t)\Delta t$ where $c(t) \geq 0$, and the coloring by $\Delta x(t) := x(t+\Delta t) - x(t) = \varepsilon(t) v(t) \sqrt{\Delta t}$, where $\varepsilon(t)$ is a ± 1 sign and $v(t) \in \mathbb{R}^n$ is a unit vector, that will be chosen suitably so that the change in potential $\Delta\Phi(t) := \Phi(t+\Delta t) - \Phi(t) \leq 0$.

We will also require that $v(t) \perp x(t)$ which ensures that $\|x(t+\Delta t)\|_2^2 = \|x(t)\|_2^2 + \Delta t$, and thus $\|x(t)\|_2^2 - \|x(0)\|_2^2 = t$ for all t.

For simplicity, we drop (t) from our notations below when the context is clear.

Change in the Potential. Now we analyze the change in the potential function. Using Lemma 5.2.1 and second-order Taylor expansion, we have

$$\begin{aligned}\Delta\Phi &\approx -p\cdot\mathsf{tr}(S^{-(p+1)}\Delta S)+\frac{p}{2}\cdot\sum_{k=1}^{p+1}\mathsf{tr}(S^{-k}\Delta S S^{-(p+2)}\Delta S)\\ &\le -p\cdot\mathsf{tr}(S^{-(p+1)}\Delta S)+\frac{p(p+1)}{2}\cdot\mathsf{tr}(S^{-(p+2)}(\Delta S)^2),\end{aligned}$$

where the second inequality follows by applying Lemma 5.2.2 with $B=\Delta S$ and $A=S^{-(p+2)}$ and $\alpha=k/(p+2)$ for $k=1,\dots,p+1$. As is completely standard (see e.g., Lemma A.1 in [BLV22]), the error in the approximation of $\Delta\Phi$ in the first step via second-order Taylor expansion can be made negligible by choosing small enough $\Delta t=n^{-\Omega(1)}$.

As $R(x+\Delta x)=R(x)+R(v)\varepsilon\sqrt{\Delta t}$ and $\Delta b=c\Delta t$, the slack in (5.1) changes as

$$\Delta S=(cI_m-\mathcal{R}(v)^\top\mathcal{R}(v))\Delta t-(\mathcal{R}(v)^\top\mathcal{R}(x)+\mathcal{R}(x)^\top\mathcal{R}(v))\varepsilon\sqrt{\Delta t}.$$

Choosing $\varepsilon\in\{\pm1\}$ uniformly at random gives

$$\frac{\mathbb{E}_\varepsilon[\Delta\Phi]}{p\Delta t}\le-\mathsf{tr}(S^{-(p+1)}(cI-\mathcal{R}(v)^\top\mathcal{R}(v)))+\frac{p+1}{2}\cdot\mathsf{tr}(S^{-(p+2)}(Z+Z^\top)^2),$$

where we denote $Z:=\mathcal{R}(v)^\top\mathcal{R}(x)$.

Let $0<s_1\le\cdots\le s_m$ be the eigenvalues of S, and consider its spectral decomposition $S=Q^\top\mathsf{diag}(\{s_j\}_{j=1}^m)Q$ for some orthogonal matrix $Q\in\mathbb{R}^{m\times m}$. Let $\widetilde{R}_i:=R_iQ$ and $\widetilde{\mathcal{R}}(x):=\mathcal{R}(x)Q$ be the matrices after diagonalizing S, and similarly let $\widetilde{Z}:=\widetilde{\mathcal{R}}(v)^\top\widetilde{\mathcal{R}}(x)$. Then

$$\frac{\mathbb{E}_\varepsilon[\Delta\Phi]}{p\Delta t}\le-\sum_{j=1}^m s_j^{-(p+1)}(c-(\widetilde{\mathcal{R}}(v)^\top\widetilde{\mathcal{R}}(v))_{j,j})+\frac{p+1}{2}\sum_{j=1}^m s_j^{-(p+2)}((\widetilde{Z}+\widetilde{Z}^\top)^2)_{j,j}.\tag{5.2}$$

We next show how to find a random vector $v\perp x$ (with support of size at most n) such that the RHS of (5.2) is at most 0 in expectation. This will imply that there exists a deterministic choice of v and $\varepsilon\in\{\pm1\}$, which can be efficiently computed, for which Φ does not increase.

For clarity, we will use j to index the coordinates in $[m]$ and i to index those in $[n]$.

5.3.2 Blocking Small Eigenvalues of the Slack Matrix

Notice that the first term in (5.2) scales as $s_j^{-(p+1)}$, while the second term scales as $s_j^{-(p+2)}$. This can be problematic when some of the s_j become tiny, as the second summand can be much larger than that the first, leading to an increase in the potential. To prevent this, we "block" these small s_j by choosing v such that $((\widetilde{Z}+\widetilde{Z}^\top)^2)_{j,j}=0$. Since $\widetilde{Z}=Q^\top ZQ$, this condition is equivalent to $(Z+Z^\top)Q_j=\mathbf{0}$, where Q_j is the jth column of Q.

Expanding out, each such j gives the following constraints on v,

$$\mathbf{0}=(Z+Z^\top)Q_j=\sum_{i=1}^{n} v_i(R_i^\top\mathcal{R}(x)Q_j+\mathcal{R}(x)^\top R_iQ_j). \tag{5.3}$$

As $R_i^\top\mathcal{R}(x)Q_j+\mathcal{R}(x)^\top R_iQ_j\in\mathbb{R}^m$, in general (5.3) gives a linear system for v with m constraints.

Remark 5.3.1. *When the matrices R_i have certain sparsity pattern (e.g. they are all block-diagonal matrices), then (5.3) can have much fewer than m constraints. Later, we will exploit this sparsity structure to recover the state-of-the-art bounds matrix discrepancy for block-diagonal matrices in [LRR17, DJR22], as well as to prove our more general bound in Theorem 5.1.5.*

The Meta-Algorithm. A coordinate $i\in[n]$ is called *alive* if $x_i\in[-1+1/n,1-1/n]$. Let n_t be the number of alive coordinates at time t, and we assume wlog these are the first n_t coordinates. As we only update the alive coordinates, we may view v as a vector in $\mathbb{R}^{n_t}$ by ignoring all its zero coordinates in $\{n_t+1,\dots,n\}$.

We block the smallest few s_j so that the total number of constraints in (5.3) from all blocked s_j is at most $n_t/2-1$. Let J_t (which will be specified in each of the later sections) be the number of s_j that can be blocked in this manner, and W be the subspace orthogonal to

these constraints and to the current coloring vector x. Note that $\dim(W) \geq n_t - (J_t + 1) \geq n_t/2$.

Recalling our convention that $0 < s_1 \leq \cdots \leq s_m$, for any vector $v \in W$, we thus have

$$\begin{aligned}\frac{\mathbb{E}_\varepsilon[\Delta\Phi]}{p\Delta t} &\leq -\sum_{j=1}^{m} s_j^{-(p+1)}(c - (\widetilde{\mathcal{R}}(v)^\top \widetilde{\mathcal{R}}(v))_{j,j}) + \frac{p+1}{2}\sum_{j=J_t+1}^{m} s_j^{-(p+2)}((\widetilde{Z} + \widetilde{Z}^\top)^2)_{j,j} \\ &\leq -\sum_{j=1}^{m} s_j^{-(p+1)}(c - (\widetilde{\mathcal{R}}(v)^\top \widetilde{\mathcal{R}}(v))_{j,j}) + (p+1)\sum_{j=J_t+1}^{m} s_j^{-(p+2)}(\widetilde{Z}\widetilde{Z}^\top + \widetilde{Z}^\top\widetilde{Z})_{j,j} \quad (5.4)\end{aligned}$$

where the second inequality uses $(\widetilde{Z} + \widetilde{Z}^\top)^2 \preceq 2(\widetilde{Z}\widetilde{Z}^\top + \widetilde{Z}^\top\widetilde{Z})$.

Choosing v. Let $u_1, \cdots, u_{\dim(W)}$ be an orthonormal basis of the subspace W, we will set v to be one of $u_1, \cdots, u_{\dim(W)}$ uniformly at random. So we have

$$\mathbb{E}[vv^\top] = \frac{1}{\dim(W)} \cdot I_W \preceq \frac{2I}{n_t}.$$

Next we analyze the expected potential change due to such a random choice of v.

5.3.3 Potential Decrease via an Averaging Argument

Let $\widetilde{R}(x)_j$ be the jth column of the matrix $\widetilde{R}(x)$. Since $S = b \cdot I_m - \mathcal{R}(x)^\top\mathcal{R}(x) \succ 0$, we have $\|\widetilde{R}(x)^\top\widetilde{R}(x)\|_{\mathsf{op}}, \|\widetilde{R}(x)\widetilde{R}(x)^\top\|_{\mathsf{op}} < b$, which in particular implies $\|\widetilde{R}(x)_j\|_2^2 < b$. Then we have

$$\begin{aligned}\mathbb{E}_v[(\widetilde{Z}\widetilde{Z}^\top)_{j,j}] &= \mathbb{E}_v[\widetilde{\mathcal{R}}(v)_j^\top\widetilde{\mathcal{R}}(x)\widetilde{\mathcal{R}}(x)^\top\widetilde{\mathcal{R}}(v)_j] \leq b \cdot \mathbb{E}_v[\|\widetilde{\mathcal{R}}(v)_j\|_2^2] \\ &= b \cdot Q_j^\top \mathbb{E}_v[\mathcal{R}(v)^\top\mathcal{R}(v)]Q_j \leq \frac{2b}{n_t} \cdot \Big\|\sum_{i=1}^{n} R_i^\top R_i\Big\|_{\mathsf{op}},\end{aligned}$$

where the last inequality uses $\mathbb{E}[vv^\top] \preceq (2/n_t)I$, and $\|Q_j\|_2 = 1$ as Q is orthogonal. Similarly,

$$\mathbb{E}_v[(\widetilde{Z}^\top\widetilde{Z})_{j,j}] = \widetilde{\mathcal{R}}(x)_j^\top \mathbb{E}_v[\widetilde{\mathcal{R}}(v)\widetilde{\mathcal{R}}(v)^\top]\widetilde{\mathcal{R}}(x)_j \leq \frac{2b}{n_t} \cdot \Big\|\sum_{i=1}^{n} R_i R_i^\top\Big\|_{\mathsf{op}}.$$

Let $\sigma(\mathcal{R})^2 := \max\left\{\left\|\sum_{i=1}^n R_i^\top R_i\right\|_{\mathsf{op}}, \left\|\sum_{i=1}^n R_i R_i^\top\right\|_{\mathsf{op}}\right\}$ be the variance parameter for the matrices $R_1, \cdots, R_n$. Then the above bounds imply that

$$\max\left\{\mathbb{E}_v[(\widetilde{Z}\widetilde{Z}^\top)_{j,j}], \mathbb{E}_v[(\widetilde{Z}^\top\widetilde{Z})_{j,j}]\right\} \leq \frac{2b}{n_t}\cdot\sigma(\mathcal{R})^2.$$

Also note that as $\mathbb{E}_v[vv^T] \preceq (2/n_t)I$, we have that $\mathbb{E}_v[(\widetilde{\mathcal{R}}_v^\top\widetilde{\mathcal{R}}_v)_{j,j}] \leq (2/n_t)\cdot\sigma(\mathcal{R})^2$.

Plugging these bounds into (5.4), we obtain

$$\frac{\mathbb{E}_{\varepsilon,v}[\Delta\Phi]}{p\Delta t} \leq -\sum_{j=1}^m s_j^{-(p+1)}\left(c - \frac{2}{n_t}\cdot\sigma(\mathcal{R})^2\right) + \frac{4b(p+1)}{n_t}\cdot\sigma(\mathcal{R})^2\sum_{j=J_t+1}^m s_j^{-(p+2)}. \tag{5.5}$$

Then it follows that to satisfy $\mathbb{E}_{\varepsilon,v}[\Delta\Phi] \leq 0$, it suffices to choose

$$c(t) \geq \frac{2}{n_t}\cdot\sigma(\mathcal{R})^2 + \frac{4b(t)\cdot(p+1)}{n_t}\cdot\sigma(\mathcal{R})^2\cdot\max_{j\geq J_t+1} s_i^{-1}. \tag{5.6}$$

When $c(t)$ is chosen to satisfy (5.6), as $\varepsilon \in \{\pm 1\}$ and v comes from a set of at most n vectors, we can choose ε and v deterministically such that $\Delta\Phi \leq 0$ at time t. For small enough Δt, this maintains the invariants $S(t) \succ 0$ and $\Phi(t) \leq \Phi(0)$.

5.3.4 Partial Coloring Matrix Discrepancy Bound

To obtain a partial coloring $x \in [-1,1]^n$ with $|\{i : x_i \in \{\pm 1\}\}| \geq n/2$, we run the above potential-based algorithm until some time T when the number of alive variables reaches $n_T = n/2$. Note that $T \leq n$ as $\|x(t)\|_2^2 = t$ for all t.

For any time step $t \leq T$, since we have by induction hypothesis that $\Phi(t) \leq \Phi(0) = m\cdot b(0)^{-p}$ and as we block the J_t smallest s_j, it follows that $\max_{j\geq J_t+1} s_j^{-p} \leq \Phi(0)/J_t$, and hence

$$\max_{j\geq J_t+1} s_j^{-1} \leq \left(\frac{\Phi(0)}{J_t+1}\right)^{1/p} = b(0)^{-1}\cdot\left(\frac{m}{J_t+1}\right)^{1/p}.$$

Therefore, for (5.6) to hold, it suffices to satisfy the condition

$$\begin{aligned} c(t) &\geq \frac{2}{n_t} \cdot \sigma(\mathcal{R})^2 + \frac{4b(t) \cdot (p+1)}{n_t} \cdot \sigma(\mathcal{R})^2 \cdot b(0)^{-1} \left(\frac{m}{J_t+1}\right)^{1/p} \\ &= \left(2 + \frac{8pb(t)}{b(0)} \left(\frac{m}{J_t+1}\right)^{1/p}\right) \cdot \frac{\sigma(\mathcal{R})^2}{n_t}. \end{aligned} \tag{5.7}$$

Suppose the initial barrier $b(0)$ and the rate $c(s)$ of increasing $b(s)$ at time s can be chosen so that for all $t \leq T$, the barrier $b(t)$ satisfies

$$b(t) = b(0) + \int_0^t c(s) \mathsf{d}s \leq 2b(0). \tag{5.8}$$

Then to satisfy (5.7) it suffices to choose

$$c(t) \geq 4\left(1 + 8p\left(\frac{m}{J_t+1}\right)^{1/p}\right) \cdot \frac{\sigma(\mathcal{R})^2}{n}. \tag{5.9}$$

To summarize, we have shown the following. If we can choose $b(0)$ and $c(t)$ to satisfy conditions (5.8) and (5.9), then the meta-algorithm above deterministically finds a partial coloring $x(T)$ with $n/2$ coordinates ± 1 and squared discrepancy $\|\mathcal{R}(x(T))^\top \mathcal{R}(x(T))\|_{\mathsf{op}} \leq 2b(0)$, and hence partial coloring discrepancy bound $\|\mathcal{R}(x(T))\|_{\mathsf{op}} \leq \sqrt{2b(0)}$.

5.4 Warm-Up: Recovering State-of-the-art Bounds

In this section, we prove Theorem 5.1.2 and 5.1.3 based on the framework described above. A common generalization and strengthening of both of these results will be given in Section 5.5.

5.4.1 Block-Diagonal Matrices

Let us recall Theorem 5.1.2 for block-diagonal matrices [LRR17, DJR22].

Theorem 4.1.3 (Matrix Spencer for Block Diagonal Matrices)**.** *Let $m \geq \sqrt{n}$ and $h \leq m$. Given block diagonal symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{\mathsf{op}} \leq 1$ and block size*

$h \times h$, one can efficiently find a coloring $x \in \{\pm 1\}^n$ with

$$\Big\| \sum_{i=1}^{n} x_i A_i \Big\|_{\mathsf{op}} \le O(\sqrt{n \cdot \max(1, \log(hm/n))}).$$

Proof. We directly apply the strategy in Section 5.3 by taking the matrices $R_i = A_i$ with dimension $R_i \in \mathbb{R}^{m \times m}$. As each R_i is h-block diagonal, the matrix $\mathcal{R}(x)^\top \mathcal{R}(x) = \mathcal{R}(x)^2$, where we recall that $\mathcal{R}(x) = \sum_{i=1}^n x_i R_i$, is also h-block diagonal for any $x \in \mathbb{R}^n$. Thus the orthogonal matrix $Q = Q(t)$ that diagonalizes $S(t) = b(t)I - \mathcal{R}(x(t))^2$ is also h-block diagonal. So the number of constraints in (5.3) for blocking one s_j is at most h, and thus blocking the $J_t = \lfloor n_t/2h \rfloor$ smallest s_j incurs only $n_t/2$ constraints (in particular the corresponding Q_j has at most h non-zero entries).

For the purpose of obtaining a partial coloring $x \in [-1, 1]^n$ with $|\{i : x_i \in \{\pm 1\}\}| \ge n/2$, we have $n_t \ge n/2$ and so $J_t \ge \lfloor n/4h \rfloor$. Also note that $\sigma(\mathcal{R})^2 = \| \sum_{i=1}^n A_i^2 \|_{\mathsf{op}} \le n$ since each $\|A_i\|_{\mathsf{op}} \le 1$. Thus to satisfy condition (5.9), it suffices to choose $p = \max\{1, \sqrt{\log(4hm/n)}\}$ and

$$c(t) = c := 4 \left(1 + 8p \left(\frac{m}{\lfloor n/4h \rfloor + 1} \right)^{1/p} \right) = O(\max\{1, \log(hm/n)\}).$$

Then setting $b(0) = cn$ guarantees $b(t) = b(0) + cT \le 2b(0)$ in condition (5.8). This gives a partial coloring $x \in [-1, 1]^n$ with $|\{i : x_i \in \{\pm 1\}\}| \ge n/2$ such that

$$\| \sum_i x_i A_i \|_{\mathsf{op}} \le \sqrt{2b_0} = O(\sqrt{n} \max\{1, \sqrt{\log(hm/n)}\}).$$

Repeatedly applying the argument above gives a full coloring $x \in \{\pm 1\}^n$ with discrepancy at most $O(\sqrt{n} \cdot \max\{1, \sqrt{\log(hm/n)}\})$. This proves the theorem. □

5.4.2 Moderate-Rank Matrices

We now prove Theorem 5.1.3 for moderate-rank matrices due to [HRS22].

Theorem 5.1.3 (Matrix Spencer for Moderate-Rank Matrices, Theorem 1.3 in [HRS22])**.**

Given symmetric matrices $A_1, \cdots, A_n \in \mathbb{R}^{m\times m}$ *with* $\|A_i\|_{\mathsf{op}} \le 1$ *and* $\sum_{i=1}^n \|A_i\|_F^2 \le nf$*, then one can efficiently find a coloring* $x \in \{\pm 1\}^n$ *with*

$$\|\sum_i x_i A_i\|_{\mathsf{op}} \le O\left(\sqrt{n}\cdot \max\{1, \sqrt{\log(f^2/n)}\}\right).$$

The remainder of this subsection is devoted to proving Theorem 5.1.3.

Proof of Theorem 5.1.3. Let $M := \sum_{i=1}^n A_i^2$ and note that $\mathsf{tr}(M) = \sum_i \mathsf{tr}(A_i^2) = \sum_i \|A_i\|_F^2 \le nf$, by our assumption. Also, we can assume wlog that $f \ge 2\sqrt{n}$ and just prove the bound $O(\sqrt{n\log(f^2/n)})$ in Theorem 5.1.3. As m can be arbitrarily large, we cannot directly apply the analysis in Section 5.3. Instead, we carefully exploit that M only has $O(f)$ eigenvalues of size $\Omega(n)$ (as $\mathsf{tr}(M) \le nf$), and in general, at most $f\cdot 10^k$ eigenvalues larger than $n10^{-k}$.

Dividing into Classes. By applying a rotation to the matrices A_i, we may assume wlog that M is diagonal. Note that each diagonal entry $M_{j,j} = \sum_{i=1}^n A_i(\cdot, j)^\top A_i(\cdot, j)$, where $A_i(\cdot, j)$ is the jth column of matrix A_i. We divide the diagonal entries of M into multiple classes as follows. Let $m_k = f\cdot 10^k$ be the dimension for class k. Let the top m_0 diagonal entries be class 0, then the next m_1 largest diagonal entries be class 1, and in general, once we have defined class $< k$, let the next m_k biggest entries be class k. This decomposes the rows/columns of the matrices A_i accordingly.

Intuitively, class 0 corresponds to the heaviest columns of the matrices A_i, which contribute the most to the discrepancy of $\sum_i x_i A_i$, class 1 to the next heaviest and so on. We ensure that the contribution of the class k columns to the discrepancy decreases geometrically with k. Roughly, this allows us to focus on the class 0 columns and view each A_i as an $m \times m_0$ matrix. As $m_0 = f$, the strategy in Section 5.3 will imply the desired $\sqrt{n\log(f^2/n)}$ bound. We now give the details.

Decomposing the matrices. Let $I_k \subseteq [m]$ be the indices for class k, and $I_{\ge k} = \cup_{u\ge k} I_u$ for classes $u \ge k$. For each class k and matrix A_i, define the L-shape class-k matrix $L_{k,i} \in$

$\mathbb{R}^{I_{\geq k} \times I_{\geq k}}$ as

$$L_{k,i}(y,z) := \begin{cases} A_i(y,z) & \text{if } y \in I_k, z \in I_{\geq k}, \text{ or if } y \in I_{\geq k}, z \in I_k \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $L_{k,i}$ is obtained from A_i as follows: First consider the $|I_{\geq k}| \times |I_{\geq k}|$ submatrix of A_i of rows and columns of class $\geq k$. Then set $A_i(y,z) = 0$ if both y and z are in class $\geq k+1$. Note that the $L_{k,i}$ for $k = 0, 1, 2, \ldots$, partition the entries of A_i. Next, let $R_{k,i} \in \mathbb{R}^{I_{\geq k} \times I_k}$ be the restriction of A_i to the entries in $I_{\geq k} \times I_k$. Note that $R_{k,i}$ is exactly the vertical $|I_{\geq k}| \times |I_k|$ rectangular part of $L_{k,i}$. Define $\mathcal{A}(x) := \sum_{i=1}^n x_i A_i$, $\mathcal{L}_k(x) := \sum_{i=1}^n x_i L_{k,i}$ and $\mathcal{R}_k(x) := \sum_{i=1}^n x_i R_{k,i}$ the corresponding discrepancy matrices.

It turns out that to bound the discrepancy of the matrices A_i, it suffices to control the discrepancy of the rectangular matrices $R_{k,i}$ for all class k.

Claim 5.4.1 (Discrepancy of $\mathcal{R}$ suffices)**.** *Let $x \in \mathbb{R}^n$ be any fractional coloring, and define $U_k := \|\mathcal{R}_k(x)^\top \mathcal{R}_k(x)\|_{\mathsf{op}}$ for all $k \geq 0$, then we have $\|\mathcal{A}(x)\|_{\mathsf{op}} \leq 2 \sum_{k \geq 0} \sqrt{U_k}$.*

Proof. We write $\mathcal{R}_k(x) = (X, Y)^\top$, where $X = \mathcal{A}(x)_{I_k \times I_k}$ is the $I_k \times I_k$ principal submatrix of $\mathcal{A}(x)$. Then we have $\mathcal{L}_k(x) = \begin{pmatrix} X & Y \\ Y^\top & 0 \end{pmatrix}$, and that

$$U_k = \|\mathcal{R}_k(x)^\top \mathcal{R}_k(x)\|_{\mathsf{op}} = \|X^2 + YY^\top\|_{\mathsf{op}}.$$

It immediately follows that $\|X\|_{\mathsf{op}} \leq \sqrt{U_k}$. Note that we also have

$$\|(\mathcal{L}_k(x) - \mathsf{diag}(X, 0))^2\|_{\mathsf{op}} = \|\mathsf{diag}(YY^\top, Y^\top Y)\|_{\mathsf{op}} \leq U_k.$$

This shows that $\|\mathcal{L}_k(x)\|_{\mathsf{op}} \leq 2\sqrt{U_k}$. Finally, note that since $L_{k,i}$ are symmetric, we have

$$\|\mathcal{A}(x)\|_{\mathsf{op}} = \Big\| \sum_i x_i A_i \Big\|_{\mathsf{op}} \leq \sum_k \Big\| \sum_i x_i L_{k,i} \Big\|_{\mathsf{op}} = \sum_k \|\mathcal{L}_k(x)\|_{\mathsf{op}} \leq 2 \sum_k \sqrt{U_k}.$$

This completes the proof of the claim. □

The key benefit of defining multiple classes comes from the following upper bound on the variance parameter for class k. Recall from Section 5.3 (e.g., see (5.6)) that the variance parameter crucially determines the final discrepancy bound.

Claim 5.4.2 (Variance Parameter for Class k)**.** *Define the variance parameter for class k as*

$$\sigma(\mathcal{R}_k)^2 := \max\left\{\left\|\sum_i R_{k,i}^\top R_{k,i}\right\|_{\mathsf{op}}, \left\|\sum_i R_{k,i} R_{k,i}^\top\right\|_{\mathsf{op}}\right\}.$$

Then we have $\sigma(\mathcal{R}_k)^2 \le \mathsf{tr}(M)/(1+\sum_{j=0}^{k-1} m_j) \le n/10^k$.

Proof. Note that we have $\sum_{i=1}^n R_{k,i}^\top R_{k,i} \preceq M_{I_{\ge k}, I_{\ge k}}$ and $\sum_{i=1}^n R_{k,i} R_{k,i}^\top \preceq M_{I_{\ge k}, I_{\ge k}}$. It then follows that $\sigma(\mathcal{R}_k)^2 \le \|M_{I_{\ge k}, I_{\ge k}}\|_{\mathsf{op}}$. Since $M_{I_{\ge k}, I_{\ge k}}$ is obtained from M by removing the largest $\sum_{j=0}^{k-1} m_j$ diagonal entries, the bound in the claim follows. □

An immediate consequence of Claim 5.4.2 is that we only need to consider classes k with $10^k \le n^{1.5}$. To see this, let k' be a class with $10^{k'} > n^{1.5}$, then by Claim 5.4.2 we have $\sigma(\mathcal{R}_{k'})^2 \le 1/\sqrt{n}$. Then for any coloring $x \in \{\pm 1\}^n$, we can bound

$$\|\mathcal{R}_{k'}(x)^\top \mathcal{R}_{k'}(x)\|_{\mathsf{op}} \le \left\|n \cdot \sum_i R_{k,i}^\top R_{k,i}\right\|_{\mathsf{op}} \le n \cdot \sigma(\mathcal{R}_{k'})^2 \le \sqrt{n}.$$

Thus an arbitrary coloring would have small discrepancy for class k'.

The Potential Function for Class k. We will apply the potential function meta analysis from Section 5.3, but define an individual potential function $\Phi_k(t)$ for each class k. In particular, let the slack matrix $S_k(t) \in \mathbb{S}^{m_k \times m_k}$ for class k be

$$S_k := b_k(t) \cdot I_{m_k} - \mathcal{R}_k(x(t))^\top \mathcal{R}_k(x(t)),$$

and the potential function be $\Phi_k(t) = \mathsf{tr}(S_k^{-p_k})$, where $p_k \ge 1$ will be specified later, and will be different for each class k. Ideally, we would like to show that there is some choice

of v and ε such that every Φ_k decreases simultaneously. However, there is no way to ensure this, and instead we will control $\Phi(t) = \sum_k \Phi_k(t)$, the sum of the potential functions of each class k. As in Section 5.3, we assume inductively that $\Phi(t) \le \Phi(0)$ and $S_k \succ 0$ for all classes k. Then we update $\Delta x(t) = \varepsilon(t)v(t)\sqrt{\Delta t}$ for unit vector $v \in \mathbb{R}(t)^{n_t}$ and $\varepsilon(t) \in \{\pm 1\}$, and $\Delta b_k(t) = c_k(t)\Delta t$. The goal is to find $v(t), \varepsilon(t)$ such that $\Phi(t)$ does not increase.

We run the same analysis as in Section 5.3.1-5.3.3, by blocking $J_{k,t} = \frac{n_t}{2^{k+2}m_k}$ of the smallest $s_{k,j}(t)$ from each class k. By (5.3), the number of constraints from blocking one $s_{k,j}(t)$ is at most m_k, and thus across all classes k, the total number of constraints incurred from blocking is at most

$$\sum_k m_k J_{k,t} = \sum_k \frac{n_t}{2^{k+2}} \le n_t/2.$$

It follows from (5.5) that the change of each $\Phi_k(t)$ is given by

$$\begin{aligned}&\frac{\mathbb{E}_{\varepsilon,v}[\Delta\Phi_k(t)]}{p_k\Delta t}\\ &\le -\sum_{j=1}^{m} s_{k,j}(t)^{-(p_k+1)}(c_k(t) - \frac{2}{n_t}\sigma(\mathcal{R}_k)^2) + \frac{4b_k(t)(p_k+1)}{n_t}\sigma(\mathcal{R}_k)^2 \sum_{j=J_{k,t}+1}^{m} s_{k,j}(t)^{-(p_k+2)}.\end{aligned}$$

Then to satisfy $\mathbb{E}_{\varepsilon,v}[\Delta\Phi_k(t)] \le 0$ for all class k, it suffices to choose

$$c_k(t) \ge \frac{2}{n_t}\cdot\sigma(\mathcal{R}_k)^2 + \frac{4b_k(t)\cdot(p_k+1)}{n_t}\cdot\sigma(\mathcal{R}_k)^2\cdot\max_{j\ge J_{k,t}+1} s_{k,j}(t)^{-1}. \tag{5.10}$$

Note that since the smallest $J_{k,t}$ of the $s_{k,j}(t)$ have been blocked, we have

$$\max_{j\ge J_{k,t}+1} s_{k,j}(t) \le \left(\frac{\Phi_k(t)}{J_{k,t}+1}\right)^{1/p_k} \le \left(\frac{\Phi(0)}{J_{k,t}+1}\right)^{1/p_k},$$

where the last inequality follows from the inductive assumption that $\sum_k \Phi_k(t) = \Phi(t) \le \Phi(0)$.

Parameter Setting for Partial Coloring. We run the above process until some time $T \le n$ such that $n_T = n/2$. By Claim 5.4.2 we have $\sigma(\mathcal{R}_k)^2 \le n/10^k$. As $n_t \ge n/2$, to satisfy

(5.10), it suffices to have

$$c_k(t) \geq \frac{4}{10^k} + \frac{8b_k(t)\cdot(p_k+1)}{10^k}\cdot\left(\frac{\Phi(0)}{J_{k,t}+1}\right)^{1/p_k}. \tag{5.11}$$

Now we fix the parameter setting so that (5.11) holds. Let us choose $p_0 = 100\log(f^2/n)$ and $b_0(0) = 100np_0$. We set $b_k(0) = b_0(0)/2^k$, and choose p_k such that $\Phi_k(0) = \Phi_0(0)/2^k$. This guarantees that $\Phi(0) = \sum_k \Phi_k(0) \leq 2\Phi_0(0) = 2^{k+1}\cdot\Phi_k(0)$. The value of p_k satisfies that

$$\Phi_k(0) = m_k(b_k(0))^{-p_k} = f\cdot 10^k\cdot(b_k(0))^{-p_k} = \Phi_0(0)/2^k = f\cdot 2^{-k}\cdot(b_0(0))^{-p_0}.$$

Using $b_k(0) = b_0(0)\cdot 2^{-k}$, we obtain the equality

$$20^k\cdot b_0(0)^{p_0} = (2^{-k}\cdot b_0(0))^{p_k}.$$

This implies that $p_k > p_0 \geq 100$. Next, we claim that $p_k \leq 4p_0$. This is because we are only considering classes k with $10^k \leq n^{1.5}$, and this together with $b_0(0) \geq 10n$ imply that $2^{-k}b_0(0) \geq b_0(0)^{1/2}$ and $20^k \leq 2^k\cdot 10^k \leq n^2$. Using these bounds, the equality above gives that $n^2 \geq b_0(0)^{p_k/2-p_0}$, and as $b_0(0) \geq 10n$ we have $p_k \leq 2p_0 + 4 \leq 4p_0$.

Finally, we set $c_k(t) = c_k := 25p_k5^{-k}$ independent of t for any class k. Note that since $b_k(0) = 100np_0\cdot 2^{-k}$, and as $p_k \leq 4p_0$, this choice of $c_k(t)$ guarantees that for any $t \leq n$,

$$b_k(t) = b_k(0) + \int_0^t c_k(s)\mathsf{d}s \leq b_k(0) + 25np_k\cdot 5^{-k} \leq 2b_k(0).$$

Now we check that (5.11) holds under these parameter settings. For any class $k \geq 0$, using $\Phi(0) \leq 2^{k+1}\Phi_k(0) = 2^{k+1}10^k f\cdot b_k(0)^{-p_k}$ and $J_k = \frac{n_t}{2^{k+2}m_k} \geq \frac{n}{2^{k+3}10^k f}$, then (5.11) is implied by

$$c_k(t) \geq 4\cdot 10^{-k} + 16(p_k+1)b_k(0)\cdot 10^{-k}\cdot b_k(0)^{-1}\left(\frac{2^{2k+4}10^{2k}f^2}{n}\right)^{1/p_k}.$$

Note that $p_k \geq p_0 \geq 100\log(f^2/n)$, so the factor $(2^{2k+4}10^{2k}f^2/n)^{1/p_k} \ll 2^k$. It follows that

our choice of $c_k(t) = 25p_k \cdot 5^{-k}$ satisfies the above inequality.

Partial Coloring Discrepancy Bound. By the parameter setting above, we have $b_k(T) \leq 2b_k(0) = 200np_0 \cdot 2^{-k}$. It follows that for each class k, the partial coloring $x(T)$ satisfies

$$\|\mathcal{R}_k(x(T))^\top \mathcal{R}_k(x(T))\|_{\mathsf{op}} \leq 2b_k(T) \leq O(2^{-k}) \cdot n\log(f^2/n).$$

Then Claim 5.4.1 implies that $\|\mathcal{A}(x)\|_{\mathsf{op}} \leq 2\sum_{k\geq 0}\sqrt{2b_k(T)} = O(\sqrt{n\log(f^2/n)})$. Finally, iterating the above partial coloring procedure, one finds a full coloring with discrepancy $O(\sqrt{n\log(f^2/n)})$, completing the proof of the theorem. □

5.5 A More General Bound for Matrix Spencer

In this section, we consider a more general class of matrices, and prove the following more general theorem that contains both Theorem 5.1.1 and Theorem 5.1.2.

Theorem 5.1.5 (General Matrix Discrepancy Bound)**.** *Given block-diagonal symmetric matrices $A_i = \mathsf{diag}(D_i^1, \cdots, D_i^\ell)$ with diagonal blocks $D_i^j \in \mathbb{R}^{h_j \times h_j}$. If $\|D_i^j\|_{\mathsf{op}} \leq 1$ and $\sum_{i=1}^n \|D_i^j\|_F^2 \leq gn$ for all $i \in [n]$ and $j \in [\ell]$, then one can efficiently find a coloring $x \in \{\pm 1\}^n$ with*

$$\Big\|\sum_i x_i A_i\Big\|_{\mathsf{op}} \leq O\left(\sqrt{n} \cdot \max\{1, \sqrt{\log(g^2\ell/n)}\}\right).$$

Notice that applying Theorem 5.1.3 directly (with $f = g\ell$) to this class of matrices gives a worse discrepancy bound of $O(\sqrt{n} \cdot \max\{1, \sqrt{\log(g^2\ell^2/n)}\})$.

As mentioned in the last paragraph of Section 5.1.1, there are instances where Theorem 5.1.5 is strictly stronger than Theorem 5.1.3 and Theorem 5.1.2. In particular, when $m = n$, $h_j = h = \sqrt{n}$ and $\|D_i^j\|_F^2 = \sqrt{h} = n^{1/4}$ for all $j \in [\ell]$, the bound in Theorem 5.1.5 is $O(\sqrt{n})$ while both Theorem 5.1.3 and Theorem 5.1.2 give $O(\sqrt{n\log n})$.

Before proving Theorem 5.1.5, we first give some intuition on how its proof differs from that of Theorem 5.1.3. At a high level, as in the proof of Theorem 5.1.3, we follow our general

framework in Section 5.3 to guarantee a discrepancy bound of $O(\sqrt{n}\cdot\max\{1, \sqrt{\log(m_0/J_0)}\})$, where m_0 is the number of heavy rows (corresponding to diagonal entries of $M := \sum_i A_i^2$ that are roughly n), and J_0 is the number of small $s_{0,i}$ we can block when at least $n/2$ coordinates $i \in [n]$ are alive.

We can only take $m_0 = g\ell$ as before, but our gain comes from being able to block more $s_{0,i}$ when $\ell > 1$. In Theorem 5.1.3, we can block at most $n/2g\ell$ smallest $s_{0,i}$. But now, since there are at most g heavy rows in each block, and since blocking each $s_{0,i}$ requires at most g constraints in (5.3), we can afford to block $J_0 = n/2g$ directions. This leads to the improved discrepancy bound of $O(\sqrt{n} \cdot \max\{1, \sqrt{\log(g^2\ell/n)}\})$ in Theorem 5.1.5. Next, we present a more formal proof.

Proof of Theorem 5.1.5. Let $M^j := \sum_{i=1}^n (D_i^j)^2$ for each block $j \in [\ell]$, which by assumption satisfies $\mathsf{tr}(M) \le gn$. Define $M := \sum_{i=1}^n A_i^2 = \mathsf{diag}(M^1, \cdots, M^\ell)$ as before, which we assume is diagonal wlog. We may also assume wlog that $g \ge 2\sqrt{n/\ell}$ so the bound in Theorem 5.1.5 is $O(\sqrt{n\log(g^2\ell/n)})$. Again, we decompose the columns of the A_i matrices into multiple classes.

Dividing into Classes. Let $\overline{m}_k = g \cdot 10^k$ be the number of class k columns for each diagonal block $j \in [\ell]$. For each block $j \in [\ell]$, we let the largest $\overline{m}_0$ diagonal entries of M^j be class 0, then the next $\overline{m}_1$ largest diagonal entries be class 1, and so on so forth. For all blocks $j \in [\ell]$, there is a total of $m_k = \overline{m}_k\ell$ class k entries. Let $I_k^j \subseteq [m]$ be the indices for class k that come from block $j \in [\ell]$. Let $I_k \subseteq [m]$ be the indices for class k, and $I_{\ge k} = \cup_{u\ge k} I_u$ the indices for classes $u \ge k$. As in the proof of Theorem 5.1.3, we define $L_{k,i} \in \mathbb{R}^{I_{\ge k}\times I_{\ge k}}$ and $R_{k,i} \in \mathbb{R}^{I_{\ge k}\times I_k}$ be the L-shape and rectangular matrices for each class k, which again satisfy Claim 5.4.1 and Claim 5.4.2.

The key observation is that the matrices $R_{k,i}$ admit the following common structure: the $\overline{m}_k$ columns $R_{k,i}^j \in \mathbb{R}^{I_{\ge k}\times I_k^j}$ from each diagonal block $j \in [\ell]$ are supported on different rows for different $j \in [\ell]$. This implies that $(R_{k,i}^j)^\top R_{k,i}^{j'} = \mathbf{0}$ whenever $j \ne j'$. Thus for any fractional coloring $x \in \mathbb{R}^n$, the squared discrepancy matrix $\mathcal{R}_k(x)^\top\mathcal{R}_k(x) \in \mathbb{R}^{\overline{m}_k\ell\times\overline{m}_k\ell}$ is

block-diagonal with block size $\overline{m}_k \times \overline{m}_k$.

The Potential Function for Class k. As in the proof of Theorem 5.1.3, we define

$$S_k(t) := b_k(t) \cdot I_{m_k} - \mathcal{R}_k(x(t))^\top \mathcal{R}_k(x(t)),$$

and use our meta analysis from Section 5.3 to bound the potential function $\Phi(t) = \sum_k \Phi_k(t)$, where each $\Phi_k(t) = \mathsf{tr}(S_k(t)^{-p_k})$. Then by (5.3), the constraints incurred by blocking one $s_{k,j}$ is

$$f \sum_{i=1}^{n} v_i(t) \cdot (R_{k,i}^\top \mathcal{R}_k(x(t)) Q_{k,j}(t) + \mathcal{R}_k(x(t))^\top R_{k,i} Q_{k,j}(t)) = \mathbf{0}.$$

Since $\mathcal{R}_k(x(t))^\top \mathcal{R}_k(x(t))$ is block-diagonal with block size $\overline{m}_k \times \overline{m}_k$, so does $S_k(t)$ and the orthogonal matrix $Q_k(t)$ in the spectral decomposition $S_k(t) = Q_k(t)^\top \mathsf{diag}(\{s_{k,j}(t)\}_{j=1}^{m_k}) Q_k(t)$. Thus the vector $(R_{k,i}^\top \mathcal{R}_k(x(t)) Q_{k,j}(t) + \mathcal{R}_k(x(t))^\top R_{k,i} Q_{k,j}(t))$ would be supported only on one of the blocks in $[\ell]$, and so the number of constraints above is at most $\overline{m}_k$, instead of $m_k = \overline{m}_k \ell$. Therefore, we can set $J_{k,t} = n_t/(\overline{m}_k 2^{k+2})$ so that the total number of constraints incurred from blocking is at most $n_t/2$.

Parameter Setting for Partial Coloring. Running the same analysis as in the proof of Theorem 5.1.3, we have the same sufficient condition (5.11) on $c_k(t)$ which we rewrite below

$$c_k(t) \geq \frac{4}{10^k} + \frac{8 b_k(t) \cdot (p_k + 1)}{10^k} \cdot \left(\frac{\Phi(0)}{J_{k,t} + 1} \right)^{1/p_k}. \tag{5.12}$$

The only difference in our parameter setting from that of Theorem 5.1.3 is that we set $p_0 = 100 \log(g^2 \ell / n)$. Then we can set $b_k(0) = 100 n p_0 / 2^k$, $c_k(t) = c_k := 25 p_k 5^{-k}$ and p_k such that $\Phi_k(0) = \Phi_0(0)/2^k$ as before. Following along the lines of the proof of Theorem 5.1.3, one can verify that these parameter settings satisfy condition (5.12).

Partial Coloring Discrepancy Bound. By the parameter setting above, we have $b_k(T) \le 2b_k(0) = 200np_0 \cdot 2^{-k}$. It follows that for each class k, the partial coloring $x(T)$ satisfies

$$\|\mathcal{R}_k(x(T))^\top \mathcal{R}_k(x(T))\|_{\mathsf{op}} \le 2b_k(T) \le O(2^{-k}) \cdot n\log(g^2\ell/n).$$

Then Claim 5.4.1 implies that $\|\mathcal{A}(x)\|_{\mathsf{op}} \le 2\sum_{k\ge 0}\sqrt{2b_k(T)} = O(\sqrt{n\log(g^2\ell/n)})$.

Finally, iterating the above partial coloring procedure, one finds a full coloring with discrepancy $O(\sqrt{n\log(g^2\ell/n)})$, completing the proof of the theorem. □

5.6 *Missing Proof*

In this section, we prove Theorem 5.1.4 which is restated below for convenience.

Theorem 5.1.4 (Theorem 3.1 in [HRS22]). *Given symmetric matrices $A_1, \cdots, A_n \in \mathbb{R}^{m\times m}$, one can efficiently find a partial coloring $x \in [-1,1]^n$ with $|\{i : x_i \in \{\pm 1\}\}| = \Omega(n)$ such that*

$$\Big\|\sum_i x_i A_i\Big\|_{\mathsf{op}} \le O\left(\Big\|\sum_i A_i^2\Big\|_{\mathsf{op}}^{1/2} \cdot \max\left\{1, \sqrt{\log\left(\frac{\sum_i \mathsf{tr}(A_i^2)}{\sqrt{n}\|\sum_i A_i^2\|_{\mathsf{op}}}\right)}\right\}\right).$$

The proof of Theorem 5.1.4 largely follows from the proof of Theorem 5.1.3 with a different parameter setting, so we only highlight their differences below.

Proof of Theorem 5.1.4 (Sketch). For simplicity, we denote $M := \sum_i A_i^2$ as before, and denote $\beta := \mathsf{tr}(M)$ and $\alpha := \|M\|_{\mathsf{op}}$ for simplicity. Note that if we multiply all matrices A_i by a factor $\lambda > 0$, then the bound in Theorem 5.1.4 also gets multiplied by λ. Therefore, we may rescale the matrices A_i such that $\alpha = n$. We may also assume without loss of generality that $\beta \ge 2\alpha\sqrt{n}$ so that the bound in Theorem 5.1.4 is $O(\sqrt{\alpha\log(\beta^2/\alpha^2 n)})$.

Our rescaling of the matrices A_i above might violate the assumptions $\|A_i\|_{\mathsf{op}} \le 1$ in the statement of Theorem 5.1.3. However, we can still repeat the proof of Theorem 5.1.3, as our analysis for obtaining a partial coloring discrepancy bound only uses the bound on $\sigma(\mathbb{R}_k)^2$, and the only way we are using $\|A_i\|_{\mathsf{op}} \le 1$ is to obtain a bound on $\sigma(\mathbb{R}_k)^2$. More precisely, we can simply repeat the proof of Theorem 5.1.3, by setting $m_k = \frac{\beta}{\alpha}\cdot 10^k$ instead. Dividing

into classes using this new m_k value, we have $\sigma(\mathbb{R}_k)^2 \leq O(\frac{\beta}{m_k}) \leq O(\frac{\alpha}{10^k})$. As in the proof of Theorem 5.1.3, we can still block $J_{k,t} = \frac{n_t}{2^{k+2}m_k}$ smallest $s_{k,j}$ from each class k as before, and (5.11) still suffices for potential decrease.

The only difference in the parameter setting is that now we use $p_0 = 100\log(m_k^2/n)$. The rest of the parameters are set in the same way as in the proof of Theorem 5.1.3, namely $b_0(0) = 100np_0$, $b_k(0) = b_0(0)/2^k$, $c_k(t) = c_k := 25p_k5^{-k}$, and p_k such that $\Phi_k(0) = \Phi_0(0)/2^k$. Following along the lines of the proof of Theorem 5.1.3, one can verify that these parameter settings satisfy (5.11) for all classes k. The discrepancy bound is then $O(\sqrt{b_0(T)}) = O(\sqrt{\alpha\log(\beta^2/\alpha^2 n)})$. □

Chapter 6

MATRIX DISCREPANCY III: MATRIX SPENCER CONJECTURE UP TO POLY-LOGARITHMIC RANK

In this chapter, we present our latest progress on the matrix Spencer conjecture (Conjecture 1.3.1). In particular, we detail a proof of the matrix Spencer conjecture up to poly-logarithmic rank, which also implies a nearly optimal lower bound for quantum random access codes using a connection obtained in [HRS22]. This chapter is based on a joint paper with Nikhil Bansal and Raghu Meka [BJM22b].

This chapter is intended to be self-contained so that readers who are interested in understanding the latest progress on the matrix Spencer conjecture can directly start reading through this chapter without having to recall any definitions and notations from the previous two chapters. To make for a self-contained presentation, this chapter starts by repeating the motivations and definitions that have already appeared in the last two chapters.

6.1 Introduction

We study discrepancy minimization in the matrix setting. Let us start with the classical discrepancy setting where given vectors $a_1, \ldots, a_n \in \mathbb{R}^d$ satisfying $\|a_i\|_\infty \leq 1$ for all $i \in [n]$, and the goal is to find signs $x_1, \ldots, x_n \in \{\pm 1\}$ to minimize the discrepancy $\|\sum_{i=1}^n x_i a_i\|_\infty$. In a seminal result, Spencer [Spe85] showed that the $O(\sqrt{n \log d})$ bound obtained by a random coloring is not tight and showed the following bound, which is also the best possible in general.

Theorem 4.1.1 (Spencer [Spe85])**.** *Let $m \geq n$. Given vectors $a_1, \ldots, a_n \in \mathbb{R}^m$ with $\|a_i\|_\infty \leq 1$, there exists $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i a_i\|_\infty \leq O(\sqrt{n \log(2m/n)})$.*

In particular for $d = O(n)$, this gives an $O(\sqrt{n})$ bound, in contrast to the $O(\sqrt{n \log n})$

bound for random coloring obtained by applying Chernoff and union bounds.

To prove this result, Spencer developed the powerful partial-coloring method via the entropy method, building on the previous work of Beck [Bec81]. Another approach to prove Theorem 4.1.1 based on convex geometry was developed independently by Gluskin [Glu89]. While these original arguments used the pigeonhole principle and were non-algorithmic, in recent years, there has been a rich line of work [Ban10, BS13, LM15a, Rot17, LRR17, ES18, RR20a] on their algorithmic versions.

Matrix Spencer Setting. A natural generalization of Spencer's problem to matrices is the following. Let $A_1, \ldots, A_n \in \mathbb{R}^{d\times d}$ be symmetric matrices with maximum singular value, or operator norm, $\|A_i\|_{\mathsf{op}} \leq 1$. Find a coloring $x \in \{\pm 1\}^n$ that minimizes $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}}$. In particular, Spencer's result corresponds to the case when all the $A_i = \mathsf{diag}(a_i)$ are diagonal.

As in the vector case, for a random coloring $x \in \{\pm 1\}^n$, the non-commutative Khintchine inequality of Lust-Piquard and Pisier [LP86, LPP91, Pis03], or the matrix Chernoff bound [Oli10, Tro15], give that

$$\mathbb{E}\Big[\big\|\sum_i x_i A_i\big\|_{\mathsf{op}}\Big] = O\Big(\sqrt{\log d}\cdot \Big\|\sum_i A_i^2\Big\|_{\mathsf{op}}^{1/2}\Big). \tag{6.1}$$

This implies a bound of $O(\sqrt{n\log d})$ on the matrix discrepancy. This inequality also holds when one picks $x \in \mathbb{R}^n$ to be standard Gaussians, which will play an important role in our results.

Matrix concentration bounds are powerful and widely used tools in mathematics and computer science, and it is natural to ask when one can beat them. In particular, whether the following natural analog of Spencer's result for matrices holds is a tantalizing open question.

Conjecture 1.3.1 (Matrix Spencer Conjecture, [Zou12, Mek14])**.** *Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m\times m}$ with each $\|A_i\|_{\mathsf{op}} \leq 1$, there exist signs $x \in \{\pm 1\}^n$ such that $\big\|\sum_{i=1}^n x_i A_i\big\|_{\mathsf{op}} \leq O(\sqrt{n}\cdot \max\{1, \sqrt{\log(m/n)}\})$. In particular, the matrix discrepancy is*

$O(\sqrt{n})$ *for* $m = n$.

While this conjecture is still open, there has been exciting progress on important special cases. Recently, Hopkins, Raghavendra and Shetty [HRS22] proved Conjecture 1.3.1 where each matrix A_i has rank $O(\sqrt{n})$; in a different direction, Levy, Ramadas and Rothvoss [LRR17] and Dadush, Jiang and Reis [DJR22] established Conjecture 1.3.1 for block-diagonal matrices with block size $h = O(n/d)$. Recently, Bansal, Jiang, and Meka [BJM22a] gave an approach based on *barrier functions* to achieve a bound that unifies and slightly strengthens the results of [HRS22, DJR22].

6.1.1 Our Results

The main result of this chapter is the following theorem.

Theorem 6.1.1 (Matrix Spencer Up to Poly-logarithmic Rank)**.** *Given $d \times d$ symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{d\times d}$ each with $\|A_i\|_{\mathsf{op}} \le 1$ and $\|A_i\|_F^2 \le n/\log^3 n$, there exist signs $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} = O(\sqrt{n})$. Moreover, these signs can be computed efficiently.*

Note that the condition $\|A_i\|_F^2 \le n/\log^3 n$ is satisfied when each A_i has rank at most $n/\log^3 n$ or in particular if $d \le n/\log^3 n$. Thus Theorem 6.1.1 resolves Conjecture 1.3.1 up to poly-logarithmic dimension or poly-logarithmic rank. We remark that even when assuming the matrices A_i have small rank (even rank 1) or small dimension (even $d = \sqrt{n}$), it is known that one cannot hope for a bound better than $\Theta(\sqrt{n})$ [DJR22]. For instance, let $e_1, \ldots, e_n \in \mathbb{R}^n$ be the standard basis vectors and take $A_i = (1/2)(e_1 + e_i)(e_1 + e_i)^T$. Then, for any $x \in \{\pm 1\}^n$, the first column of $\sum_i x_i A_i$ has norm $\Omega(\sqrt{n})$ so its spectral norm is $\Omega(\sqrt{n})$. This is in sharp contrast to the diagonal case, where an $O(\sqrt{r \log n})$ bound for rank r matrices holds [Ban98], and a $O(\sqrt{r})$ bound was conjectured [BF81].

Further, when matrices A_i have dimension $d = \omega(n)$ but $\mathsf{rank}(A_i) \le n/\log^3 n$, the bound in Theorem 6.1.1 is $O(\sqrt{n})$ and is stronger than the $\omega(\sqrt{n})$ bound suggested by Conjecture 1.3.1.

A new ingredient in our proof is a recent strengthening of the non-commutative Khintchine inequality for Gaussian random matrices of the form $\sum_i g_i A_i$ where $g_1, \ldots, g_n$ are independent standard Gaussian random variables due to Bandeira, Boedihardjo, and van Handel [BBvH21]. The central idea is to pick a suitable projection of the random matrix to a subspace so that the bound of [BBvH21] matches $O(\sqrt{n})$. We defer the details to the full proof.

The same proof strategy also implies the following improvement over the random coloring bound of $O(\sqrt{n \log d})$ whenever the matrices have $\mathsf{rank}(A_i) = o(n)$, and in particular whenever the dimension is $d = o(n)$, by using the result of [Tro18] together with [BBvH21]. Previously, nothing better than the random coloring bound was known even when $d = n^{1/2+\varepsilon}$ for any small constant $\varepsilon > 0$.

Corollary 6.1.2 (Improvement Over Random Coloring)**.** *Given $d \times d$ symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ each with $\|A_i\|_{\mathsf{op}} \leq 1$ and $\|A_i\|_F^2 \leq r$, there exist signs $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} = O\left(\sqrt{n}(\log d)^{1/4} + (nr)^{1/4}\sqrt{\log d}\right)$. Moreover, these signs can be computed efficiently.*

We leave the details of the proof of Corollary 6.1.2 to Section 6.4.

Implications for Quantum Random Access Codes. [HRS22] identified a beautiful connection between the matrix Spencer conjecture and quantum random access codes that achieve advantage $C/\sqrt{n}$ for a big enough constant. They use this connection in their proof of the conjecture for matrices of rank $O(\sqrt{n})$.

Consider the following two-party communication problem: Alice is given a vector $x \in \{\pm 1\}^n$ and Bob an index $i \in [n]$. We are interested in the one-way quantum communication complexity (from Alice to Bob) of computing x_i. That is, Alice gets to send a quantum message to Bob and Bob must use this message to compute a guess for x_i. For a protocol Π, let $\mathsf{adv}_\Pi(x, i) = \max(0, \mathbb{P}[\Pi(x, i) = x_i] - 1/2)$ be the advantage over random guessing that Alice and Bob have. Note that the randomness is over that of the protocol.

The seminal works of [ANTSV02] showed that for any protocol Π with $\mathbb{E}_{x,i}[\mathsf{adv}_\Pi(x, i)] =$

$\Omega(1)$, Alice must send $\Omega(n)$ qubits to Bob[1]. In [HRS22], the following elegant connection between Conjecture 1.3.1 and the above communication problem is made: The conjecture is true if and only if there is some constant C such that any protocol Π with $\min_x \mathbb{E}_i[\mathsf{adv}_\Pi(x,i)] > C/\sqrt{n}$ must send at least $\log_2 n - O(1)$ qubits from Alice to Bob.

As our main result, Theorem 6.1.1, proves the conjecture for matrices of dimension $n/\log^3 n$, this combined with Claim 1.6 in [HRS22] immediately imply the following corollary:

Corollary 6.1.3 (QRAC Lower Bound)**.** *There exists a universal constant $C > 0$ such that the following holds. Any quantum one-way protocol Π as above with $\min_x \mathbb{E}_i[\mathsf{adv}_\Pi(x,i)] > C/\sqrt{n}$ requires at least $\log_2 n - 3\log_2\log_2 n - O(1)$ qubits of communication from Alice to Bob.*

Note that the leading constant of 1 in front of $\log_2 n$ is right for the first time and is the best possible (for sufficiently large constant $C > 0$). Previously, the results of [HRS22, DJR22] imply a lower bound of $(1/2)\log_2 n - O(1)$ on the quantum one-way communication complexity.

Further, a modification of the example in [DJR22] shows that there exists a protocol Π such that for all $x \in \{\pm 1\}^n, i \in [n]$, $\mathsf{adv}_\Pi(x,i) > c/\sqrt{n}$ for some constant $c > 0$ and involves at most $(1/2)\log_2 n + O(1)$ qubits of communication. Combined with our lower bound, this shows a somewhat sharp transition in the communication required for protocols as in Corollary 6.1.3: for some constants $0 < c < C$, achieving an advantage of $C/\sqrt{n}$ requires $\log_2 n - O(\log\log n)$ qubits, whereas one can achieve $c/\sqrt{n}$ advantage with $(1/2)\log_2 n + O(1)$ qubits. Interestingly, the transition is a quantum phenomenon and is absent for classical randomized communication; a tight bound of $\log_2 n + \Theta(\alpha^2)$ bits of communication is known for achieving advantage $\alpha/\sqrt{n}$ for all $\alpha > 0$.

---

[1] On a related note, if one is not interested in the exact constant, one can obtain an $\Omega(n)$ bound easily from the matrix Chernoff bound in (6.1) (without using any quantum information theory).

6.1.2 Further Related Works

Discrepancy Theory. Discrepancy theory is widely studied and has applications to many other mathematics and computer science areas. We refer readers to the excellent books [Cha00, Mat99, CST+14] for a more comprehensive account of the rich history of discrepancy. Recent developments in discrepancy have led to several applications in approximation algorithms, differential privacy, fair allocation, experimental design, and more [MN12, Rot13, NTZ13, BCKL14a, Ban19, JKS19, HSSZ19, BJSS20, BRS22].

Matrix Discrepancy and Non-Commutativity Random Matrix Theory. Many natural problems in the study of spectra of matrices can be viewed as questions about matrix discrepancy, e.g., graph sparsification [BSS12, RR20b], the Kadison-Singer problem [MSS15] and its generalization [KLS20], and the design of quantum random access codes [ANTSV02, HRS22].

Matrix discrepancy is also closely related to non-commutative random matrix theory, where the typical value of $\|\sum_i x_i A_i\|_{\mathsf{op}}$ for a random coloring x has received significant attention. The bound of $\mathbb{E}[\|\sum_i x_i A_i\|_{\mathsf{op}}] \leq O(\sqrt{n \log m})$ by matrix Chernoff [AW02] or matrix Khintchine [LP86, LPP91, Pis03] that is generally tight for commutative matrices, can be often improved in the non-commutative case (e.g. [Ver18, Tro18, BBvH21] and the references therein). We refer readers to the book [Tao12, Vu14] for a more comprehensive account of random matrix theory.

6.2 Preliminaries

We first recall some basic facts about matrices and describe the notations. For a square matrix $A \in \mathbb{R}^{m \times m}$ with entries a_{ij}, its trace $\mathsf{tr}(A) = \sum_i a_{ii}$ and Frobenius norm $\|A\|_F = \sqrt{\mathsf{tr}(A^T A)} = (\sum_{ij} a_{ij}^2)^{1/2}$. If A is symmetric with eigenvalues $\lambda_1, \ldots, \lambda_n$, then we have $\mathsf{tr}(A) = \sum_i \lambda_i$, $\|A\|_F = (\sum_i \lambda_i^2)^{1/2}$ and its operator norm $\|A\|_{\mathsf{op}} = \max_{\|x\|_2=1} \|Ax\|_2 = \max_i |\lambda_i|$. A symmetric matrix A is positive semidefinite (PSD) if all its eigenvalues $\lambda_i \geq 0$.

For a linear subspace $H \subseteq \mathbb{R}^n$, let $H^\perp$ denote its orthogonal complement. For any matrix

$A \in \mathbb{R}^{d\times d}$, we let $\vec{(}A) \in \mathbb{R}^{d^2}$ be the vector formed by the d^2 entries of A in a fixed order. For a subspace H and convex set $K \subseteq H$, denote $\gamma_H(K)$ the Gaussian measure of K restricted to H, i.e. the probability that a standard Gaussian vector on H lies in K.

6.2.1 Matrix Concentration

Let $X \in \mathbb{R}^{d\times d}$ be a symmetric multi-variate Gaussian random matrix (i.e., the entries of X are jointly Gaussian). Equivalently, we can assume that X is of the form $X = \sum_{i=1}^n g_i A_i$ where g_i are independent standard Gaussians and $A_1, \ldots, A_n \in \mathbb{R}^{d\times d}$ are symmetric matrices[2]. Note that this representation of X is not unique, and by the rotational invariance of the Gaussians, one also has $X = \sum_j g_j B_j$ where $B_j = \sum_i (v^j)_i A_i$ for any $n \times n$ orthogonal matrix with columns v^j.

Let $\sigma(X)^2 = \|\mathbb{E}[X^2]\|_{\mathsf{op}} = \|\sum_i A_i^2\|_{\mathsf{op}}$. The fundamental matrix-Chernoff inequality or non-commutative Khintchine inequality implies, among other things, that for X as above, we have

$$\mathbb{E}[\|X\|_{\mathsf{op}}] = O(\sigma(X) \cdot \sqrt{\log d}).$$

Note that this bound is tight in general, for instance, if X is a suitable diagonal matrix. Much attention has been given to finding special cases where the $\sqrt{\log d}$ factor in the bound above can be improved. Of particular note is the work of Tropp [Tro18] where he introduced a specific matrix alignment parameter to capture the non-commutativity of the matrices A_i.

Recently, Bandeira, Boedihardjo, and van Handel made substantial progress in this direction in [BBvH21]. In particular, they related the matrix alignment parameter of Tropp to the following more natural parameter. Let

$$\mathsf{Cov}(X) := \mathbb{E}[\vec{(}X)\vec{(}X)^\top] = \mathbb{E}\Big[\sum_{i=1}^n \vec{(}A_i)\vec{(}A_i)^\top\Big] \tag{6.2}$$

[2]It will be useful to think of the matrix X by itself as a random matrix, and only use the specific representation $\sum_i g_i A_i$ when needed.

be the $d^2 \times d^2$ covariance matrix of its d^2 scalar entries and define

$$v(x)^2 := \|\mathsf{Cov}(X)\|_{\mathsf{op}}.$$

Bandeira, Boedihardjo, and van Handel [BBvH21] showed the following refinement of the non-commutative Khintchine inequality of Lust-Piquard and Pisier [LP86, LPP91, Pis03].

Theorem 6.2.1 ([BBvH21], Theorem 1.2)**.** *Given symmetric matrices $A_1, \cdots, A_n \in \mathbb{R}^{d\times d}$, let $X = \sum_{i=1}^n g_i A_i$ where g_i are i.i.d. standard Gaussians. Then*

$$\mathbb{E}[\|X\|_{\mathsf{op}}] \leq C \cdot (\sigma(X) + (\log^{3/4} d)\sigma(X)^{1/2} v(X)^{1/2}),$$

where C is some universal constant. In particular, $\mathbb{E}[\|X\|_{\mathsf{op}}] = O(\sigma(X) + (\log^{3/2} d)v(X))$.

We remark that the bound in [BBvH21] is substantially more potent and, in particular, gives the optimum constant for the $\sigma(X)$ term and even control over the full spectrum of X. However, the weaker version above suffices for our purposes.

6.2.2 Partial Colorings in Convex Sets

The seminal work of Gluskin [Glu89] introduced the idea of finding partial colorings via techniques from convex geometry. At the core is the idea that any symmetric convex set $K \subseteq \mathbb{R}^n$ with sufficiently large Gaussian volume must contain a vector from $\{-1, 0, 1\}^n$ with $\Omega(n)$ non-zero coordinates (i.e., a *good partial coloring*). In particular, Giannopoulos [Gia97] showed that if $\gamma(K) \geq e^{-\delta n}$ for a sufficiently small constant δ, then K must contain a good partial coloring. Rothvoss [Rot13] gave an algorithmic version of Giannopoulos's result and extended it to subspaces with dimension close to n. This extension will be useful for our purposes.

Lemma 6.2.2 ([Rot17], Lemma 9)**.** *Let $\varepsilon \leq 1/60000$ and $\delta := \frac{3}{2}\varepsilon \log_2(1/\varepsilon)$. Given a subspace $H \subseteq \mathbb{R}^n$ of dimension at least $(1-\delta)n$, a symmetric convex set $K \subseteq H$ with*

$\gamma_H(K) \geq e^{-\delta n}$ *and a point* $x_0 \in (-1,1)^n$. *There exists a polynomial time algorithm to find a point* $x \in (x_0 + K) \cap [-1,1]^n$ *so that* $|\{i : x_i \in \{\pm 1\}\}| \geq \varepsilon n/2$.

6.3 Proof of the Main Result

Following the standard approach, it suffices to find a partial coloring with $O(n^{1/2})$ discrepancy. We show this below in Section 6.3.1, and then show how Theorem 6.1.1 follows from it in Section 6.3.2.

6.3.1 Main Partial Coloring Lemma

Lemma 6.3.1 (Main Partial Coloring Lemma)**.** *There exist constants* $c, c' > 0$ *such that the following holds. Given symmetric matrices* $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ *that satisfy* $\|\sum_{i=1}^n A_i^2\|_{\mathsf{op}} \leq \sigma^2$ *and* $\sum_{i=1}^n \|A_i\|_F^2 \leq nf^2$ *and a point* $x_0 \in (-1,1)^n$, *there exists a point* $x \in [-1,1]^n$ *such that*

$$\Big\| \sum_{i=1}^n (x_i - x_{0,i}) A_i \Big\|_{\mathsf{op}} \leq c(\sigma + (\log^{3/4} d)\sqrt{\sigma f}\,),$$

and $|\{i : x_i \in \{\pm 1\}\}| > c'n$. *Moreover, such a point can be found in polynomial time.*

The partial coloring upper bound could be changed to the clearer bound of $O(\sigma + (\log d)^{3/2} f)$ without too much loss; but the above is better for our recursion. In particular, note that if $\sigma \leq \sqrt{n}$ and $f^2 \leq n/\log^3 d$ (which will be true when $\|A_i\|_{\mathsf{op}} \leq 1$, and $\mathsf{rank}(A_i) \leq n/\log^3 d$), we get a partial coloring with a spectral norm bound of $O(\sqrt{n})$.

The idea behind the proof is as follows. Let $X = \sum_{i=1}^n g_i A_i$ where g_i are i.i.d standard Gaussian random variables. Consider the convex body

$$K = \Big\{ x \in \mathbb{R}^n : \Big\| \sum_{i=1}^n x_i A_i \Big\|_{\mathsf{op}} \leq c\sigma \Big\} \subseteq \mathbb{R}^n$$

for some suitably large constant $c > 0$. If K had Gaussian measure $\gamma(K) \geq \exp(-\Omega(n))$, then we could directly use Rothvoss's partial coloring result [Rot17]. As $\sigma(X) \leq \sigma$, one may hope that the improved concentration bound in Theorem 6.2.1 can be used to show such a

lower bound on $\gamma(K)$. However, it is unclear how to do this directly, as we do not have any control on $v(X)$ and it might even be larger than $\sigma(X)$. So our key idea is to work with a suitable slice of the body K.

A key observation is that even if $v(X)$ itself is large, the number of large eigenvalues of $\mathsf{Cov}(X)$ must be small as $\mathsf{tr}(\mathsf{Cov}(X)) = \sum_{i=1}^{n} \|A_i\|_F^2 \leq nf^2$. In particular, if we set $\Delta^2 \geq f^2/\delta$, then the number of *bad* eigenvectors of $\mathsf{Cov}(X)$ with eigenvalue greater than Δ^2 is at most δn. The main idea is to restrict the g_i's to lie in a subspace $H \subseteq \mathbb{R}^n$ so that if $y \in H$ is drawn from the standard Gaussian distribution on H, the resulting matrix $Y = \sum_i y_i A_i$ is perpendicular to each of the *bad* eigenvectors of $\mathsf{Cov}(X)$. This ensures that $v(Y) \leq \Delta$ and by Theorem 6.2.1, $\mathbb{E}[\|Y\|_{\mathsf{op}}] = O(\sigma + (\log^{3/4} d)\sqrt{\sigma \cdot f})$. Further, as the number of such bad eigenvectors of $\mathsf{Cov}(X)$ is small, we can ensure that H has dimension at least $(1-\delta)n$. We can now apply Lemma 6.2.2 to get the desired partial coloring. We now give the details.

Proof of Lemma 6.3.1. We define constants $\varepsilon := 1/60000$ and $\delta := \frac{3}{2}\varepsilon \log_2(1/\varepsilon)$ to be as in Lemma 6.2.2. We define $X = \sum_{i=1}^{n} g_i A_i$ where g_i are i.i.d. standard Gaussian random variables. Consider the PSD matrix $\mathsf{Cov}(X) \in \mathbb{R}^{d^2 \times d^2}$ defined in (6.2). Note that by assumption,

$$\mathsf{tr}(\mathsf{Cov}(X)) = \sum_{i=1}^{n} \|\vec{(}A_i)\|_2^2 = \sum_{i=1}^{n} \|A_i\|_F^2 \leq \delta n \Delta^2,$$

where $\Delta^2 := f^2/\delta$. This implies that there can be at most $k := \delta n$ eigenvalues of $\mathsf{Cov}(X)$ exceeding Δ^2. Let $V_1, \cdots, V_k \in \mathbb{R}^{d \times d}$ be such that $\vec{(}V_j)$ is the eigenvector for the jth largest eigenvalue of $\mathsf{Cov}(X)$. Define the subspace

$$H := \Big\{ y \in \mathbb{R}^n : \sum_{i=1}^{n} y_i \cdot \mathsf{tr}(A_i V_j) = 0, \forall j \in [k] \Big\}.$$

Now we sample the standard Gaussian vector $g \in \mathbb{R}^n$ as follows: first sample a standard Gaussian vector $y \in H$, then sample an independent standard Gaussian vector $r \in H^{\perp}$, and finally let $g = y + r$. We define $Y := \sum_{i=1}^{n} y_i A_i$ and $R := \sum_{i=1}^{n} r_i A_i$, which implies $X = Y + R$. Since Y and R are independent and have zero mean, we immediately have that

$\mathbb{E}[X^2] = \mathbb{E}[Y^2] + \mathbb{E}[R^2]$ and therefore $\sigma(Y) \le \sigma(X) \le \sigma$ by the assumption in the Lemma.

We next show that $v(Y) \le \Delta$. As $\mathsf{tr}(YV_j) = 0$ for any $j \in [k]$, we have $\vec{(V_j)}^\top \mathsf{Cov}(Y)\vec{(V_j)} = 0$. As $\mathsf{Cov}(Y)$ is a PSD matrix, $W := \mathsf{span}\{\vec{(V_1)}, \cdots, \vec{(V_k)}\} \subseteq \mathbb{R}^{d^2}$ must be a subspace of the eigenspace corresponding to the eigenvalue 0 of the matrix $\mathsf{Cov}(Y)$. For any vector $v \in \mathbb{R}^{d^2}$ with $v \perp W$, we thus have that

$$v^\top \mathsf{Cov}(Y) v \le v^\top \mathsf{Cov}(X) v \le \Delta^2,$$

as $\mathsf{Cov}(X) = \mathsf{Cov}(Y) + \mathsf{Cov}(R)$, and the $(k+1)$th eigenvalue of $\mathsf{Cov}(X)$ is at most Δ^2. This proves that $\|\mathsf{Cov}(Y)\|_{\mathsf{op}} \le \Delta^2$, or equivalently $v(Y) \le \Delta$.

Now we want to apply Theorem 6.2.1 to $Y = \sum_{i=1}^n y_i A_i$. A crucial but elementary fact is that Theorem 6.2.1 holds for any (symmetric) matrix-valued random variable whose entries are jointly Gaussian and the final bound only depends on the overall distribution of the random matrix and not on the specific representation as a sum of independent matrices. Clearly, the matrix Y we have is a multi-variate Gaussian random variable so we can apply their result.

To be precise, we can justify its validity even though the vector y does not have independent coordinates as follows. Let $v^1, \ldots, v^k \in \mathbb{R}^n$ be an orthonormal basis for H. Then, we can write $y = \sum_{j=1}^k h_j v^j$ where h_i are i.i.d standard Gaussian variables. We can now write

$$Y = \sum_{i=1}^n y_i A_i = \sum_{j=1}^k h_j \Big(\sum_{i=1}^n (v^j)_i A_i \Big) = \sum_{j=1}^k h_j B_j,$$

where we define $B_j = \sum_{i=1}^n (v^j)_i A_i$. Thus Y can be written in the form of a Gaussian matrix series in terms of the i.i.d. standard Gaussians h_i.

Thus we can apply Theorem 6.2.1 to Y and obtain for universal constant $C > 0$,

$$\mathbb{E}[\|Y\|_{\mathsf{op}}] \le C \cdot (\sigma(Y) + (\log^{3/4} d) \cdot \sqrt{\sigma(Y) v(Y)}) \le c(\sigma + (\log^{3/4} d)\sqrt{\sigma \cdot f}), \tag{6.3}$$

for a sufficiently big constant $c > C(1 + 1/\sqrt{\delta})$. Let us consider the convex body

$$K' := \Big\{ x \in H : \Big\| \sum_{i=1}^{n} x_i A_i \Big\|_{\mathsf{op}} \leq 2c(\sigma + (\log^{3/4} d)\sqrt{\sigma \cdot f}) \Big\}.$$

By Markov's inequality and (6.3), it follows that $\gamma_H(K') \geq 1/2 \geq e^{-\delta n}$. Also note that $\dim(H) \geq (1-\delta)n$ since H is defined by δn constraints. It then follows from Lemma 6.2.2 that we can efficiently find a point $x \in (x_0 + K') \cap [-1, 1]^n$ such that $|\{i : |x_i| = 1\}| \geq \varepsilon n/2 = \Omega(n)$. By the definition of K, the guarantee that $x \in x_0 + K'$ translates to

$$\Big\| \sum_{i=1}^{n} (x_i - x_{0,i}) A_i \Big\|_{\mathsf{op}} \leq 2c(\sigma + (\log^{3/4} d)\sqrt{\sigma \cdot f}).$$

This completes the proof of the lemma. As $(\log^{3/4} d)\sqrt{\sigma \cdot f} \leq (\sigma + f(\log d)^{3/2})/2$, this implies a partial coloring discrepancy bound of at most $O(\sigma + (\log d)^{3/2} f)$. □

6.3.2 Proof of Main Theorem

We can now prove Theorem 6.1.1 (restated below) by recursively applying Lemma 6.3.1.

Theorem 6.1.1 (Matrix Spencer Up to Poly-logarithmic Rank)**.** *Given $d \times d$ symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ each with $\|A_i\|_{\mathsf{op}} \leq 1$ and $\|A_i\|_F^2 \leq n/\log^3 n$, there exist signs $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} = O(\sqrt{n})$. Moreover, these signs can be computed efficiently.*

Proof of Theorem 6.1.1. Denote $f^2 := n/\log^3 n$. First, without loss of generality, we can assume that $d \leq n^2$. Indeed, suppose to the contrary that $d > n^2$. Define $M := \sum_{i=1}^n A_i^2$ and note that $\mathsf{tr}(M) = \sum_{i=1}^n \|A_i\|_F^2 \leq nf^2$. By a change of basis, we may assume without loss of generality that M is diagonal and its diagonal entries are in descending order. Note that $M_{n^2,n^2} \leq \mathsf{tr}(M)/n^2 \leq f^2/n$. Define $B_i \in \mathbb{R}^{(d-n^2)\times d}$ the matrix obtained by removing

the first n^2 rows of A_i. We have for any coloring $x \in \{\pm 1\}^n$,

$$\Big\| \sum_{i=1}^n x_i B_i \Big\|_{\mathsf{op}}^2 = \Big\| \Big(\sum_{i=1}^n x_i B_i \Big)^\top \Big(\sum_{i=1}^n x_i B_i \Big) \Big\|_{\mathsf{op}} \leq n \cdot \Big\| \sum_{i=1}^n B_i^\top B_i \Big\|_{\mathsf{op}} \leq f^2,$$

where the inequality follows as $x_i x_j (B_i^T B_j + B_j^T B_i) \preceq (B_i^T B_i + B_j^T B_j)$ for all i, j. Now we let $L_i \in \mathbb{R}^{d \times d}$ be the matrix obtained by zeroing out the top left $n^2 \times n^2$ block of A_i. Since matrices A_i are symmetric, it follows that for any coloring $x \in \{\pm 1\}^n$,

$$\Big\| \sum_{i=1}^n x_i L_i \Big\|_{\mathsf{op}} \leq 2 \Big\| \sum_{i=1}^n x_i B_i \Big\|_{\mathsf{op}} \leq 2f.$$

This shows that we only need to keep the top left $n^2 \times n^2$ block of each matrix A_i without affecting the discrepancy by more than an additive term of $2f$. We thus assume henceforth that $d \leq n^2$.

By assumption, the matrices A_i satisfy $\| \sum_{i=1}^n A_i^2 \|_{\mathsf{op}} \leq n$ and $\sum_{i=1}^n \|A_i\|_F^2 \leq nf^2$. Therefore, we can apply Lemma 6.3.1 with $x_0 = 0$ to obtain a partial coloring $x^{(1)} \in [-1, 1]^n$ with $\| \sum_{i=1}^n x_i^{(1)} A_i \|_{\mathsf{op}} = O(\sqrt{n})$ and $|\{i : |x_i^{(1)}| = 1\}| = \Omega(n)$. Next we let $I_1 := \{i \in [n] : |x_i^{(1)}| < 1\}$, and recursively apply Lemma 6.3.1 to the set of matrices $\{A_i\}_{i \in I_1}$ with point $x^{(1)}|_{I_1}$. Continuing this process of recursively applying Lemma 6.3.1 to the set of coordinates i such that $|x_i| < 1$, the number of such coordinates decreases by a constant factor in each iteration.

Let $x^{(t)} \in [-1, 1]^n$ be the resulting vector in the tth iteration and let n_t denote the number of coordinates in $x^{(t)}$ that are in $(-1, 1)$. Then, we have $n_{t+1} < \lambda n_t$ for some constant $\lambda < 1$ and by using Lemma 6.3.1 with $\sigma \leq \sqrt{n_t}$, we get that the discrepancy increases additively by at most $c(\sqrt{n_t} + (\log^{3/4} d) \cdot f^{1/2} n_t^{1/4})$. Therefore, repeating it for $O(\log n)$ iterations, we get a full coloring with discrepancy at most

$$c \sum_t (\sqrt{n_t} + (\log^{3/4} d) \cdot f^{1/2} n_t^{1/4}) = O(\sqrt{n}) + O((\log^{3/4} n) \cdot f^{1/2} n^{1/4}),$$

where we have used that $d \le n^2$ and n_t's form a geometrically decreasing series. The theorem now follows since we have chosen $f^2 = n/\log^3 n$. □

Remark 6.3.2. *One can also use the version of Lemma 6.2.2 without defining the subspace as in the proof above, but this requires assuming $\|A_i\|_F^2 \le n/\log^4 d$ in Theorem 6.1.1. In particular, by taking $\vec{(A_i')}$ to be the eigenvectors of the matrix $\mathsf{Cov}(X) = \sum_{i=1}^n \vec{(A_i)}\vec{(A_i)}^\top$ in descending order of eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n$, the random matrix $\sum_{i=1}^n g_i \lambda_i^{1/2} A_i'$ has the same distribution as $\sum_{i=1}^n g_i A_i$. One can then guarantee that $\gamma_n(K) \ge 2^{-O(n)}$ by considering the event that $g_1, \cdots, g_k$ are all $1/\mathsf{poly}(n)$ small for $k = \Theta(n/\log n)$, and applying Theorem 6.2.1 to control $\|\sum_{i=k+1}^n g_i \lambda_i^{1/2} A_i'\|_{\mathsf{op}}$.*

6.4 Improvement Over Random Coloring for $o(n)$-rank Matrices

In this section, we sketch how the strategy in the proof of Theorem 6.1.1 can be used to prove Corollary 6.1.2, which we restate below for convenience. In particular, Corollary 6.1.2 improves over the random coloring bound of $O(\sqrt{n \log d})$ whenever the matrices have $\mathsf{rank}(A_i) = o(n)$, and in particular for all dimension $d = o(n)$.

Corollary 6.1.2 (Improvement Over Random Coloring)**.** *Given $d \times d$ symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ each with $\|A_i\|_{\mathsf{op}} \le 1$ and $\|A_i\|_F^2 \le r$, there exist signs $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} = O\left(\sqrt{n}(\log d)^{1/4} + (nr)^{1/4}\sqrt{\log d}\right)$. Moreover, these signs can be computed efficiently.*

Proof of Corollary 6.1.2 (sketch). The main observation is that instead of using the bound given by Theorem 6.2.1, one can combine Corollary 3.6 in [Tro18] with Proposition 4.6 in [BBvH21] to obtain the bound

$$\mathbb{E}[\|X\|_{\mathsf{op}}] = O\left((\log d)^{1/4}\sigma(X) + (\log d)^{1/2}\sqrt{v(X)\sigma(X)}\right) \tag{6.4}$$

for any symmetric Gaussian random matrix $X \in \mathbb{R}^{d \times d}$. Notice the improved $(\log d)^{1/2}$ factor in the second term here compared to the factor of $(\log d)^{3/4}$ in Theorem 6.2.1, but at the

expense of the worse $(\log d)^{1/4}$ factor in the first term.

As $\mathsf{tr}(\mathsf{Cov}(X)) \leq nr$ for $r = \max_i \mathsf{rank}(A_i)$, by the argument in the proof of Lemma 6.3.1 we can assume that $v(X)^2 = O(r)$ by restricting to the subspace H orthogonal to the large eigenvectors of $\mathsf{Cov}(X)$. Plugging this into (6.4), the argument in Lemma 6.3.1 implies a partial coloring with discrepancy $O\left(\sqrt{n}(\log d)^{1/4} + (nr)^{1/4}\sqrt{\log d}\right)$, improving upon the random coloring bound of $O(\sqrt{n \log d})$ whenever $r = o(n)$. Finally, as in the proof of Theorem 6.1.1, the $O(\log n)$ iterations of partial coloring to get a full coloring only leads to an $O(1)$ factor loss overall. □

Chapter 7

ONLINE DISCREPANCY I: CHANGE OF BASIS

In this chapter and the next two chapters, we turn to the study of online discrepancy. The online discrepancy question was first studied by Spencer [Spe77] in the 1970s, who showed that the $\sqrt{T}$ discrepancy by random coloring cannot be improved against adaptive adversaries. In this chapter, we give an algorithm with $\mathsf{poly}(\log T)$ discrepancy for this problem in the stochastic setting. We also give the first poly-logarithmic discrepancy algorithm for online geometric discrepancy problems. This chapter is based on a joint paper with Nikhil Bansal, Sahil Singla, and Makrand Sinha [BJSS20] that appeared in the *52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC 2020)*.

7.1 Introduction

Consider the following online vector balancing question, originally proposed by Spencer [Sho77]: vectors $v_1, v_2, \ldots, v_T \in [-1, 1]^n$ arrive online, and upon the arrival of v_t, a sign $\chi_t \in \{\pm 1\}$ must be chosen irrevocably, so that the ℓ_∞-norm of the signed sum $d_t = \chi_1 v_1 + \ldots + \chi_t v_t$ remains as small as possible. That is, find the smallest B such that $\max_{t \in [T]} \|d_t\|_\infty \leq B$. As we shall see later, the problem arises naturally in various contexts where one wants to divide an incoming stream of objects, so that the split is as even as possible along each of the various dimensions that one might care about.

A naïve algorithm is to pick each sign χ_t randomly and independently, which by standard tail bounds gives $B = \Theta((T \log n)^{1/2})$ with high probability. In most of the interesting settings, $T \gg n$, and a natural question is whether the dependence on T can be improved from $T^{1/2}$ to say, $\log T$, or removed altogether (possibly with a worse dependence on n).

Offline setting. The offline version of the problem, where the vectors $v_1, \ldots, v_T$ are given in advance and the goal is to minimize $\max_{t\in[T]} \|d_t\|_\infty$, is known as the signed-series problem. It was first studied by Spencer [Sho77], who obtained a bound independent of T, but exponential in n. This was later improved by Bárány and Grinberg [BG81] to $B \leq 2n$. Chobanyan [Cho94] showed a beautiful connection between the signed-series problem and the classic Steinitz problem on the rearrangement of vector sequences—any upper bound on B also holds for the latter problem. Steinitz problem has a much longer history, originating from a question of Riemann and Lévy in the 19th century (c.f. the survey [Bár08] for some fascinating history). A long-standing conjecture for both the problems, still open, is that $B = O(n^{1/2})$. Another notable bound is due to Banaszczyk [Ban12], who showed that $B = O((n \log T)^{1/2})$. While the original argument in [Ban12] was non-constructive, a polynomial time algorithm to find such a signing was recently given in [BG17].

In general, there has been extensive work on various offline discrepancy problems over last several decades, and several powerful techniques such as the partial coloring method [Spe85] and convex geometric methods [Gia97, Ban98, Ban12, MNT14] have been developed, which significantly improve upon the bounds given by random coloring. While these initial methods were mostly non-algorithmic, several new algorithmic techniques and insights have been developed in recent years [Ban10, LM15b, Rot17, ES18, BDG16, LRR17, BDGL18, DNTTJ18].

Online setting. The online setting was first studied in the 70's and 80's, but it did not receive much interest later as it was realized that the best guarantees are already achieved by trivial algorithms. In particular, the $T^{1/2}$ dependence on T achieved by random coloring cannot be improved [Sho77]. See [Spe87, Bár79] for even more specific lower bounds. The difficulty is that the all-powerful adversary, upon seeing the signs chosen by the algorithm until time $t-1$, can choose the next input vector v_t to be *orthogonal* to d_{t-1}. Now, irrespective

of the choice of the sign χ_t, the resulting signed sum d_t satisfies

$$\|d_t\|_2^2 \;=\; \|d_{t-1} + \chi_t v_t\|_2^2 \;=\; \|d_{t-1}\|_2^2 + 2\chi_t\langle d_{t-1}, v_t\rangle + \|v_t\|_2^2 \;=\; \|d_{t-1}\|_2^2 + \|v_t\|_2^2. \qquad (7.1)$$

For any d_{t-1}, one can always pick v_t with[1] $\|v_t\|_\infty \le 1$ and $\|v_t\|_2^2 \ge n-1$, resulting in $\|d_t\|_2^2 \ge (n-1)t$, and hence $\|d_t\|_\infty = \Omega(t^{1/2})$ for all $t \in [T]$ (as long as $n > 1$).

It is therefore natural to ask if relaxing the power of the adversary, or making additional assumptions on the input sequence, can lead to interesting new ideas and to algorithms that perform much better, and in particular, give bounds that only mildly depend on T.

A natural assumption is that of *stochasticity*: if the arriving vectors are chosen in an i.i.d. manner from some distribution p, can we maintain that the ℓ_∞ norm of the current signed-sum d_t—henceforth, referred to as discrepancy—is $\mathsf{poly}(n)$ or $\mathsf{poly}(n, \log T)$?

Previous work and challenges. Recently, this stochastic setting was studied by Bansal and Spencer [BS20], where they considered the case where p is the uniform distribution on all $\{-1, 1\}^n$ vectors. They give an online algorithm achieving a bound of $O(\sqrt{n})$ on the *expected* discrepancy, matching the best possible offline bound, and an $O(\sqrt{n}\log T)$ discrepancy bound at all times $t \in [T]$, with high probability.

In general, the algorithmic discrepancy approaches developed in the last decade do not seem to help in the online setting. This is because in the offline setting, the algorithms can ensure that the discrepancy stays low by *simultaneously* updating the colors of various elements in a correlated way. In the online setting, however, the discrepancy must necessarily rise (in the ℓ_2 sense) whenever the incoming vector v_t is almost orthogonal to d_{t-1}, which can happen quite often. The only thing that the online algorithm can do is to *actively* try to *cancel* this increase, whenever possible, by choosing the sign χ_t cleverly.

The algorithm of [BS20] crucially uses that if the coordinates of v_t are independently

[1] For any $d \in \mathbb{R}^n$, any basic feasible solution to $\langle d, x\rangle = 0$ with $x \in [-1, 1]^n$ has at least $n-1$ coordinates ± 1.

distributed and mean-zero[2], then for any d_{t-1} the incoming vector v_t will typically be far from being orthogonal to d_{t-1}. More quantitatively, the *anti-concentration* property for independent random variables gives that for any $d_{t-1} = (d_1, \ldots, d_n)$, the random vector $v_t = (X_1, \ldots, X_n)$ with $X_1, \ldots, X_n$ being independent and mean-zero satisfies

$$\mathbb{E}_v\Big[|\langle d_{t-1}, v_t\rangle|\Big] = \Omega\left(\Big(\sum_{i=1}^{n} d_i^2 \cdot \mathbb{E}[X_i]^2\Big)^{1/2}\right).$$

Whenever $|\langle d_{t-1}, v_t\rangle|$ is large, the algorithm can choose χ_t appropriately to create a *negative drift* in (7.1), to offset the increase due to the $\|v_t\|^2$ term. We give a more detailed description below in §7.2.1.

In many interesting settings, however, the X_i's can be *dependent*. For example, motivated by an envy minimization problem, Jiang, Kulkarni, and Singla [JKS19] considered the following natural online interval discrepancy problem: points $x_1, \ldots, x_T$ arrive uniformly in the interval $[0, 1]$, and the goal is to assign them signs online to minimize the discrepancy of every sub-interval of $[0, 1]$. (For adversarial arrivals, [JKS19] show $\mathsf{poly}(T)$ lower bounds.) Viewing the sub-intervals (after proper discretization) as coordinates, this becomes a stochastic online vector balancing problem, but where the random variables X_i corresponding to the various sub-intervals are dependent (details in §7.2.2). They give a non-trivial algorithm that achieves $T^{1/\log\log T}$ discrepancy, which is much better than the $T^{1/2}$ bound obtained by random coloring, but still substantially worse than $\mathrm{polylog}(T)$.

In general, the difficulty with dependent coordinates X_i is that even a small correlation can destroy anti-concentration, which makes it difficult to create a negative drift. For example, suppose the distribution p is mostly supported on vectors with an equal number of $+1$ and -1 coordinates. Now if d has the form $d = c(1, \ldots, 1)$, then the incoming vector v_t is almost always orthogonal to it, and $\|d_T\|_2$ can potentially increase as fast as $\Omega(T^{1/2})$.

In this chapter, we focus on the stochastic setting where the coordinates have depen-

[2]Note that this holds in the case of uniform distribution over $\{-1, 1\}^n$.

dencies, and give several results both for specific geometric problems and for general vector balancing under arbitrary distributions. In general, there are various other ways in which one can relax the power of the adversary, and in §7.8 we describe several interesting open questions and directions in this area.

7.1.1 Our Discrepancy Bounds

We first consider the following interval discrepancy problem. Let $x = x_1, \ldots, x_T$ be a sequence of points drawn uniformly in $[0, 1]$ and let $\chi_1, \ldots, \chi_T \in \{\pm 1\}$ be a signing. For an interval $I \subseteq [0, 1]$, let $\mathbf{1}_I$ denote the indicator function of the interval I. For any time $t \in [T]$, we define the discrepancy of interval I to be

$$\mathsf{disc}_t(I) := \left|\chi_1 \mathbf{1}_I(x_1) + \cdots + \chi_t \mathbf{1}_I(x_t)\right|.$$

We show the following bounds on discrepancy.

Theorem 7.1.1 (Interval Discrepancy)**.** *There is an online algorithm which selects signs $\chi_t \in \{\pm 1\}$ such that, with high probability*[3]*, for every interval $I \subseteq [0, 1]$ we have $\max_{t\in[T]} \mathsf{disc}_t(I) = O(\log^3 T)$. Moreover, with constant probability, for any online algorithm,*

$$\max_{I\subseteq[0,1]} \max_{t\in[T]} \mathsf{disc}_t(I) = \Omega\left(\sqrt{\log T}\right).$$

This gives an exponential improvement over the $T^{1/\log\log T}$ bound of [JKS19], and is tight up to polynomial factors. The lower bound also improves a previous bound of $\Omega(\log^{1/4} T)$ of [JKS19].

There are two natural d-dimensional generalizations of the interval discrepancy problem, and our framework, which we will describe in §4.3, can handle both of them.

[3]Throughout the chapter, "with high probability" means with $1 - 1/\mathsf{poly}(n, T)$ probability where the exponent of the polynomial can be made as large as desired, depending on the constant in the discrepancy upper bound.

d-dimensional Online Interval Discrepancy: Consider a sequence of points $x_1, \ldots, x_T$ drawn uniformly from the unit cube $[0,1]^d$. The goal is to simultaneously minimize the discrepancy of every interval for all the d-coordinates. In other words, to minimize the following for every interval I and every coordinate $i \in [d]$:

$$\mathsf{disc}_t^i(I) := \Big|\chi_1 \mathbf{1}_I(x_1(i)) + \ldots + \chi_t \mathbf{1}_I(x_t(i))\Big|.$$

The offline version of this problem for $d \geq 2$ is equivalent to the classic d-permutations problem, where an upper bound of $O(\sqrt{d}\log T)$ [SST97] and a breakthrough lower bound of $\Omega(\log T)$ [NNN12, Fra21] for $d \geq 3$, and $\Omega(\sqrt{d})$ in general is known for the worst-case placement of points.

We show the following generalization of Theorem 7.1.1 that matches the best offline bounds, up to polynomial factors.

Theorem 7.1.2 (d-dimensional Interval Discrepancy)**.** *There is an online algorithm which selects signs $\chi_t \in \{\pm 1\}$ such that, with high probability, for each $i \in [d]$ and $I \subseteq [0,1]$, we have $\max_{t\in[T]} \mathsf{disc}_t^i(I) = O(d\log^3 T)$. Moreover, with constant probability, for any online algorithm there exists an interval I and a coordinate $i \in [d]$, such that $\max_{t\in[T]} \mathsf{disc}_t^i(I)$ $= \Omega\big(\sqrt{d\log(T/d)}\big)$.*

Previously, Jiang et al. [JKS19] could extend their analysis for online interval discrepancy to the $d = 2$ case and prove the same $T^{1/\log\log T}$ bound. However, their proof is rather ad-hoc and does not seem to generalize to higher d. In contrast, our bound holds for any d, and is tight up to polynomial factors.

The second natural generalization of interval discrepancy is to d-dimensional axis-parallel boxes, which gives the following online version of the extensively studied Tusnády's Problem.

d-dimensional Online Tusnády's Problem: Consider a sequence of points $x_1, \ldots, x_T$ drawn uniformly from the unit cube $[0,1]^d$. The goal is to simultaneously minimize the discrepancy

of all axis-parallel boxes. In other words, to minimize the following for every box B:

$$\mathsf{disc}_t(B) := \left|\chi_1 \mathbf{1}_B(x_1) + \ldots + \chi_t \mathbf{1}_B(x_t)\right|.$$

The (offline) Tusnády's problem has a fascinating history (see [Mat99] and references there in), and after a long line of work, it is known that for the worst-case placement of points, the offline discrepancy is at most $O_d(\log^{d-\frac{1}{2}} T)$ [Nik17] and at least $\Omega_d(\log^{d-1} T)$ [MN15]. We show the following result in the online setting, which is tight to within polynomial factors.

Theorem 7.1.3 (Tusnády's Problem)**.** *There is an online algorithm which selects signs $\chi_t \in \{\pm 1\}$ such that, with high probability, for every axis-parallel box B, we have $\max_{t\in[T]} \mathsf{disc}_t(B) = O_d(\log^{2d+1} T)$. Moreover, for any online algorithm, with constant probability, there exists a box B such that $\max_{t\in[T]} \mathsf{disc}_t(B) = \Omega_d(\log^{d/2} T)$.*

In contrast, the proof approach of [JKS19] completely breaks down for the Tusnády's problem even in two dimensions and does not give any better lower bounds in terms of d. We recently learned that results similar to Theorems 7.1.1 and 7.1.3 were also obtained by Dwivedi et al. [DFGGR19], in the context of understanding the power of online thinning in reducing discrepancy.

Remark: Although all the problems above are stated for uniform distributions, one can use the probability integral transformation to reduce any product distribution to the uniform distribution without increasing the discrepancy, so our results in Theorems 7.1.2 and 7.1.3 also apply to any product distribution over $[0,1]^d$.

Finally, note that Theorem 7.1.1 follows as a direct corollary of either of the above theorems.

General distributions. We now consider the setting of *arbitrary* distributions for the online vector balancing problem. Here we need to tackle the orthogonality issue which gave $\Omega(T^{1/2})$ lower bounds discussed in (7.1). As discussed earlier, for the uniform distribution over $\{-1,+1\}^n$, Bansal and Spencer [BS20] get around this issue since this does not happen

for the uniform distribution reasonably often, and hence, $\mathbb{E}[\langle d_{t-1}, v_t\rangle]$ is large for any vector d_{t-1}. Using this, they obtain the bound $O(n^{1/2}\log T)$. Our next result shows that such a $\mathsf{poly}(n, \log T)$ upper bound is possible even for arbitrary distributions.

Theorem 7.1.4. (Vector Balancing Under Dependencies) *For any sequence of vectors $v_1, \ldots, v_T \in [-1,1]^n$ sampled i.i.d. from some arbitrary distribution p, there is an online algorithm which selects signs $\chi_t \in \{\pm 1\}$ such that, with high probability, we have*

$$\max_{t\in[T]} \left\|\chi_1 v_1 + \ldots + \chi_t v_t\right\|_\infty = O(n^2(\log T + \log n)).$$

In §7.4.2 we show that the dependencies on n and $\log T$ in this theorem are tight up to polynomial factors as there is an $\Omega(n^{1/2} + (\log T/\log\log T)^{1/2})$ lower bound.

All of the above results follow from a general framework that we discuss next. In addition to the framework below, the key new technical ingredient is an anti-concentration inequality for dependent random variables, which we describe below in Theorem 7.1.5. This may be of independent interest.

7.1.2 Our Framework

To tackle the orthogonality issue, one of our key idea is to work with a *different basis* for the discrepancy vectors. More specifically, instead of maintaining bounds on the individual coordinate discrepancies $d_t(i)$, we maintain bounds on suitable linear combinations of them. This basis ensures that the (new) coordinates of the incoming vector are *uncorrelated*, i.e., $\mathbb{E}[X(i)\cdot X(j)] = \mathbb{E}[X(i)]\cdot\mathbb{E}[X(j)]$ for distinct coordinates i, j. Note that this condition is only on the expected values, and is much weaker, e.g., even pairwise independence. Once one finds a suitable new basis, which turns out to be an eigenbasis of the covariance matrix, the anti-concentration bound for such random variables (proved below in Theorem 7.1.5), together with the standard exponential penalty based framework used in previous works [BS20, JKS19], gives Theorem 7.1.4.

For our results on geometric discrepancy problems, there is an additional challenge, we

cannot afford to lose a $\mathsf{poly}(n)$ factor, as in Theorem 7.1.4 above, since the dimension $n = \Theta(T)$. In this case, however, the update vectors are $(\log T)$-sparse in the original basis (see §7.2) and one could hope to utilize this sparsity. Yet another challenge in this case is that bounding the discrepancy in a new basis preserves ℓ_2-discrepancy in the original basis, but could lead to a $\sqrt{n}$ loss in ℓ_∞-discrepancy. To get $\mathsf{polylog}(T)$ bounds, we use a natural basis from wavelet theory, called the *Haar system*, which simultaneously has sparsity, uncorrelation, and avoids the ℓ_2 to ℓ_∞ loss. This also easily extends to higher dimensions as these wavelets can be tensorized in a natural way to get a suitable basis for higher dimensional versions of the problems. A more detailed description of our framework is given in §7.2. Next we discuss our anti-concentration results.

7.1.3 Our Anti-Concentration Results for Non-Independent Random Variables

Suppose $X_1, \ldots, X_n$ are independent $\{-1, +1\}$ random variables with mean zero. Then, it is well-known that $|\sum_i X_i|$ has mean $\Theta(n^{1/2})$, and moreover, this value is at least $\Omega(n^{1/2})$ with constant probability.

Now, on the other hand, consider the following distribution. Let H_n be $n \times n$ Hadamard matrix and let $H_n(i)$ denote its i-th row for $i \in [n]$. Consider the random vector $X = (X_1, \ldots, X_n)$, where $X = \xi \cdot H_n(i)$ for a Rademacher random variable $\xi \in \{-1, +1\}$ and a uniformly chosen $i \in [n]$. Then the X_i's are still mean-zero and $\{-1, +1\}$, and in fact, pairwise independent. However, the magnitude of the sum $|\sum_i X_i|$ behaves very differently from the i.i.d. setting above. It takes value n with probability only $1/n$ (if $X = \xi \cdot H_n(1)$, the row of all 1's) and is 0 otherwise. In particular the mean is $\mathbb{E}[|\sum_i X_i|] = 1$ (instead of $n^{1/2}$ above), and moreover the entire contribution to the mean comes from an event with probability only $1/n$.

Nevertheless, we can say interesting things about the anti-concentration of sums of such random variables. In particular, we show the following results for uncorrelated or pairwise independent random variables.

Theorem 7.1.5. (Uncorrelated anti-concentration) *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$, let $X_1, \ldots, X_n$ be uncorrelated random variables that are bounded $|X_i| \le c$, satisfy $\mathbb{E}[X_i X_j] = 0$ for all $i \neq j$, and have sparsity s (the number of non-zero X_i's in any outcome). Then*

$$\mathbb{E}\Big|\sum_i a_i X_i\Big| \;\ge\; \mathbb{E}\Big[\sum_i |a_i| X_i^2\Big] \cdot \frac{1}{cs}. \tag{7.2}$$

Moreover, this bound is tight, even for pairwise independent random variables.

The tightness holds for the Hadamard example above, where $\mathbb{E}|\sum_i X_i| = 1$, $s = n$, $c = 1$, and $\mathbb{E}[\sum_i X_i^2] = n$.

Theorem 7.1.6. (Pairwise independent anti-concentration) *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$, let $X_1, \ldots, X_n$ be mean-zero pairwise independent random variables with sparsity $s \le n$. Then*

$$\mathbb{E}\Big[\Big|\sum_i a_i X_i\Big|\Big] \;\ge\; \mathbb{E}\Big[\sum_i |a_i X_i|\Big] \cdot \frac{1}{s}. \tag{7.3}$$

Note that this bound is also tight for the Hadamard example. In general, the bound (7.3) is stronger than in (7.2); and a simple example in §7.3.2 shows that (7.3) cannot hold for uncorrelated random variables.

Although the anti-concentration properties and the small-ball probabilities for independent variables have been extensively studied (c.f. [NV13]), the uncorrelated and pairwise independent setting does not seem to have been studied before, and Theorems 7.1.5 and 7.1.6 do not seem to be known, to the best of our knowledge.

7.1.4 Applications to Envy Minimization

A classic measure of fairness in the field of fair division is envy [Fol66, TV85, LMMS04, Bud11]. A recent work of Benade et al. [BKPP18] introduced the *online envy minimization* problem where T items arrive one-by-one. In the two player setting, on arrival of item $t \in \{1, \ldots, T\}$ we get to see the valuations $v_{it} \in [0, 1]$ for both the players $i \in \{1, 2\}$.

The goal is to immediately and irrevocably allocate the item to one of the players while minimizing the maximum *envy*. There are two natural notions of envy: cardinal and ordinal (see §7.7 for definitions). Benade et al. [BKPP18] show an $\Omega(T^{1/2})$ lower bound for online envy minimization in the *adversarial* model—the reason is similar to Bárány's [Bár79] lower bound for online discrepancy. Can we obtain better bounds when the player valuations are drawn from a distribution?[4]

In the special case of product distributions (each player independently draws their value), Jiang et al. [JKS19] observed that the 2-dimensional interval discrepancy bounds also hold for online envy minimization. In particular, they obtained a $T^{1/\log\log T}$ bound on the ordinal envy. Our new interval discrepancy bound from Theorem 7.1.2 immediately improves this to an $O(\log^3 T)$ bound on ordinal envy. Moreover, we use our vector balancing result to obtain an $O(\log T)$ bound on the cardinal envy even for general distributions.

Corollary 7.1.7. *Suppose valuations of two players are drawn i.i.d. from some distribution* p *over* $[0,1]\times[0,1]$. *Then, for an arbitrary distribution* p *(i.e., player valuations for the same item could be correlated), the online cardinal envy is* $O(\log T)$. *Moreover, if* p *is a product distribution (i.e., player valuations for the same item are independent) then the online ordinal envy is also* $O(\log^3 T)$.

Chapter Organization

The rest of the chapter is organized as follows: in §7.2, we give an overview of previous challenges and our main ideas. In §7.3, we prove our key anti-concentration theorems that are necessary for our upper bounds on discrepancy. In §7.4, we give upper and lower bounds for online discrepancy under certain "uncorrelation" assumptions on the distribution. Then, we apply these bounds in §7.5 to obtain our vector balancing result (Theorem 7.1.4). In §7.6, we again apply these bounds to obtain our geometric discrepancy results (Theorems 7.1.2

[4] If we make a simplifying assumption that the distribution does not depend on the time horizon T, better bounds are known [ZP19, DGK+14].

and 7.1.3). In §7.7, we show why our results immediately apply to online envy minimization. Finally, in §7.8 we end with some discussion of open problems and directions.

7.2 Proof Overview

Let us start by reviewing the approach considered by Bansal and Spencer [BS20] in the case of independent coordinates. We also discuss the challenges involved in extending it to the setting of dependent coordinates.

7.2.1 Independent Coordinates: Bansal and Spencer

Consider the online vector balancing problem, when each arriving vector is uniformly chosen from $\{\pm 1\}^n$, so that all the coordinates are independent. To design an online algorithm, it is natural to keep a potential function that keeps track of the discrepancy and chooses a sign χ_t for the current vector v_t that minimizes the increase in the potential. Formally, let $d_t = \chi_1 v_1 + \ldots + \chi_t v_t$ denote the discrepancy vector at time t. For a parameter $0 < \lambda < 1$, define the potential function

$$\Phi_t = \sum_{i \in [n]} \cosh(\lambda d_t(i)),$$

where $d_t(i)$ denotes the ith coordinate of d_t and $\cosh(x) = \frac{1}{2} \cdot (e^x + e^{-x})$ for all $x \in \mathbb{R}$. One should think of the above potential function as a proxy for the maximum discrepancy as Φ_t is dominated by the maximum discrepancy: $\Phi_t \approx e^{\lambda \|d_t\|_\infty}$.

On the arrival of vector v_t, the algorithm chooses a sign $\chi_t \in \{\pm 1\}$, which updates the discrepancy vector to $d_t = d_{t-1} + \chi_t v_t$ and changes the potential from Φ_{t-1} to Φ_t. If we can show that whenever $\Phi_t > 2n$, the *drift* $\Delta\Phi_t := \Phi_t - \Phi_{t-1}$ is negative in expectation for the sign χ_t chosen by the algorithm, then we can say that the potential after T arrivals, Φ_T, is bounded by $\mathsf{poly}(nT)$ with high probability. This implies $\cosh(\lambda \|d_T\|_\infty)$ is bounded by $\mathsf{poly}(nT)$, which means a bound of $O(\lambda^{-1} \log T)$ on the maximum discrepancy.

Let us try to compute the expected drift. Define $d = d_{t-1}$. By considering the Taylor expansion, we get $\cosh(x+\delta) \leq \cosh(x) + \sinh(x)\delta + \cosh(x)\delta^2$ where $\sinh(x) = \frac{1}{2} \cdot (e^x - e^{-x})$

for all $x \in \mathbb{R}$. So,

$$\Delta\Phi_t \;\approx\; \sum_{i\in[n]} \Big(\lambda \sinh(\lambda d(i)) \cdot (\chi_t v_t(i)) + \lambda^2 \cosh(\lambda d(i)) \cdot (\chi_t v_t(i))^2\Big) \;=\; \chi_t \lambda L + \lambda^2 Q,$$

where $L = \sum_{i\in[n]} \sinh(\lambda d(i)) \cdot v_t(i)$ is the *linear* term and $Q = \sum_{i\in[n]} \cosh(\lambda d(i))$ is the *quadratic* term from the Taylor expansion (note that $(\chi_t v_t(i))^2 = 1$). Since the algorithm is free to choose the sign χ_t to minimize the drift, $\Delta\Phi_t \approx -\lambda|L| + \lambda^2 Q$. Now if one can show that $\mathbb{E}_{v_t}[|L|] \geq \frac{\mathbb{E}[Q]}{2\lambda}$, we would get that the expected drift $\mathbb{E}[\Delta\Phi_t] < 0$, and this would translate to a good discrepancy bound of $O(\lambda^{-1} \log T)$ if λ is large as described above.

Since $\cosh(x)$ and $|\sinh(x)|$ only differ by at most 1, we can make the approximation $Q \approx \sum_{i\in[n]} |\sinh(\lambda d(i))|$ up to some small error. So, denoting $\beta = 1/\lambda$ and $a_i = \sinh(\lambda d(i))$, our task reduces to proving the following anti-concentration statement:

Question. Let $X_1, \ldots, X_n$ be independent random variables with $|X_i| \leq 1$. What is the smallest β such that the following holds:

$$\mathbb{E}\Big[\Big|\sum_{i\in[n]} a_i X_i\Big|\Big] \geq \frac{1}{\beta} \cdot \mathbb{E}\Big[\sum_{i\in[n]} |a_i| X_i^2\Big]. \tag{7.4}$$

In the case where the X_i's are independent Rademacher (± 1) random variables, classical Khintchine's inequality and Cauchy-Schwarz tell us that

$$\mathbb{E}\Big[\Big|\sum_{i\in[n]} a_i X_i\Big|\Big] \;\geq\; \frac{1}{\sqrt{2}} \cdot \Big(\sum_{i\in[n]} a_i^2\Big)^{1/2} \;\geq\; \frac{1}{\sqrt{2n}}\Big(\sum_{i\in[n]} |a_i|\Big) \;=\; \frac{1}{\sqrt{2n}} \cdot \mathbb{E}\Big[\sum_{i\in[n]} |a_i| X_i^2\Big],$$

so $\beta = O(\sqrt{n})$, which suffices for the discrepancy application. In general, when X_i's are not Rademacher but are still bounded ($|X_i| \leq 1$), mean-zero, and *independent*, then following [BS20] one can still show that $\beta = O(\sqrt{n})$.

The above gives a bound of $O(\sqrt{n} \log T)$ on the maximum discrepancy at every time $t \in [T]$. However, when the input distribution has dependencies across coordinates, i.e. the X_i's are dependent, one can not take β to be small in general. For example, $\beta \to \infty$ when

all a_i's are one and a random set of coordinates $S \subset [n]$ of size $n/2$ (say n is even) take value $+1$ and the remaining coordinates in $[n] \setminus S$ take value -1.

Next we discuss the simplest geometric discrepancy problem—the interval discrepancy problem in one dimension—where such a situation already arises if we use the same approach as above.

7.2.2 Interval Discrepancy: Previous Barriers

Recall, we have T points $x_1, \ldots, x_T$ chosen uniformly from $[0, 1]$ which need to be given ± 1 signs online. Consider the *dyadic* intervals $I_{j,k} := [k2^{-j}, (k+1)2^{-j}]$ where $0 \leq k < 2^j$ and $0 \leq j \leq \log T$. For intuition, imagine embedding the unit interval on a complete binary tree of height $\log T$; now sub-intervals corresponding to every node of the binary tree are dyadic intervals. Note that the smallest dyadic interval has size $2^{-\log T} = 1/T$. By a standard reduction, every sub-interval of $[0, 1]$ is contained in a union of some $O(\log T)$ dyadic intervals, so it suffices to track the discrepancy of these dyadic intervals.

Denoting by $\mathbf{1}_I$ the indicator function for an interval I, define

$$d_t(I) := \chi_1 \mathbf{1}_I(x_1) + \ldots + \chi_t \mathbf{1}_I(x_t).$$

Note that $|d_t(I_{j,k})|$ is the discrepancy of the interval $I_{j,k}$ at time t. A natural choice of algorithm is to use the potential function

$$\Phi_t = \sum_{j,k} \cosh(\lambda d_t(I_{j,k})),$$

which is a proxy for the maximum discrepancy of any dyadic interval. Ideally, we want to set $0 < \lambda < 1$ as large as possible. Defining $d_{j,k} = d_{t-1}(I_{j,k})$, and doing a similar analysis as before, we derive

$$\Delta\Phi_t \approx \chi_t \lambda L + \lambda^2 Q,$$

where $L = \sum_{j,k} \sinh(\lambda d_{j,k}) \cdot \mathbf{1}_{I_{j,k}}(x_t)$ and $Q = \sum_{j,k} \cosh(\lambda d_{j,k}) \cdot \mathbf{1}_{I_{j,k}}(x_t)^2$. The problem

again reduces to showing an anti-concentration statement as in Eq. (7.4) with X_i's being the indicators $\mathbf{1}_{I_{j,k}}$ for all j, k. It turns out that the smallest β one can hope for this setting is exponential in the height of the tree (see Appendix 7.9 for an example), which for binary trees of height $\log T$ only yields a $\mathsf{poly}(T)$ bound on the discrepancy.

One can still leverage something out of this approach—letting $B = T^{1/\log\log T}$, it was shown by Jiang, Kulkarni, and Singla [JKS19] that by embedding B-adic intervals on a B-ary tree of height $\log\log T$, the above approach gives a sub-polynomial $T^{1/\log\log T}$ bound for the interval discrepancy problem. However, this cannot be pushed to give a $\mathsf{polylog}(T)$ bound because the above obstruction does not allow us to handle trees of height $\log T$.

7.2.3 Interval Discrepancy: A New Potential and the BDG Inequality

To get around the previous problem, we take a different approach and instead of directly using the discrepancies in the potential Φ_t, we work with linear combinations of discrepancies with the following desirable properties. First, if there is a bound on these linear combinations then it should imply a bound on the original discrepancies. Second, and more importantly, the term L in $\Delta\Phi_t$ can be viewed as a martingale, which leads to much better anti-concentration properties, i.e., smaller β in (7.4).

More specifically, consider the previous embedding of the dyadic intervals of length at least $1/T$ on the complete binary tree of depth $\log T$. For any interval $I_{j,k}$, let the left half interval be $I^l_{j,k}$ and the right half interval be $I^r_{j,k}$, and consider the difference (see Figure 7.1) of their discrepancies

$$d^-_t(I_{j,k}) := d_t(I^l_{j,k}) - d_t(I^r_{j,k}).$$

Note that if $|d_t(I_{j,k})| \le \alpha$ and also $|d^-_t(I_{j,k})| \le \alpha$, then both $|d_t(I^l_{j,k})| \le \alpha$ and $|d_t(I^r_{j,k})| \le \alpha$. A simple inductive argument now shows that if $|d_t([0,1])| \le \alpha$ and the differences of discrepancy for every dyadic interval $I_{j,k}$ satisfies $|d^-_t(I_{j,k})| \le \alpha$, then every dyadic interval also has discrepancy at most α, thus satisfying the first property above. So let us consider a different

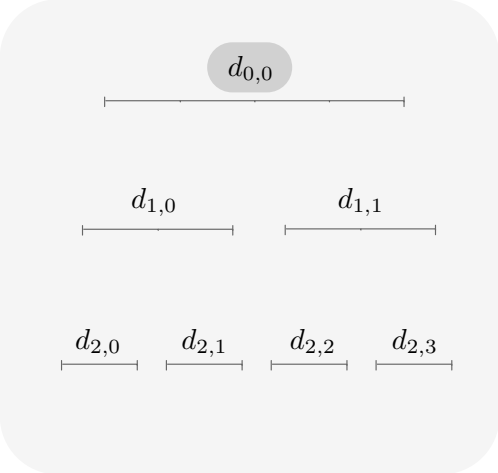
potential function:

$$\Xi_t := \cosh(\lambda d_t(I_{0,0})) + \sum_{j,k} \cosh(\lambda d_t^-(I_{j,k}))$$

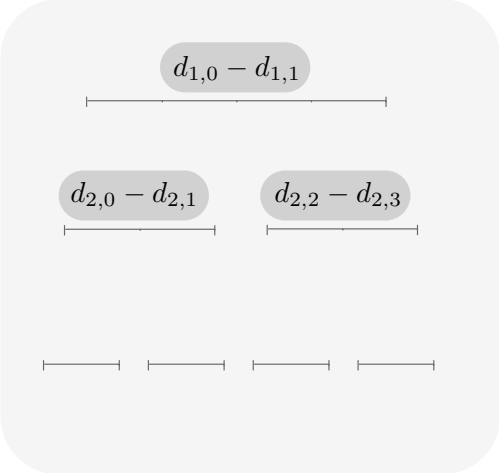
with j, k ranging over all the dyadic intervals (corresponding to internal nodes of the tree) and $0 < \lambda < 1$ is a parameter that we want to set as large as possible. Denoting $d_{j,k}^- = d_{t-1}^-(I_{j,k})$, as before, we can write $\Delta\Xi_t \approx \chi_t \lambda L + \lambda^2 Q$, with

$$L = \sinh(\lambda d_t(I_{0,0})) + \sum_{j,k} \sinh(\lambda d_{j,k}^-) \cdot X_{j,k}(x_t) \quad \text{and}$$
$$Q = \cosh(\lambda d_t(I_{0,0})) + \sum_{j,k} \cosh(\lambda d_{j,k}^-) \cdot X_{j,k}(x_t)^2,$$

where $X_{j,k} = \mathbf{1}_{I_{j,k}^l} - \mathbf{1}_{I_{j,k}^r}$ for any interval $I_{j,k}$. Note that $X_{j,k}$ takes value 1 on the left half of $I_{j,k}$, and -1 on the right half of $I_{j,k}$, and is zero otherwise.



(a) The discrepancy $d_{j,k}$ terms for intervals $I_{j,k}$



(b) The difference of discrepancy $d_{j,k}^- := d_t^-(I_{j,k})$ terms for intervals $I_{j,k}$

Figure 7.1: Some terms appearing in the new potential function Ξ_t. Note that the hyperbolic cosine for the highlighted terms appears in Ξ_t.

Anti-concentration via Martingale analysis. Now we show how the random variable L can be viewed as a $(\log T)$-step martingale. Let us view a uniform point $x \in [0, 1]$ as being

sampled one *bit* at a time, starting with the most significant bit. At any point where j bits of x have been revealed, the interval $I_{j,k}$ on the j^{th} level of the dyadic tree is determined. Now, consider the process that starts with the value $Y_0 = \sinh(\lambda d_{0,0})$ at the root and at any time $0 \leq j \leq \log T$, the process is on some node of the j^{th} level. Conditioned on this node being $I_{j,k}$, the payoff $Y_j := a_j X_j$ where $a_j = \sinh(d^-_{j,k})$ and X_j equals 1 if the process moves to the left child and equals -1 otherwise. Defining $L_j = Y_0 + Y_1 \ldots + Y_j$, it follows that the sequence $L_0, \ldots, L_{\log T}$ is a martingale and $L = L_{\log T}$.

Moreover, by the approximation $\cosh(x) \approx |\sinh(x)|$, we get that $Q = |Y_0| + |Y_1| + \ldots + |Y_{\log T}|$. Letting $a_0 = Y_0$ and $X_0 = 1$, the question then becomes—what is the smallest β such that the following holds:

$$\mathbb{E}\left|\sum_{i=0}^{\log T} a_i X_i\right| \;\geq\; \frac{1}{\beta} \cdot \mathbb{E}\left[\sum_{i=0}^{\log T} |a_i| X_i^2\right] \;=\; \frac{1}{\beta} \cdot \mathbb{E}\left[\sum_{i=0}^{\log T} |a_i|\right].$$

For martingales, a statement similar to Khintchine's inequality is implied by the well-known Burkholder-Davis-Gundy (BDG) inequality (see Theorem 7.10.1 in Appendix 7.10):

$$\mathbb{E}\left[\max_{t \leq \log T} \left|\sum_{i=0}^{t} a_i X_i\right|\right] \;\geq\; c \cdot \mathbb{E}\left[\left(\sum_{i=0}^{\log T} a_i^2\right)^{1/2}\right]$$

for a positive constant c. One can also prove (see Lemma 7.10.2 in Appendix 7.10) that

$$(1 + \log T) \cdot \mathbb{E}\left|\sum_{i=0}^{\log T} a_i X_i\right| \;\geq\; \mathbb{E}\left[\max_{t \leq \log T} \left|\sum_{i=0}^{t} a_i X_i\right|\right].$$

Then, similar to the analysis for independent Rademacher random variables, using Cauchy-Schwarz,

$$(1 + \log T) \cdot \mathbb{E}\left|\sum_{i=0}^{\log T} a_i X_i\right| \;\geq\; c \cdot \mathbb{E}\left[\left(\sum_{i=0}^{\log T} a_i^2\right)^{1/2}\right] \;\geq\; \frac{c}{\sqrt{\log T}} \cdot \mathbb{E}\left[\sum_{i=0}^{\log T} |a_i|\right].$$

So we can conclude that $\beta = \mathsf{polylog}(T)$, which gives a $\mathsf{polylog}(T)$ bound on interval discrep-

ancy.

How to extend this analysis to d-dimensional Tusńady's problem? The martingale analysis above strongly relied on the interval structure of the problem, which is not clear even for the two-dimensional Tusńady's problem. To answer this question, we take a much more general view of our online discrepancy problem.[5]

7.2.4 A More General View of Changing Basis

One can also view the above analysis of the interval discrepancy problem as a more general underlying principle—that of working with a different *basis*. For example, let us take a linear algebraic approach to interval discrepancy and consider it as a vector balancing problem in $\mathbb{R}^{\mathcal{D}}$, where $\mathcal{D} = \{I_{j,k} \mid 0 \le j \le \log T, 0 \le k < 2^j\}$ is the set of all dyadic intervals. When a new point $x \in [0,1]$ arrives, the coordinate $I \in \mathcal{D}$ of the update vector v_t is given by

$$v_t(I) = \mathbf{1}_I(x).$$

Note that the update v_t lives in a T-dimensional subspace $\mathcal{V}$ of the $(2T-1)$-dimensional space $\mathbb{R}^{\mathcal{D}}$ since the T-intervals, $I_{\log T,k}$, at the bottom layer determine the rest of the coordinates.

The original potential function Φ from §7.2.1 corresponded to working with the original basis, but with the potential function Ξ from §7.2.3, our approach consisted of bounding the ℓ_∞-discrepancy in a different basis of the subspace $\mathcal{V}$. In general, we may choose any basis and then define a potential function as the sum of hyperbolic-cosines of the coordinates. To choose the right basis, we need several properties from it, but most importantly we need uncorrelation.

Uncorrelation and anti-concentration via the Eigenbasis. Recall that we say random variables X, Y are *uncorrelated* if $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, which is a condition only on the expected values of the random variables. Using Theorem 7.1.5, to show anti-concentration

[5] The more general view in fact gives a (slightly) better bound for interval discrepancy than the martingale based argument above. However, we include this martingale argument here, as it is insightful and could be useful for other problems.

it suffices that the coordinates in the *new basis* are mean-zero and uncorrelated, i.e.,

$$\mathbb{E}_v[v(i)v(j)] = 0$$

for distinct coordinates i, j.

For our vector balancing results under arbitrary distributions in Theorem 7.1.4, we work in an *eigenbasis* of the covariance matrix. As will be shown in the proof later, standard results from linear algebra imply that the coordinates are uncorrelated in any eigenbasis. Our next lemma uses this anti-concentration (along with the hyperbolic cosine potential) to bound discrepancy in the *new basis* in terms of *sparsity*—number of non-zero coordinates—of the incoming vectors.

Lemma 7.2.1. (Bounded discrepancy) *Let* p *be a distribution supported over* s*-sparse vectors in* $[-1,1]^n$ *satisfying* $\mathbb{E}_{v\sim\mathsf{p}}[v(i)v(j)] = 0$ *for all* $i \neq j \in [n]$*. Then for vectors* $v_1, \ldots, v_T$ *sampled i.i.d. from* p*, there is an online algorithm that maintains* $O(s(\log n + \log T))$ *discrepancy with high probability.*

Even though this lemma implies low discrepancy in the new basis, we need to be careful in bounding discrepancy in the original basis.

Sparsity and going back to the original basis. As discussed briefly in §4.3, although working in an eigenbasis allows us to obtain polynomial bounds for vector balancing, this is apriori not sufficient for our polylogarithmic geometric discrepancy bounds. There are two main challenges—firstly, working in a new basis might lose any sparsity that we might have in the original basis; e.g., in the one-dimensional interval discrepancy problem the arriving vectors are $(\log T)$-sparse (dyadic intervals) in the original basis, but could be $\Omega(T)$-sparse in the new basis; and secondly, even if one can find a new basis where the coordinates are uncorrelated and have low sparsity, Lemma 7.2.1 only implies low ℓ_∞-discrepancy in the new basis. So going back to the original basis might lose us a factor $\sqrt{n}$ more (we can only claim ℓ_2-discrepancy is the same). Recall, when we view interval discrepancy as vector balancing,

$n = \Theta(T)$, so we cannot afford losing $\sqrt{n}$. Fortunately, there is a special basis consisting of *Haar wavelets* that allows us to prove $\mathsf{polylog}(T)$ geometric discrepancy bounds.

7.2.5 Haar Wavelets: Polylogarithmic Geometric Discrepancy

There is a natural orthogonal basis associated with the unit interval—the basis of Haar wavelet functions. These consist of the functions $\Psi_{j,k}$'s shown in Figure 7.2. Together these functions are known to form an orthogonal basis for functions on the unit interval with bounded L_2-norm.

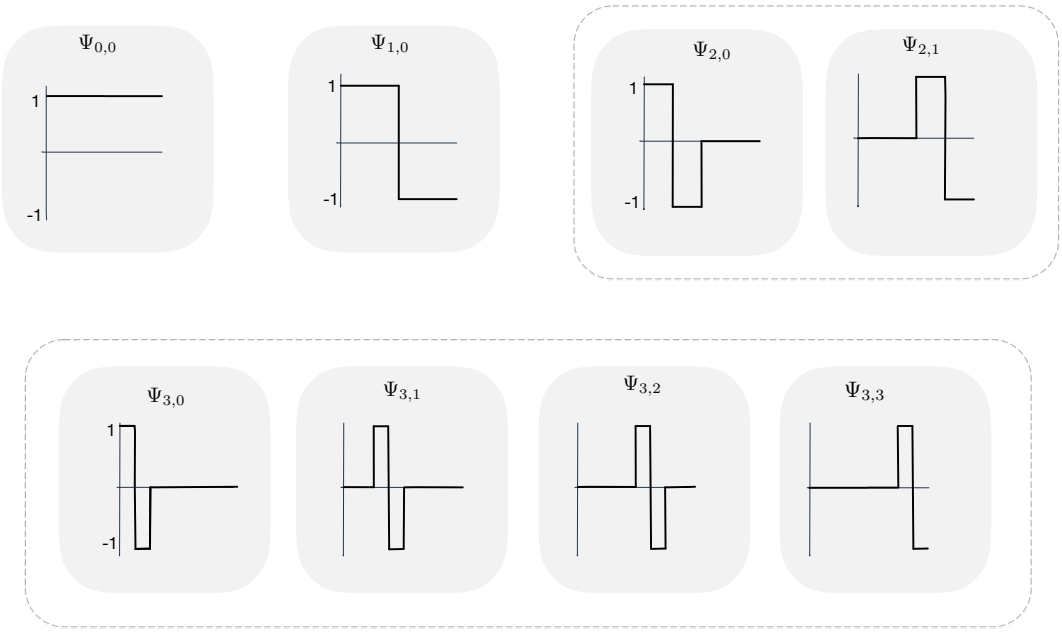


Figure 7.2: Haar wavelets in one dimension

Associated with the one-dimensional Haar wavelets is a natural martingale, which is the same martingale that our previous analysis in §7.2.3 relied on (e.g., $X_{j,k} = \Psi_{j+1,k}$ in the notation of §7.2.3.). It turns out that the Haar wavelets have nice orthogonality and sparsity properties that allow us to use Lemma 7.2.1—in particular, $\mathbb{E}_x[h(x)h'(x)] = 0$ for distinct Haar wavelet functions $h \neq h'$ and x sampled uniformly from $[0, 1]$. Moreover, moving from

the basis of Haar wavelets to the original basis does not incur any additional loss in the discrepancy bound, since for any dyadic interval I, one can show that its discrepancy

$$|d_t(I)| \quad \leq \quad \alpha \|\widehat{\mathbf{1}}_I\|_1,$$

where α is a bound on the discrepancy in the Haar basis and $\|\widehat{\mathbf{1}}_I\|_1$ is the ℓ_1-norm of the function $\mathbf{1}_I$ in the Haar basis. We prove that this ℓ_1-norm is one, so $|d_t(I)| \leq \alpha$. This gives a more direct proof of the $\mathsf{polylog}(T)$ interval discrepancy bound and also extends easily to the d-dimensional interval discrepancy problem.

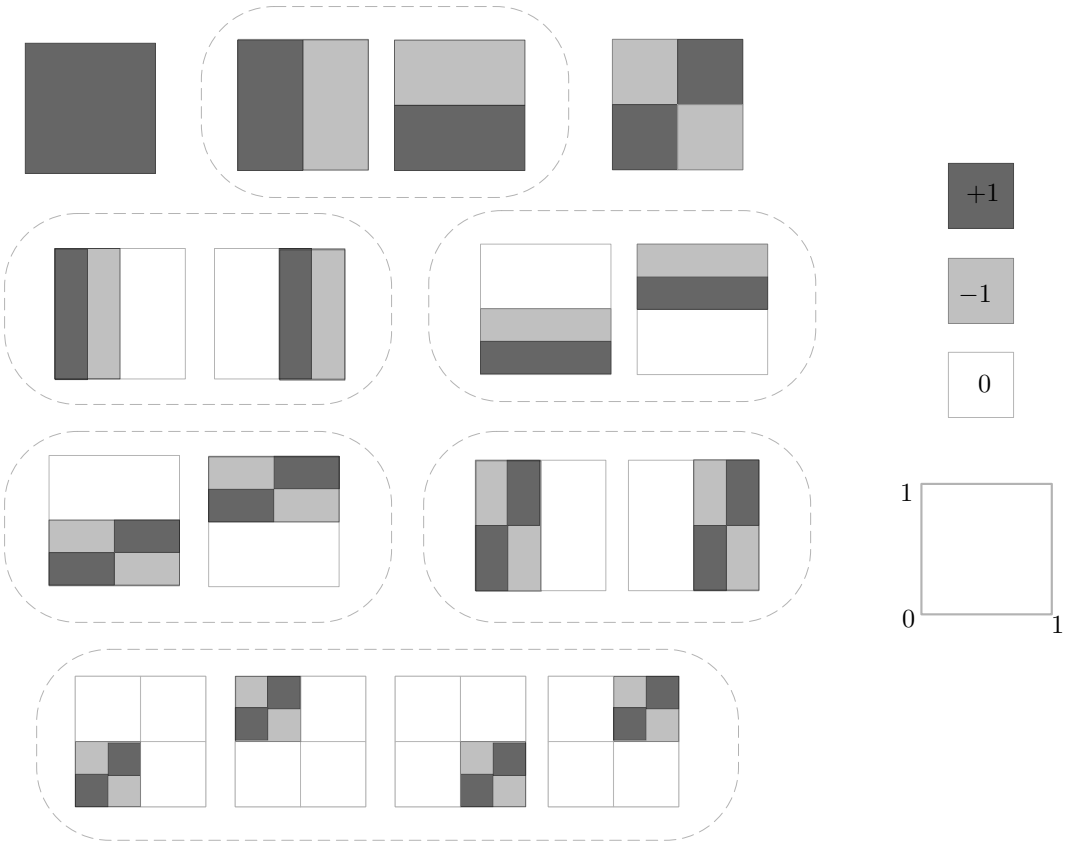


Figure 7.3: Haar wavelets in two dimensions

Tusńady's problem. Given the above framework of working in the Haar basis, our ex-

tension to the d-dimensional Tusńady's problem now naturally follows. For example, in two dimensions, we work with the basis of Haar wavelet functions which is formed by a taking tensor product $\Psi_{j,k} \times \Psi_{j',k'}$ of the one dimensional wavelets (see Figure 7.3). These functions form an orthogonal basis for all bounded *product* functions over $[0,1]^2$ and have nice sparsity properties. Moreover, we prove that for any axis-parallel box, the ℓ_1-norm of the Haar basis coefficients is one, so we do not lose any additional factor in the discrepancy bound while moving from the Haar basis to the original basis. This gives a polylogarithmic bound for two-dimensional Tusńady's problem, and also extends easily to higher dimensions.

7.2.6 Notations

All logarithms in this chapter will be base two. For any integer k, throughout the chapter $[k]$ will denote the set $\{1, \ldots, k\}$. For a vector $u \in \mathbb{R}^d$, we use $u(i)$ to denote the i^{th} coordinate of u for $i \in [d]$. Given another vector $v \in \mathbb{R}^d$, the notation $u \leq v$ denotes that $u(i) \leq v(i)$ for each $i \in [d]$. The all ones vector is denoted by $\mathbf{1}$. Given a distribution p, we use the notation $x \sim \mathsf{p}$ to denote an element x sampled from the distribution p. For a real function f, we will write $\mathbb{E}_{x \sim \mathsf{p}}[f(x)]$ to denote the expected value of $f(x)$ under x sampled from p. If the distribution is clear from the context, then we will abbreviate the above as $\mathbb{E}_x[f(x)]$.

7.3 Anti-Concentration Estimates

In this section we prove the anti-concentration results: we first prove it for uncorrelated random variables, and then give an improved bound for pairwise independent random variables. Although in the rest of this chapter we only use the weaker bound for uncorrelated random variables, we think the improved anti-concentration for pairwise independent random variables is of independent interest and will find applications in the future.

7.3.1 Pairwise Uncorrelated Random Variables

The following anti-concentration bound will be used in our discrepancy applications.

Theorem 7.1.5. (Uncorrelated anti-concentration) *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$, let $X_1, \ldots, X_n$ be uncorrelated random variables that are bounded $|X_i| \leq c$, satisfy $\mathbb{E}[X_i X_j] = 0$ for all $i \neq j$, and have sparsity s (the number of non-zero X_i's in any outcome). Then*

$$\mathbb{E}\Big|\sum_i a_i X_i\Big| \;\geq\; \mathbb{E}\Big[\sum_i |a_i| X_i^2\Big] \cdot \frac{1}{cs}. \tag{7.2}$$

Moreover, this bound is tight, even for pairwise independent random variables.

Note that if we have pairwise uncorrelated mean-zero random variables $X_1, \ldots, X_n$, then we get $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i] \cdot \mathbb{E}[X_j] = 0$, so the above lemma implies anti-concentration in this case. The bound in the above lemma is tight because of the Hadamard example described previously in §7.1.3.

The following is the main claim in the proof of Theorem 7.1.5. Roughly it says that $\mathbb{E}\Big|\sum_i a_i X_i\Big| \geq \frac{1}{c} \cdot \max_{k \in [n]} \mathbb{E}[|a_k| X_k^2]$. Combined with the observation that $\max_{k \in [n]} \mathbb{E}[|a_k| X_k^2] \geq \frac{1}{n} \cdot \mathbb{E}\left[\sum_k |a_k| X_k^2\right]$ this implies Theorem 7.1.5 when sparsity $s = n$. However, to get inequality (7.2) in terms of sparsity s, the statement of the claim has to be more refined.

Claim 7.3.1. *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and random variables $X_1, \ldots, X_n$ satisfying $|X_i| \leq c$ and $\mathbb{E}[X_i X_j] = 0$ for distinct i, j, the following holds for any $k \in [n]$,*

$$\mathbb{E}\left[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\right] \;\geq\; \frac{1}{c} \cdot \mathbb{E}[|a_k| X_k^2].$$

Proof. Using that $|X_k| \leq c$, we have

$$\begin{aligned} c \cdot \mathbb{E}\left[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\right] &\geq \mathbb{E}\left[\Big|\sum_i a_i X_i\Big| \cdot |X_k|\right] \\ &= \mathbb{E}\left[\Big|a_k X_k^2 + \sum_{i \neq k} a_i X_i X_k\Big|\right] \;\geq\; \mathbb{E}\left[\mathsf{sign}(a_k)\Big(a_k X_k^2 + \sum_{i \neq k} a_i X_i X_k\Big)\right]. \end{aligned}$$

Since $\mathbb{E}[X_i X_k] = 0$ for $i \neq k$, it follows that

$$c \cdot \mathbb{E}\left[\left|\sum_i a_i X_i\right| \cdot 1_{X_k \neq 0}\right] \geq \mathbb{E}\left[|a_k| X_k^2\right] + \sum_{i \neq k} a_i \cdot \mathsf{sign}(a_k) \cdot \mathbb{E}\left[X_i X_k\right]$$
$$= \mathbb{E}\left[|a_k| X_k^2\right].$$

□

When combined with the following easy claim, this will prove Theorem 7.1.5.

Claim 7.3.2. *Let $Y_1, \ldots, Y_n$ be correlated random variables such that for any outcome at most s of them are non-zero. Moreover, suppose there is a random variable L which satisfies*

$$\mathbb{E}\Big[|L| \cdot 1_{Y_k \neq 0}\Big] \geq \mathbb{E}\Big[|Y_k|\Big] \quad \textit{for all } k \in [n].$$

Then, $\mathbb{E}[|L|] \geq \frac{1}{s} \sum_k \mathbb{E}[|Y_k|]$.

Proof. Sum the given inequality for all $k \in [n]$ to get

$$\sum_k \mathbb{E}\Big[|Y_k|\Big] \quad \leq \quad \sum_k \mathbb{E}\Big[|L| \cdot 1_{Y_k \neq 0}\Big] \quad = \quad \mathbb{E}\Big[|L| \cdot \sum_k 1_{Y_k \neq 0}\Big] \quad \leq \quad \mathbb{E}\Big[|L| \cdot s\Big].$$

□

Proof of Theorem 7.1.5. Applying Claim 7.3.1 and Claim 7.3.2 (with $L = \sum_i a_i X_i$ and $Y_i = \frac{1}{c} \cdot |a_i| X_i^2$), we get that

$$\mathbb{E}\left[\left|\sum_i a_i X_i\right|\right] \quad \geq \quad \mathbb{E}\left[\sum_k |a_k| X_k^2\right] \cdot \frac{1}{cs}.$$

□

7.3.2 Pairwise Independent Random Variables

In the special case of pairwise independent random variables, it is possible to obtain an improved inequality over Theorem 7.1.5.

Theorem 7.1.6. (Pairwise independent anti-concentration) *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$, let $X_1, \ldots, X_n$ be mean-zero pairwise independent random variables with sparsity $s \leq n$. Then*

$$\mathbb{E}\Big[\Big|\sum_i a_i X_i\Big|\Big] \;\geq\; \mathbb{E}\Big[\sum_i |a_i X_i|\Big] \cdot \frac{1}{s}. \tag{7.3}$$

Notice, (7.3) immediately implies (7.2) for mean-zero pairwise independent random variables with $|X_i| \leq c$. One cannot hope to prove the stronger statement (7.3) for uncorrelated random variables due to the following example.

Example. Let $0 < \delta \ll 1$. Suppose X_1, X_2 are real random variables distributed over four outcomes:

$$(X_1, X_2) = \begin{cases} \left(\frac{1}{\delta}, \frac{1}{\delta}\right) \text{ or } \left(-\frac{1}{\delta}, -\frac{1}{\delta}\right) & \text{w.p. } \frac{\delta^2}{2(1+\delta^2)} \text{ each,} \\ (1,-1) \text{ or } (-1,1) & \text{w.p. } \frac{1}{2} - \frac{\delta^2}{2(1+\delta^2)} \text{ each.} \end{cases}$$

Here X_1 and X_2 are uncorrelated because

$$\mathbb{E}[X_1 X_2] \;=\; \frac{1}{\delta^2} \cdot \frac{\delta^2}{1+\delta^2} - 1 \cdot \left(1 - \frac{\delta^2}{1+\delta^2}\right) \;=\; 0.$$

Now it is easy to verify that X_1 and X_2 are mean zero, and

$$\mathbb{E}[|X_1 + X_2|] = \frac{2\delta}{1+\delta^2} \quad \text{and} \quad \mathbb{E}[|X_1| + |X_2|] = \frac{2+2\delta}{1+\delta^2}.$$

Therefore, the ratio between the two expectations can be made arbitrarily bad by making $\delta \to 0$.

Next, we prove Theorem 7.1.6. We start with the following claim.

Claim 7.3.3. *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and mean-zero pairwise independent random variables $X_1, \ldots, X_n$, the following holds for any $k \in [n]$,*

$$\mathbb{E}\Big[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\Big] \;\geq\; \mathbb{E}[|a_k X_k|].$$

Proof. We have

$$\begin{aligned}
\mathbb{E}\Big[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\Big] &= \mathbb{E}\Big[\Big|a_k X_k + \sum_{i \neq k} a_i X_i \cdot 1_{X_k \neq 0}\Big|\Big] \\
&\geq \mathbb{E}\Big[\mathsf{sign}(a_k X_k)\Big(a_k X_k + \sum_{i \neq k} a_i X_i \cdot 1_{X_k \neq 0}\Big)\Big] \\
&= \mathbb{E}\Big[|a_k X_k| + \mathsf{sign}(a_k X_k) \sum_{i \neq k} a_i X_i \cdot 1_{X_k \neq 0}\Big].
\end{aligned}$$

Since X_i and X_k are mean-zero and pairwise independent for $i \neq k$, we have $\mathbb{E}[X_i f(X_k)] = \mathbb{E}[X_i] \cdot \mathbb{E}[f(X_k)] = 0$ for any function f. Therefore,

$$\mathbb{E}\Big[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\Big] \;\geq\; \mathbb{E}[|a_k X_k|] + \sum_{i \neq k} \mathbb{E}\Big[\mathsf{sign}(a_k X_k) \cdot a_i X_i \cdot 1_{X_k \neq 0}\Big] \;=\; \mathbb{E}[|a_k X_k|].$$

□

Proof of Theorem 7.1.6. Combining Claim 7.3.3 with Claim 7.3.2 completes the proof of Theorem 7.1.6. □

7.4 Online Discrepancy under Uncorrelated Arrivals

In this section we consider the vector balancing problem in the special case when the input distribution has uncorrelated coordinates. All our upper and lower bounds will then follow from choosing a suitable basis to reduce the original problem to a basis with uncorrelated coordinates.

7.4.1 Upper Bounds

We say a vector in $\mathbb{R}^d$ is *s-sparse* if it has at most s non-zero coordinates. The following lemma bounds the discrepancy for uncorrelated sparse distributions.

Lemma 7.2.1. (Bounded discrepancy) *Let* p *be a distribution supported over s-sparse vectors in* $[-1,1]^n$ *satisfying* $\mathbb{E}_{v\sim\mathsf{p}}[v(i)v(j)] = 0$ *for all* $i \neq j \in [n]$. *Then for vectors* $v_1, \ldots, v_T$ *sampled i.i.d. from* p, *there is an online algorithm that maintains* $O(s(\log n + \log T))$ *discrepancy with high probability.*

Proof of Lemma 7.2.1. Our algorithm will use the same potential function approach described in §7.2, and uses our anti-concentration lemma from §7.3 to argue that the potential always remains polynomially bounded.

Algorithm. At any time step t, let $d_t = \chi_1 v_1 + \ldots + \chi_t v_t$ denote the current discrepancy vector after the signs $\chi_1, \ldots, \chi_t \in \{\pm 1\}$ have been chosen. Set $\lambda = \frac{1}{2s}$ and define the potential function

$$\Phi_t := \sum_{i\in[n]} \cosh(\lambda d_t(i)).$$

When the vector v_t arrives, the algorithm chooses the sign χ_t that minimizes the increase $\Phi_t - \Phi_{t-1}$.

Bounded Positive Drift. Let us fix a time t. To simplify the notation, let $\Delta\Phi = \Phi_t - \Phi_{t-1}$, let $d = d_{t-1}$, and let $v = v_t$.

After choosing the sign χ_t, the discrepancy vector $d_t = d + \chi_t v$. To bound the change $\Delta\Phi$, since $\cosh'(x) = \sinh(x)$ and $\sinh'(x) = \cosh(x)$, using Taylor expansion

$$\begin{aligned}\Delta\Phi &= \sum_i \Big(\lambda \sinh(\lambda d(i)) \cdot (\chi_t v(i)) + \frac{\lambda^2}{2!}\cosh(\lambda d(i)) \cdot (\chi_t v(i))^2 + \cdots\Big),\\ &\leq \sum_i \Big(\lambda \sinh(\lambda d(i)) \cdot (\chi_t v(i)) + \lambda^2 \cosh(\lambda d(i)) \cdot (\chi_t v(i))^2\Big),\end{aligned}$$

where the last inequality follows since $|\sinh(x)| \leq \cosh(x)$ for all $x \in \mathbb{R}$, and since $|\chi_t v(i)| \leq 1$ and $\lambda < 1$, the higher order terms in the Taylor expansion are dominated by the first and second order terms.

Set $L = \sum_i \sinh(\lambda d(i))v(i)$, and $Q^* = \sum_i \cosh(\lambda d(i))v(i)^2$, and $Q = \sum_i |\sinh(\lambda d(i))|v(i)^2$. Since $\cosh(x) \leq |\sinh(x)| + 1$ for $x \in \mathbb{R}$ and $|v(i)| \leq 1$, we have $Q^* \leq Q + n$. Therefore,

$$\Delta\Phi \leq \chi_t \cdot \lambda \cdot L + \lambda^2 \cdot Q + \lambda^2 n.$$

Since, the algorithm chooses χ_t to minimize the increase in the potential:

$$\Delta\Phi \leq -\lambda \cdot |L| + \lambda^2 \cdot Q + \lambda^2 n.$$

Now, since $\mathbb{E}_v[v(i)v(j)] = 0$ for all $i, j \in [n]$, we can apply Theorem 7.1.5 with $X_i = v(i)$ and $a_i = \sinh(\lambda d(i))$ to get that $\mathbb{E}_v[|L|] \geq \frac{1}{s} \cdot \mathbb{E}[Q] = 2\lambda \cdot \mathbb{E}[Q]$, which yields that

$$\mathbb{E}_v[\Delta\Phi] \quad \leq \quad -\lambda \cdot \mathbb{E}_v[|L|] + \lambda^2 \cdot \mathbb{E}_v[Q] + \lambda^2 n \quad \leq \quad -\lambda^2 \cdot \mathbb{E}_v[Q] + \lambda^2 n \quad \leq \quad n.$$

Discrepancy Bound. The above implies that for any time $t \in [T]$, the expectation $\mathbb{E}[\Phi_t] \leq nT$. By Markov's inequality and a union bound over the T time steps, with probability at least $1 - T^{-2}$, the potential $\Phi_t \leq nT^4$ for every time $t \in [T]$. Since at any time t, we have $\cosh(\lambda \|d_t\|_\infty) \leq \Phi_t$, this implies that with probability at least $1 - T^{-2}$, the discrepancy at every time is

$$O\left(\frac{\log(nT^4)}{\lambda}\right) = O(s(\log n + \log T)),$$

which finishes the proof of Lemma 7.2.1. □

7.4.2 Lower Bounds

We now show that the dependence on s and $\log T$ in Lemma 7.2.1, cannot be improved up to polynomial factors. In particular, a lower bound of $\Omega(s^{1/2})$, even when the time horizon

is $T = n$, follows directly from the following more general statement for the vector balancing problem under distributions with uncorrelated coordinates. This general version will later also imply our lower bounds for geometric discrepancy.

Lemma 7.4.1. *Let* p *be a distribution supported over vectors in* $[-1,1]^n$ *with* ℓ_2*-norm* k*, such that for every* $i \neq j \in [n]$ *we have* $\mathbb{E}_{v\sim\mathsf{p}}[v(i)v(j)] = 0$*. Then, for any online algorithm that receives as input vectors* $v_1, \ldots, v_n$ *sampled i.i.d. from* p*, with probability at least* $3/4$*, the discrepancy is* $\Omega(k)$ *at some time* $t \in [n]$*.*

We remark that the above lower bound may not hold if the algorithms are offline.

Proof of Lemma 7.4.1. Since the distribution p over inputs is fixed, we may assume that the algorithm is deterministic. Let $d_t = \chi_1 v_1 + \ldots + \chi_t v_t$ denote the discrepancy vector at any time $t \in [n]$. Consider the quadratic potential function:

$$\Phi_t \quad := \quad \|d_t\|_2^2 \quad = \quad \sum_{i\in[n]} |d_t(i)|^2.$$

We will need the following claim that shows Φ_t increases in expectation for any online algorithm. Let us define $\Delta\Phi_t = \Phi_t - \Phi_{t-1}$.

Claim 7.4.2. *Conditioned on any* $v_1, \ldots, v_{t-1}$ *and signs* $\chi_1, \ldots, \chi_{t-1}$ *such that* $\|d_{t-1}\|_\infty \leq \frac{k}{4}$*, we have*

$$\mathbb{E}_{v_t}[\Delta\Phi_t] \geq k^2/2 \tag{7.5}$$

where the expectation is taken only over the update $v_t \sim \mathsf{p}$*.*

Proof. Set $\Delta\Phi = \Delta\Phi_t$, vector $v = v_t$, and $d = d_{t-1}$. When the update v arrives, note that

$d_t = d + \chi_t v$. Therefore, the increase in the potential is given by

$$\Delta\Phi \;=\; \sum_{i=1}^{n}\Big(2d(i)\cdot\chi_t v(i) + \big(\chi_t v(i)\big)^2\Big) \;=\; 2\chi_t\Big(\sum_{i=1}^{n} d(i)v(i)\Big) + \|v\|_2^2 \;=\; 2L + k^2, \tag{7.6}$$

where $L = \chi_t\left(\sum_{i=1}^{n} d(i)v(i)\right)$.

To bound the expected value of L, we use Jensen's inequality and $\mathbb{E}_v[v(i)v(j)] = 0$ for $i \neq j$ to get:

$$\begin{aligned}(\mathbb{E}_v[L])^2 \;\leq\; \mathbb{E}_v[L^2] \;&=\; \sum_{i=1}^{n}|d(i)|^2\cdot\mathbb{E}_v[v(i)^2] + \sum_{i\neq j} d(i)d(j)\cdot\mathbb{E}_v[v(i)v(j)]\\ &=\; \sum_{i=1}^{n}|d(i)|^2\cdot\mathbb{E}_v[v(i)^2] \;\leq\; \|d\|_\infty^2\cdot\sum_{i=1}^{n}\mathbb{E}_v[v(i)^2] \;=\; \|d\|_\infty^2\,k^2 \;\leq\; \frac{k^4}{16}.\end{aligned}$$

Therefore, plugging the above in (7.6), we get

$$\mathbb{E}_v[\Delta\Phi] \;\geq\; -2\cdot|\mathbb{E}_v[L]| + k^2 \;\geq\; -2\cdot\Big(\frac{k^4}{16}\Big)^{1/2} + k^2 \;\geq\; \frac{k^2}{2}.$$

□

To prove Lemma 7.4.1 using the last claim, we define τ to be the first time that $\|d_\tau\|_\infty > k/4$ if such a τ exists, or $\tau = n$ otherwise. Let us define a new potential Φ_t^* which remains the same as Φ_t for $t \leq \tau$ and increases by $k^2/2$ deterministically for every $t > \tau$.

Note that for all possible random choices,

$$\Phi_n^* \;\leq\; \Phi_{\tau-1} + \frac{nk^2}{2} \;\leq\; \frac{nk^2}{16} + \frac{nk^2}{2},$$

where the second inequality holds since $\|d_{\tau-1}\|_\infty \leq k/4$ and therefore, $\Phi_{\tau-1} \leq \frac{1}{16}\cdot nk^2$.

Moreover, let $\mathcal{E}$ be the event that $\|d_t\|_\infty \leq k/4$ for every $t \leq n$. Note that when $\mathcal{E}$ occurs

then the final potential $\Phi_n^* \leq \frac{1}{16} \cdot nk^2$. Defining $p = \mathbb{P}[\mathcal{E}]$, we have

$$\mathbb{E}[\Phi_n^*] \;\leq\; p \cdot \frac{nk^2}{16} + (1-p)\left(\frac{nk^2}{16} + \frac{nk^2}{2}\right) \;=\; \frac{nk^2}{16} + (1-p)\frac{nk^2}{2}. \tag{7.7}$$

Moreover, from Claim 7.4.2 and the definition of Φ_t^*, it follows that $\mathbb{E}[\Phi_n^*] \geq \frac{1}{2} \cdot nk^2$. Comparing this with (7.7) yields that $p \leq 1/8$. Hence, with probability at least $7/8$, the discrepancy must be $k/4$ at some point. □

Dependence on T. We next show that the discrepancy must be $\Omega((\log T/\log\log T)^{1/2})$ with high probability even when $n = O(1)$ (we assume $n \geq 2$ throughout this discussion). We only sketch the proof here as the arguments are standard. The idea is that for large T, there is a high probability of getting a long enough run of consecutive vectors with each v_t almost orthogonal to d_{t-1}.

Let p be the uniform distribution[6] over vectors on the unit sphere S^{n-1}. For any vector $u \in \mathbb{R}^n$, and v sampled from p, there is a universal constant c so that for all $\delta \leq 1$, we have $\mathbb{P}[|\langle u, v\rangle| \leq \delta\|u\|_2/n^{1/2}] \geq c\delta$.

Let $\beta \geq 1$ be some parameter that we optimize later. Setting $\delta = 1/(4\beta)$ gives that whenever $\|d_{t-1}\|_2 \leq \beta n^{1/2}$, there is at least $c/(4\beta)$ probability that $|\langle d_{t-1}, v_t\rangle| \leq 1/4$, and hence irrespective of the sign χ_t,

$$\|d_t\|_2^2 \;\;\geq\;\; \|d_{t-1}\|_2^2 - 2|\langle d_{t-1}, v_t\rangle| + \|v_t\|_2^2 \;\;\geq\;\; \|d_{t-1}\|_2^2 + 1/2.$$

So for any τ consecutive steps, with at least $(c/4\beta)^\tau$ probability, this happens at every step (or the ℓ_2-discrepancy already exceeds $\beta n^{1/2}$ at some step), and hence the discrepancy has ℓ_2-norm at least $\Omega(\tau^{1/2})$.

Partitioning the time horizon T into T/τ disjoint blocks, and setting $\beta = \log(T)$, and

[6] Our argument works for a wide class of distributions p, as long as for any $d_{t-1} \in \mathbb{R}^n$, the random incoming vector v_t sampled from p has a non-trivial probability of having a small inner product with d_{t-1}. We only give the argument for the uniform distribution on the unit sphere for simplicity.

$\tau = \Omega(\log T / \log\log T)$, the probability such a run does not occur in any block is at most $(1-(c/4\beta)^\tau)^{(T/\tau)} = T^{-\Omega(1)}$ by our choice of the parameters. This gives the claimed lower bound.

7.5 Online Vector Balancing: Polynomial Bounds

In this section, we prove our vector balancing result for arbitrary distributions.

Theorem 7.1.4. (Vector Balancing Under Dependencies) *For any sequence of vectors $v_1, \ldots, v_T \in [-1,1]^n$ sampled i.i.d. from some arbitrary distribution p, there is an online algorithm which selects signs $\chi_t \in \{\pm 1\}$ such that, with high probability, we have*

$$\max_{t\in[T]} \left\| \chi_1 v_1 + \ldots + \chi_t v_t \right\|_\infty = O(n^2(\log T + \log n)).$$

Proof of Theorem 7.1.4. Without loss of generality, we may assume that the distribution p is *symmetric*, i.e. both v and $-v$ have the same probability density, since we can always multiply the incoming vector v with a Rademacher ± 1 random variable without changing the problem. Let $P \in \mathbb{R}^{d\times d}$ denote the covariance matrix of our input distribution, and since p is symmetric, we get $P = \mathbb{E}_{v\sim\mathsf{p}}[vv^T]$. Let U denote the orthogonal matrix whose columns $u_1, \ldots, u_n$ form an eigenbasis for P. Note that in terms of its spectral decomposition, $P = \sum_{k=1}^n \lambda_k u_k u_k^T$ for $\lambda_k \in \mathbb{R}$.

To prove our discrepancy bound, instead of working in the original basis, we will view our problem as a vector balancing problem in the basis given by the columns of U. Now the update sequence is given by $w_1, \ldots, w_T$ where $w_t = \frac{1}{\sqrt{n}} \cdot U^T v$ is the normalized update vector in the basis U.

Since $\|v\|_2 \le \sqrt{n}$ and orthogonal matrices preserve ℓ_2-norm, we have $\|U^T v\|_2 = \|v\|_2 \le \sqrt{n}$. It follows that for any t, we have $\|w_t\|_\infty \le \|w_t\|_2 = \frac{1}{\sqrt{n}} \cdot \|U^T v\|_2 \le 1$. Furthermore, any two coordinates of the update vectors w_t's are uncorrelated, i.e., for any $i \ne j \in [n]$ we have

$$\mathbb{E}[w_t(i)\cdot w_t(j)] = \frac{1}{n}\mathbb{E}[\langle u_i, v\rangle\langle u_j, v\rangle] = \frac{1}{n}\mathbb{E}[u_i^T v v^T u_j] = \frac{1}{n} u_i^T P u_j = 0,$$

where the last equality holds since $P = \sum_{k=1}^{n} \lambda_k u_k^T u_k$.

Thus, we can use the online algorithm from Lemma 7.2.1 to select signs $\chi_1, \ldots, \chi_T \in \{\pm 1\}$. Let $d_t = \chi_1 v_1 + \ldots + \chi_t v_t$ denote the discrepancy in the original basis. Now using the trivial bound of $s \leq n$ on sparsity in Lemma 7.2.1, we get that with high probability,

$$\frac{1}{\sqrt{n}} \left\| U^T d_t \right\|_\infty \;=\; O(n(\log n + \log T)).$$

Again, using that orthogonal matrices preserve ℓ_2-norm,

$$\|d_t\|_\infty \;\leq\; \|d_t\|_2 \;=\; \|U^T d_t\|_2 \;\leq\; \sqrt{n} \cdot \left\| U^T d_t \right\|_\infty \;=\; O(n^2(\log n + \log T)).$$

□

7.6 Online Geometric Discrepancy: Polylogarithmic Bounds

In this section, we will prove our results on geometric discrepancy problems. For this, we will need a special basis of orthogonal functions on the unit interval called the Haar system. We briefly review its properties.

7.6.1 Preliminaries: Haar System

Let $\Psi : \mathbb{R} \to \mathbb{R}$ denote the *mother wavelet* function

$$\Psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The *unnormalized* Haar wavelet functions (recall Figure 7.2) are defined as follows: let $\Psi_{0,0}(x) = 1$ for all $x \in \mathbb{R}$, and for any $j \in \mathbb{N}^*$ and $0 \leq k < 2^{j-1}$ define

$$\Psi_{j,k}(x) := \Psi(2^{j-1}x - k).$$

We call j as the *scale* and k as the *shift* of the wavelet.

The Haar wavelet functions have nice orthogonality properties. In particular, let x be drawn uniformly from the unit interval $[0,1]$. Then, one can easily check that

$$\begin{aligned} &\mathbb{E}_x[\Psi_{j,k}(x)^2] = 2^{-(j-1)} && \text{for } j > 0, \\ &\mathbb{E}_x[\Psi_{j,k}(x)] = 0 && \text{for } j > 0, \\ &\mathbb{E}_x[\Psi_{j,k}(x)\Psi_{j',k'}(x)] = 0 \quad && \text{unless } j = j' \text{ and } k = k'. \end{aligned} \tag{7.8}$$

The Haar wavelet functions are not just orthogonal, but they form an orthogonal basis (not orthonormal), called the *Haar system*, for the class of functions on the unit interval with bounded L_2-norm. In particular, we have the following proposition where for $j \in \mathbb{Z}_{\geq 0}$ we denote $\mathcal{H}_j = \bigcup_{0 \leq k < 2^{j-1}} \{\Psi_{j,k}\}$ and let $\mathcal{H} = \bigcup_{j \geq 0} \mathcal{H}_j$.

Proposition 7.6.1 ([Wal04], Chapter 5)**.** *For any $f : [0,1] \to \mathbb{R}$ such that $\mathbb{E}_x[f(x)^2] < \infty$, we have*

$$f = \sum_{h \in \mathcal{H}} \widehat{f}(h) \cdot h(x)$$

where $\widehat{f}(h) = \frac{\mathbb{E}_x[f(x)h(x)]}{\mathbb{E}_x[h(x)^2]}$ is the corresponding coefficient in the Haar system basis for $h \in \mathcal{H}$.

Indeed, since the Haar system forms an orthogonal basis, we also have that

$$\mathbb{E}_x[f(x)^2] = \sum_{h \in \mathcal{H}} \widehat{f}(h)^2 \cdot \mathbb{E}_x[h(x)^2].$$

A simple corollary of Proposition 7.6.1 is that $\mathcal{H}^{\otimes d}$ is an orthogonal basis for the linear space spanned by all functions over the unit cube $[0,1]^d$ that have a product structure and bounded L_2-norm. In particular, let $\boldsymbol{h} = (h_1, \ldots, h_d)$ be an element of $\mathcal{H}^{\otimes d}$ which we will view as a function from $[0,1]^d \to \mathbb{R}$ by defining $\boldsymbol{h}(x) = \prod_{i=1}^d h_i(x(i))$ for $x \in [0,1]^d$. Note that distinct $\boldsymbol{h}$ and $\boldsymbol{h}'$ are orthogonal since for x drawn uniformly from $[0,1]^d$,

$$\mathbb{E}_x[\boldsymbol{h}(x)\boldsymbol{h}'(x)] \quad = \quad \prod_{i=1}^{d} \mathbb{E}_{x(i)}[h_i(x(i))h_i'(x(i))] \quad = \quad 0. \tag{7.9}$$

Moreover, any product function can be expressed by functions in $\mathcal{H}^{\otimes d}$ as given in the following proposition[7].

Proposition 7.6.2. *For any $f : [0,1]^d \to \mathbb{R}$ such that $f(x) = \prod_{i=1}^d f_i(x(i))$ for some $f_i : [0,1] \to \mathbb{R}$ satisfying $\mathbb{E}_{x(i)}[f_i(x(i))^2] < \infty$, we have that*

$$f = \sum_{\boldsymbol{h} \in \mathcal{H}^{\otimes d}} \widehat{f}(\boldsymbol{h})\boldsymbol{h},$$

where $\widehat{f}(\boldsymbol{h}) = \frac{\mathbb{E}_x[f(x)\boldsymbol{h}(x)]}{\mathbb{E}_x[\boldsymbol{h}(x)^2]}$.

Proof. Expressing each f_i in the Haar system basis using Proposition 7.6.1, we get the statement of the proposition by tensoring. □

Let $\mathcal{H}_{\leq j} = \bigcup_{j' \leq j} \mathcal{H}_{j'}$, and define $\mathcal{H}_{<j}, \mathcal{H}_{>j}, \mathcal{H}_{\geq j}$ analogously. Then, we have the following lemma about the Haar system decomposition of indicator functions of dyadic intervals.

Proposition 7.6.3. *Let $\mathbf{1}_{I_{\ell,m}}$ denote the indicator function for the interval $I_{\ell,m} = \left[m2^{-\ell}, (m+1)2^{-\ell}\right)$. Then,*

$$\sum_{h \in \mathcal{H}_0} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 2^{-\ell},$$
$$\sum_{h \in \mathcal{H}_j} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 2^{-(\ell+1-j)} \text{ for any } 1 \leq j \leq \ell \text{ and}$$
$$\widehat{\mathbf{1}}_{I_{\ell,m}}(h) = 0 \text{ for any } h \in \mathcal{H}_{>\ell}.$$

In particular, we have $\sum_{h \in \mathcal{H}} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = \sum_{h \in \mathcal{H}_{\leq \ell}} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 1$.

Proof. First, observe that for any $j > \ell$, either $\Psi_{j,k}(x) = 0$ identically on the interval $I_{\ell,m}$ or it takes $+1$ and -1 values on equal size sub-intervals of $I_{\ell,m}$, so that $\mathbb{E}_x[1_{I_{\ell,m}}(x)\Psi_{j,k}(x)] = 0$.

[7] More generally, Proposition 7.6.2 holds for any L_2-integrable function $f \in L_2([0,1]^d)$, as the linear span of product functions with domain $[0,1]^d$ is dense in $L_2([0,1]^d)$.

For $\Psi_{0,0}$, notice that $\mathbb{E}_x[1_{I_{\ell,m}}(x)\Psi_{0,0}(x)] = 2^{-\ell}$ and $\mathbb{E}_x[\Psi_{0,0}(x)^2] = 1$. Therefore, we have

$$\sum_{h\in\mathcal{H}_0} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 2^{-\ell}.$$

Now consider any $1 \le j \le \ell$. Then, there exists a unique $0 \le k^* < 2^{j-1}$ such that Ψ_{j,k^*} takes the constant value $+1$ or -1 identically on the interval $I_{\ell,m}$, and the function $\Psi_{j,k}$ is identically zero on the interval $I_{\ell,m}$ for any $k \neq k^*$. It follows that $\mathbb{E}_x[1_{I_{\ell,m}}(x)\Psi_{j,k^*}(x)] = \pm 2^{-\ell}$, $\mathbb{E}_x[\Psi_{j,k^*}(x)^2] = 2^{-(j-1)}$ and $\mathbb{E}_x[1_{I_{\ell,m}}(x)\Psi_{j,k}(x)] = 0$ for any $k \neq k^*$. Therefore, for $1 \le j \le \ell$, we have

$$\sum_{h\in\mathcal{H}_j} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 2^{-(\ell+1-j)}.$$

From the above, it also follows that

$$\sum_{h\in\mathcal{H}} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| \;=\; \sum_{h\in\mathcal{H}_{\le\ell}} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| \;=\; 2^{-\ell} + \sum_{j=1}^{\ell} 2^{-(\ell+1-j)} \;=\; 2^{-\ell} + (1-2^{-\ell}) \;=\; 1.$$

□

We also get a similar proposition about dyadic boxes. In particular, let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d)$ for non-negative integers ℓ_i's and let $\boldsymbol{m} = (m_1, \ldots, m_d)$ for integers $0 \le m_i < 2^{\ell_i}$. Let $\mathcal{H}_{\le\boldsymbol{\ell}}^{\otimes d} = \mathcal{H}_{\le\ell_1} \times \cdots \times \mathcal{H}_{\le\ell_d}$. Then, for the dyadic box

$$I_{\boldsymbol{\ell},\boldsymbol{m}} = I_{\ell_1,m_1} \times \cdots \times I_{\ell_d,m_d},$$

we have the following proposition. Below we write $\min\{\mathbf{e}, \mathbf{f}\}$ to denote the vector whose i^{th} coordinate is $\min\{\mathbf{e}(i), \mathbf{f}(i)\}$ for $\mathbf{e}, \mathbf{f} \in \mathbb{R}^d$.

Proposition 7.6.4. *Let $\mathbf{1}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}$ denote the indicator function for the dyadic box $I_{\boldsymbol{\ell},\boldsymbol{m}}$. Then,*

$$\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{j}}^{\otimes d}} |\widehat{\mathbf{1}}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}(\boldsymbol{h})| = 2^{-\|\min\{\boldsymbol{\ell},\boldsymbol{\ell}+\mathbf{1}-\boldsymbol{j}\}\|_1} \text{ for any } \boldsymbol{j} \le \boldsymbol{\ell} \text{ and}$$

$$\widehat{\mathbf{1}}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}(\boldsymbol{h}) = 0 \text{ for any } \boldsymbol{h} \notin \mathcal{H}_{\le\boldsymbol{\ell}}.$$

In particular, we have $\sum_{\boldsymbol{h}\in\mathcal{H}^{\otimes d}} |\widehat{\mathbf{1}}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}(\boldsymbol{h})| = \sum_{\boldsymbol{h}\in\mathcal{H}_{\le\boldsymbol{\ell}}^{\otimes d}} |\widehat{\mathbf{1}}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}(\boldsymbol{h})| = 1.$

The proof of the above proposition follows from Proposition 7.6.3 by tensoring.

7.6.2 Online Interval Discrepancy Problem

Now we prove Theorem 7.1.2 for the d-dimensional interval discrepancy problem. Let $\mathbf{x} = (x_1, \dots, x_T)$ be a sequence of points in $[0,1]^d$ and let $\chi \in \{\pm 1\}^T$ be a signing. For any interval $I \subseteq [0,1]$ and time $t \in [T]$, recall that the discrepancy of interval I along coordinate direction i at time t is denoted

$$\mathsf{disc}_t^i(I, \mathbf{x}, \chi) := \Big|\chi_1 \mathbf{1}_I(x_1(i)) + \cdots + \chi_t \mathbf{1}_I(x_t(i))\Big|.$$

We will just write $\mathsf{disc}_t^i(I)$ when the input sequence and signing is clear from the context.

Upper Bounds. To maintain the discrepancy of all intervals, it will suffice to bound the discrepancy of every dyadic interval $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ of length at least $1/T$ along every coordinate direction i. Let $\mathcal{D} = \{I_{j,k} \mid 0 \le j \le \log T, 0 \le k < 2^j\}$. Then, we prove the following.

Lemma 7.6.5. *Given any sequence $x_1, \dots, x_T$ sampled independently and uniformly from $[0,1]^d$, there is an online algorithm that chooses a signing such that w.h.p. for every time $t \in [T]$, we have*

$$\max_{i\in[d]} \mathsf{disc}_t^i(I) = O(d \log^2 T) \quad \text{for all } I \in \mathcal{D}.$$

Before proving Lemma 7.6.5, we first show why it implies the upper bound in Theorem 7.1.2.

Proof of the upper bound in Theorem 7.1.2. Without loss of generality, it suffices to consider half-open intervals. Every half-open interval $I \subseteq [0,1]$ can be decomposed as a union of at most $2\log T$ disjoint dyadic intervals in $\mathcal{D}$ and two intervals $I_1 \subseteq I_{\log T,k}$ and $I_2 \subseteq I_{\log T,k'}$ for some $0 \le k, k' < T$. Note that the length of I_1 and I_2 is at most $2^{-\log T} = 1/T$. We can then write,

$$\mathsf{disc}_t^i(I) \le (2\log T) \cdot \max_{I \in \mathcal{D}} \mathsf{disc}_t^i(I) + \mathsf{disc}_t^i(I_1) + \mathsf{disc}_t^i(I_2).$$

Applying the algorithm from Lemma 7.6.5, the discrepancy of every dyadic interval can be bounded w.h.p. by $O(d\log^2 T)$. The last two terms can be bounded by N_1 and N_2 respectively where N_1 (resp. N_2) is the number of points whose projections on any of the i coordinates is in I_1 (resp. I_2).

The probability that a random point z drawn uniformly from $[0,1]^d$ has some coordinate $z(i)$ for $i \in [d]$ in I_1 or I_2 is at most $2d/T$. It follows that $\mathbb{E}[N_1 + N_2] \le 2d$, so by Chernoff bounds, with probability at least $1 - T^{-4}$, the number $N_1 + N_2 \le 4d\log T$.

Overall, w.h.p. for any interval I, we have

$$\max_{i \in [d]} \mathsf{disc}_t^i(I) \quad \le \quad 2\log T \cdot (d\log^2 T) + 4d\log T \quad = \quad O(d\log^3 T).$$

□

Next, we prove the missing Lemma 7.6.5.

Proof of Lemma 7.6.5. We will consider the d-dimensional interval discrepancy problem as a vector balancing problem in $|[d] \times \mathcal{H}_{\le \log T}|$ dimensions, where $\mathcal{H}_{\le \log T}$ are the Haar wavelet functions with scale parameter at most $\log T$. Note that $|\mathcal{H}_{\le \log T}| = T$, so the update vector in the vector balancing version will be Td-dimensional. Let us abbreviate $\mathcal{H}' = \mathcal{H}_{\le \log T}$.

At any time when the point $x_t \in [0,1]$ arrives, then the (i,h) coordinate of the update vector $v_t \in [-1,1]^{d \times \mathcal{H}'}$ is given by

$$v_t(i,h) = h(x_t(i)).$$

Note that all the coordinates $(i, \Psi_{0,0})$ for $i \in [d]$ will always have the same value where $\Psi_{0,0}$ is constant Haar wavelet. So, to apply the online algorithm given by Lemma 7.2.1 we will only consider the subspace spanned by the coordinates (i,h) where $i \in [d]$ and $h \neq \Psi_{0,0}$ and the extra coordinate $(1, \Psi_{0,0})$.

Let us check first that we satisfy the conditions Lemma 7.2.1. First, note that the $\|v_t\|_\infty \leq 1$ and the vector v_t has at most $d \log T + 1$ non-zero coordinates, since for any fixed scale $0 \leq j \leq \log T$ and any point $z \in [0,1]$, all but one of the values $\{h(z)\}_{h \in \mathcal{H}_j}$ are zero. The last condition to check is that the coordinates of the vector v_t are uncorrelated. This is a consequence of (7.8), since whenever coordinates (i,h) and (i',h') satisfy $i \neq i'$ or $h \neq h'$, we have

$$\mathbb{E}_{v_t}[v_t(i,h) \cdot v_t(i',h')] \quad = \quad \mathbb{E}_{x_t}[h(x_t(i)) \cdot h'(x_t(i'))] \quad = \quad 0.$$

To elaborate more, first note that we cannot have $h = h' = \Psi_{0,0}$ since we are working in the aforementioned subspace. Now, if $i \neq i'$ then the coordinates $x_t(i)$ and $x_t(i)$ are sampled independently from $[0,1]$, and $\mathbb{E}_z[h(z)] = 0$ for $h \neq \Psi_{0,0}$ when z is drawn uniformly from $[0,1]$. Otherwise, for $i = i'$ but $h \neq h'$, it follows from the orthogonality of the Haar system that $\mathbb{E}_z[h(z)h'(z)] = 0$.

Next, applying the online algorithm from Lemma 7.2.1, we select signs $\chi_1, \ldots, \chi_T$ such that we get an ℓ_∞ bound on the vector $d_t = \sum_{l \leq t} \chi_l v_l$. In particular, with high probability we have

$$|d_t(i,h)| \quad = \quad \left| \sum_{l \leq t} \chi_l h(x_l(i)) \right| \quad = \quad O(d \log^2 T) \quad \text{for any } i \in [d], h \in \mathcal{H}'.$$

Note that the bound on $|d_t(i, \Psi_{0,0})|$ for $i \neq 1$ follows since $|d_t(i, \Psi_{0,0})| = |d_t(1, \Psi_{0,0})|$.

To finish the proof, we need to bound the discrepancy of every dyadic interval in terms of $\|d_t\|_\infty$. Note that for any dyadic interval $I \in \mathcal{D}$, its coefficients in the Haar system basis $\widehat{\mathbf{1}}_I(h) = 0$ for $h \in \mathcal{H}_{>\log T}$ using Proposition 7.6.3. Now, for any $i \in [d]$ and dyadic interval $I \in \mathcal{D}$, we can write

$$\begin{aligned}
\mathsf{disc}_t^i(I) = \Big|\sum_{l\le t} \chi_l \mathbf{1}_I(x_l(i))\Big| &= \Big|\sum_{l\le t}\chi_l \sum_{h\in\mathcal{H}'} \widehat{\mathbf{1}}_I(h)h(x_l(i))\Big| \\
&= \Big|\sum_{h\in\mathcal{H}'} \widehat{\mathbf{1}}_I(h)\Big(\sum_{l\le t}\chi_l h(x_l(i))\Big)\Big| = \Big|\sum_{h\in\mathcal{H}'}\widehat{\mathbf{1}}_I(h)d_t(i,h)\Big| \\
&\le \|d_t\|_\infty \cdot \Big(\sum_{h\in\mathcal{H}'}|\widehat{\mathbf{1}}_I(h)|\Big) \le \|d_t\|_\infty = O(d\log^2 T),
\end{aligned}$$

where the second last inequality follows again from Proposition 7.6.3. □

Lower Bounds. Next we present the proof the lower bound in Theorem 7.1.2.

Proof of the lower bound in Theorem 7.1.2. Set $A = T/d$. We will again consider the d-dimensional interval discrepancy problem as a vector balancing problem in $|[d] \times \mathcal{H}_{\le \log A}|$ dimensions where $\mathcal{H}_{\le \log A}$ are the Haar wavelet functions with scale parameter at most $\log A$. Note that $|\mathcal{H}_{\le \log T}| = A$, so the update vector in the vector balancing version will be T-dimensional. Let us abbreviate $\mathcal{H}' = \mathcal{H}_{\le \log A}$.

At any time when the point $x_t \in [0,1]^d$ arrives, then the (i,h) coordinate of the update vector v_t is given by

$$v_t(i,h) = \begin{cases} 0 & \text{if } h = \Psi_{0,0} \\ h(x_t(i)) & \text{otherwise.} \end{cases}$$

Here we are essentially ignoring the coordinates (i,h) with $h = \Psi_{0,0}$. Since for any fixed scale $0 < j \le \log A$ and any point $z \in [0,1]$, all but one of the values $\{h(z)\}_{h\in\mathcal{H}_j}$ are zero, the vector v_t has $d\log A$ non-zero coordinates all of which take value ± 1. It follows that the Euclidean norm of any update vector v_t is $\sqrt{d\log A}$.

Furthermore, from the orthogonality of the Haar system, it follows that the coordinates

of the vector v_t are uncorrelated:

$$\mathbb{E}_{v_t}[v_t(i,h)v_t(i',h')] = \mathbb{E}_{x_t}[h(x_t(i))h'(x_t(i'))] = 0.$$

Then, applying Lemma 7.4.1, we get that with probability at least $3/4$, there is a $t \in [T]$ and a coordinate (i,h) with $h \neq \Psi_{0,0}$ such that $|d_t(i,h)| = \Omega(\sqrt{d \log A})$.

Let $h = \Psi_{j,k}$ for some j,k where $j > 0$ (recall that coordinates (i,h) where $h = \Psi_{0,0}$ are always 0). Then, by definition $h = \mathbf{1}_{I_1} - \mathbf{1}_{I_2}$ where I_1 and I_2 are the first and second halves of the interval $I_{j-1,k}$. In this case,

$$|d_t(i,h)| \;=\; \Big|\Big(\sum_{s\le t}\chi_t \mathbf{1}_{I_1}(x_s)\Big) - \Big(\sum_{s\le t}\chi_t \mathbf{1}_{I_2}(x_s)\Big)\Big| \;\le\; 2\max\Big\{|\mathsf{disc}_t(I_1)|, |\mathsf{disc}_t(I_2)|\Big\}.$$

Therefore, substituting $A = T/d$, there exists an interval I such that $\mathsf{disc}_t^i(I) = \Omega\left(\sqrt{d\log\left(\frac{T}{d}\right)}\right)$. □

7.6.3 Online Tusnády's Problem

Let $\mathbf{x} = (x_1, \ldots, x_T)$ be a sequence of points in $[0,1]^d$ and let $\chi \in \{\pm 1\}^T$ be a signing. For any axis-parallel box $B \subseteq [0,1]^d$ and any time $t \in [T]$, recall that the discrepancy of axis-parallel box B at time t is denoted

$$\mathsf{disc}_t(B, \mathbf{x}, \chi) := \Big|\chi(1)\cdot \mathbf{1}_B(x_1) + \ldots + \chi(t)\cdot \mathbf{1}_B(x_t)\Big|.$$

We will just write $\mathsf{disc}_t(B)$ when the input sequence and signing is clear from the context.

Upper Bounds. As in the interval case, it will be sufficient to work with dyadic boxes. Recall that $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ for $j \in \mathbb{Z}_{\ge 0}$ and $0 \le k < 2^j$. To maintain the discrepancy of all boxes, it will suffice to bound the discrepancy of every dyadic box

$$B_{\boldsymbol{j},\boldsymbol{k}} := I_{\boldsymbol{j}(1),\boldsymbol{k}(1)} \times \ldots \times I_{\boldsymbol{j}(d),\boldsymbol{k}(d)},$$

with $\boldsymbol{j}, \boldsymbol{k} \in \mathbb{Z}^d$ with $0 \le \boldsymbol{j}$ and $0 \le \boldsymbol{k} < 2^{\boldsymbol{j}}$ with each side length at least $1/T$. In particular, let $\mathcal{D} = \{B_{\boldsymbol{j},\boldsymbol{k}} \mid 0 \le \boldsymbol{j} \le (\log T)\mathbf{1}\ ,\ 0 \le \boldsymbol{k} < 2^{\boldsymbol{j}}\}$ where $\mathbf{1} \in \mathbb{R}^d$ is the all ones vector. Then, we prove the following lemma to bound the discrepancy of every dyadic box.

Lemma 7.6.6. *Given any sequence $x_1, \ldots, x_T$ sampled independently and uniformly from $[0,1]^d$, there is an online algorithm that chooses a signing such that w.h.p. for every time $t \in [T]$,*

$$\mathsf{disc}_t(B) = O\left(\log^{d+1} T\right), \text{ for all } B \in \mathcal{D}.$$

Before proving Lemma 7.6.6, we first show why it implies Theorem 7.1.3.

Proof of the upper bound in Theorem 7.1.3. Without loss of generality, it suffices to consider axis-parallel boxes $B = I_1 \times \cdots \times I_d$ where I_j's are half-open sub-intervals of $[0,1]$. Recall that every half-open interval $I \subseteq [0,1]$ can be decomposed as a union of at most $2\log T$ disjoint dyadic intervals in $\mathcal{D}$ and two intervals $I' \subseteq I_{\log T, k}$ and $I'' \subseteq I_{\log T, k'}$ for some $0 \le k, k' < T$ (note that the length of I' and I'' is at most $2^{-\log T} = 1/T$).

From this, it follows that for any axis-parallel box B, there exists a set of dyadic boxes $\mathcal{D}' \subseteq \mathcal{D}$ of size $|\mathcal{D}'| = (2\log T)^d$ and a set $\mathcal{I}$ of size $|\mathcal{I}| = 2d$ of disjoint intervals of length at most $1/T$, such that B can be decomposed as the union of boxes in $\mathcal{D}'$ and some other boxes of the form $I'_1 \times \cdots \times I'_d$, where $I'_i \in \mathcal{I}$ for at least one $i \in [d]$. We can therefore bound,

$$\mathsf{disc}_t(B) \le (2\log T)^d \cdot \left(\max_{B \in \mathcal{D}} \mathsf{disc}_t(B)\right) + N,$$

where N is the number of points z in the input sequence such that $z(i) \in I$ for some $i \in [d]$ and $I \in \mathcal{I}$.

Applying the algorithm from Lemma 7.6.6, the discrepancy of every dyadic box can be bounded by $O(\log^{d+1} T)$ with high probability. Also, since the length of every interval in $\mathcal{I}$ is at most $1/T$, for z drawn uniformly from $[0,1]^d$, we have

$$\mathbb{P}_z\Big[\exists i \in [d], \exists I \in \mathcal{I} \text{ such that } z(i) \in I\Big] \le \frac{2d^2}{T}.$$

Therefore, we have that $\mathbb{E}[N] \leq 2d^2$ and applying Chernoff bounds, it follows that with probability at least $1 - T^{-4}$, the number $N \leq 4d^2 \log T$.

Overall, with high probability for any axis-parallel box B, we have

$$\mathsf{disc}_t(B) \quad \leq \quad (2\log T)^d(\log^{d+1} T) + 4d^2 \log T \quad = \quad O_d(\log^{2d+1} T).$$

□

Next, we prove the missing Lemma 7.6.6.

Proof of Lemma 7.6.6. We will consider the d-dimensional Tusnády's problem as a vector balancing problem in $\mathcal{H}_{\leq \log T}^{\otimes d}$ dimensions where $\mathcal{H}_{\leq \log T}$ are the Haar wavelet functions with scale parameter at most $\log T$. Note that $|\mathcal{H}_{\leq \log T}| = T$, so the update vector in the vector balancing version will be T^d-dimensional. Let us abbreviate $\mathcal{H}' = \mathcal{H}_{\leq \log T}$ and also recall that for any $\boldsymbol{h} = (h_1, \ldots, h_d)$ in $\mathcal{H}'^{\otimes d}$, we view it as a function from the cube $[0,1]^d$ to $\mathbb{R}$ by defining $\boldsymbol{h}(x) = \prod_{i=1}^d h_i(x(i))$.

At any time when the point $x_t \in [0,1]^d$ arrives, then the $\boldsymbol{h} := (h_1, \ldots, h_d)$ coordinate of the update vector $v_t \in [-1,1]^{\mathcal{H}'^{\otimes d}}$ is given by

$$v_t(\boldsymbol{h}) \quad = \quad \boldsymbol{h}(x_t) \quad = \quad \prod_{i=1}^d h_i(x_t(i)).$$

We will apply the online algorithm given by Lemma 7.2.1. Let us check first that we satisfy the conditions of that lemma. First, note that the $\|v_t\|_\infty \leq 1$ and the vector v_t has at most $(\log T + 1)^d$ non-zero coordinates, since for any fixed scale $0 \leq j \leq \log T$ and any point $z \in [0,1]$, all but one of the values $\{h(z)\}_{h \in \mathcal{H}_j}$ are zero. The last condition to check is that the coordinates of the vector v_t are uncorrelated. This follows from the orthogonality of $\boldsymbol{h}$ and $\boldsymbol{h}'$. In particular, if $\boldsymbol{h} \neq \boldsymbol{h}'$, then

$$\mathbb{E}_{v_t}[v_t(\boldsymbol{h})v_t(\boldsymbol{h}')] \quad = \quad \mathbb{E}_{x_t}[\boldsymbol{h}(x_t)\boldsymbol{h}'(x_t)] \quad = \quad 0.$$

Applying the online algorithm from Lemma 7.2.1, we select signs $\chi_1, \ldots, \chi_T$ such that we get an ℓ_∞ bound on the vector $d_t = \chi_1 v_1 + \ldots \chi_t v_t$. In particular, with high probability

$$|d_t(\boldsymbol{h})| \quad = \quad \Big|\sum_{l \le t} \chi_l \boldsymbol{h}(x_l)\Big| \quad = \quad O(\log^{d+1} T) \text{ for any } \boldsymbol{h} \in \mathcal{H}'^{\otimes d}.$$

To finish the proof, we next bound the discrepancy of every dyadic box in terms of $\|d_t\|_\infty$. For any dyadic box $B \in \mathcal{D}$, since each side consists of dyadic interval $I_{j,k}$ where $j \le \log T$, Proposition 7.6.4 implies that $\widehat{\mathbf{1}}_I(\boldsymbol{h}) = 0$ for any $\boldsymbol{h} \notin \mathcal{H}'^{\otimes d}$. Therefore, we have

$$\begin{aligned}
\mathsf{disc}_t(B) = \Big|\sum_{l \le t} \chi_l \mathbf{1}_B(x_l)\Big| \quad &= \quad \Big|\sum_{l \le t} \chi_l \sum_{\boldsymbol{h} \in \mathcal{H}'^{\otimes d}} \widehat{\mathbf{1}}_B(\boldsymbol{h}) \boldsymbol{h}(x_l)\Big| \\
= \Big|\sum_{\boldsymbol{h} \in \mathcal{H}'^{\otimes d}} \widehat{\mathbf{1}}_B(\boldsymbol{h}) \Big(\sum_{l \le t} \chi_l \boldsymbol{h}(x_l)\Big)\Big| \quad &= \quad \Big|\sum_{\boldsymbol{h} \in \mathcal{H}'^{\otimes d}} \widehat{\mathbf{1}}_B(\boldsymbol{h}) d_t(\boldsymbol{h})\Big| \\
\le \|d_t\|_\infty \cdot \Big(\sum_{\boldsymbol{h} \in \mathcal{H}'^{\otimes d}} |\widehat{\mathbf{1}}_B(\boldsymbol{h})|\Big) \quad &\le \quad \|d_t\|_\infty \quad = \quad O(\log^{d+1} T),
\end{aligned}$$

where the second last inequality follows again from Proposition 7.6.4. □

Lower Bounds.

Proof of the lower bound in Theorem 7.1.3. Set $A = T^{1/d}$. We will consider the d-dimensional interval discrepancy problem as a vector balancing problem in $\mathcal{H}^{\otimes d}_{\le \log A}$ dimensions where $\mathcal{H}_{\le \log A}$ are the Haar wavelet functions with scale parameter at most $\log A$. Note that $|\mathcal{H}_{\le \log A}| = A$, so the update vector in the vector balancing version will be A^d-dimensional. Let us abbreviate $\mathcal{H}' = \mathcal{H}_{\le \log A}$ and also recall that for any $\boldsymbol{h} = (h_1, \ldots, h_d)$ in $\mathcal{H}'^{\otimes d}$, we view it as a function from the cube $[0,1]^d$ to $\mathbb{R}$ by defining $\boldsymbol{h}(x) = \prod_{i=1}^d h_i(x(i))$.

At any time when the point $x_t \in [0,1]^d$ arrives, then the $\boldsymbol{h} := (h_1, \ldots, h_d)$ coordinate of the update vector $v_t \in [-1,1]^{\mathcal{H}'^{\otimes d}}$ is given by

$$v_t(\boldsymbol{h}) \quad = \quad \boldsymbol{h}(x_t) \quad = \quad \prod_{i=1}^d h_i(x_t(i)).$$

We will apply Lemma 7.4.1. Let us check first that we satisfy the conditions of that lemma. Similar to the proof of Lemma 7.6.6, we note that the vector v_t has exactly $(\log A+1)^d$ non-zero coordinates that take the value ± 1. This implies that that Euclidean norm of any update v_t is $(\log A + 1)^{d/2}$. Also from the orthogonality of $\boldsymbol{h}$ and $\boldsymbol{h}'$, the coordinates of the vector v_t are uncorrelated — if $\boldsymbol{h} \neq \boldsymbol{h}'$, then

$$\mathbb{E}_{v_t}[v_t(\boldsymbol{h})v_t(\boldsymbol{h}')] \quad = \quad \mathbb{E}_{x_t}[\boldsymbol{h}(x_t)\boldsymbol{h}'(x_t)] \quad = \quad 0.$$

Applying Lemma 7.4.1 tells us that with probability at least $3/4$, there exists a time $t \in [T]$ and a $\boldsymbol{h} \in \mathcal{H}'$ such that $|d_t(\boldsymbol{h})| = \Omega\left((\log A + 1)^{d/2}\right)$. Note that since $\boldsymbol{h}(x) = \prod_{i=1}^{d} h_i(x(i))$ and h_i can always be expressed as $\mathbf{1}_{I_i}$ or $\mathbf{1}_{I_i} - \mathbf{1}_{I_i'}$ for some intervals I_i and I_i', it follows that there exists a set $\mathcal{B}$ of at most 2^d axis-parallel boxes and some $\chi \in \{\pm 1\}^{\mathcal{B}}$ such that

$$d_t(\boldsymbol{h}) = \sum_{B \in \mathcal{B}} \chi_B \cdot \mathsf{disc}_t(B).$$

By averaging, it follows that there is an axis-parallel box $B \in \mathcal{B}$ such that $\mathsf{disc}_t(B) \geq \dfrac{|d_t(\boldsymbol{h})|}{2^d}$.

Substituting $A = T^{1/d}$, we get that for some box B,

$$\mathsf{disc}_t(B) \quad = \quad \Omega\Big(\frac{1}{2^d} \cdot \log^{d/2} A\Big) \quad = \quad \Omega_d(\log^{d/2} T).$$

□

7.7 Applications to Online Envy Minimization

In this section we use our vector balancing and two-dimensional interval discrepancy results to bound online envy. Let us first give the formal definition of envy.

Recall that there are two players and T items where for item $t \in \{1, \ldots, T\}$, the valuation of the player $i \in \{1, 2\}$ is $v_{it} \in [0, 1]$. The *cardinal envy* is the standard notion of envy studied in fair division, which is the max over every player the difference between the

player's valuation for the other player's allocation and the player's valuation for their own allocation [LMMS04, Bud11]. Formally, if Player i is allocated set S_i by an algorithm, the cardinal envy is defined as

$$\mathsf{envy}_C(\mathbf{v}_1, \mathbf{v}_2, S_1, S_2) := \max\Big\{ \sum_{t\in S_2} v_{1t} - \sum_{t\in S_1} v_{1t} \ , \ \sum_{t\in S_1} v_{2t} - \sum_{t\in S_2} v_{2t} \Big\}.$$

The notion of ordinal envy is defined ignoring the precise item valuations, but only with respect to the relative ordering of the items. Roughly, it is the worst possible cardinal envy for $[0,1]$ valuations consistent with any given relative ordering. Thus for valuations in $[0,1]$ the ordinal envy is always at least the cardinal envy [JKS19]. For $i \in \{1,2\}$, let π_i denote the decreasing order with respect to the valuations v_{it}. Denote π_i^t the first t items in the order π. If Player i is allocated set S_i, the ordinal envy is defined as

$$\mathsf{envy}_O(\pi_1, \pi_2, S_1, S_2) := \max_{t\geq 0} \Big\{ |S_2 \cap \pi_1^t| - |S_1 \cap \pi_1^t| \ , \ |S_1 \cap \pi_2^t| - |S_2 \cap \pi_2^t| \Big\}.$$

Jiang et al. [JKS19] discuss three equivalent definitions of ordinal envy.

Next, we prove Corollary 7.1.7, which is restated below.

Corollary 7.1.7. *Suppose valuations of two players are drawn i.i.d. from some distribution* p *over* $[0,1]\times[0,1]$. *Then, for an arbitrary distribution* p *(i.e., player valuations for the same item could be correlated), the online cardinal envy is* $O(\log T)$. *Moreover, if* p *is a product distribution (i.e., player valuations for the same item are independent) then the online ordinal envy is also* $O(\log^3 T)$.

Proof. When the player valuations are drawn independently in $[0,1]$, the "moreover" part is immediate from the following lemma of [JKS19] along with our Theorem 7.1.2 for 2-dimensional interval discrepancy.

Lemma 7.7.1 (Lemma 26 in [JKS19])**.** *For two players with independent valuations, any upper bound for 2-dimensional interval discrepancy problem also holds for 2-player online ordinal envy minimization.*

Next, we bound online cardinal envy under arbitrary distributions. In the following lemma we reduce this problem to 2-dimensional vector balancing.

Lemma 7.7.2. *For two players taking values from an arbitrary distribution* p *over* $[0,1] \times [0,1]$*, any upper bound for* 2*-dimensional vector balancing problem also holds for* 2*-player online cardinal envy minimization.*

Proof. For $i \in \{1,2\}$, let u_{it} denote the valuation of Player i for t^{th} item. We define the corresponding vector $v_t = (u_{1t}, -u_{2t})$. If our online vector balancing algorithm assigns the next vector v_t a $+$ sign, we give the item to Player 2, and otherwise we give it to Player 1. The crucial observation is that $d_t(1)$ and $d_t(2)$ capture precisely the cardinal envy of Players 1 and 2, respectively. Thus, any bound $\|d_t\|_\infty$ implies a bound on the maximum cardinal envy. □

The last lemma when combined with Theorem 7.1.4 finishes the proof of Corollary 7.1.7. □

7.8 Open Problems and Directions

We close this chapter by mentioning some interesting open problems that seem to require fundamental new techniques, and new directions in online discrepancy that remain unexplored.

Improving the dependence on n for general distributions. Theorem 7.1.4 gives a bound of $O(n^2 \log T)$ for online vector balancing problem under inputs sampled from an arbitrary distribution. However, an optimal dependence of $O(n^{1/2})$ on n is achievable in the special case where the distribution has independent coordinates [BS20], and also in the offline setting with worst-case inputs [Ban12]. This motivates the following question.

Question 1. *Given an arbitrary distribution* p *supported over* $[-1,1]^n$*, is there an online algorithm that maintains discrepancy* $\sqrt{n} \cdot \mathsf{polylog}(T)$ *on a sequence of* T *inputs sampled i.i.d. from* p*?*

As the anti-concentration bound in Theorem 7.1.5 for uncorrelated variables is a $n^{1/2}$ factor worse than that for independent random variables, even getting a dependence of $n \cdot \mathsf{polylog}(T)$ is an interesting first step.

Bounds in terms of sparsity. Several natural problems such as the d-dimensional interval discrepancy and d-dimensional Tusnády's problem are best viewed as vector balancing problems where the input vectors are sparse. This motivates the following online version of the Beck-Fiala problem, where the online sequence $x_1, \ldots, x_T$ is chosen independently from some distribution p supported over s-sparse n-dimensional vectors over $[-1, 1]^n$. In the offline setting with worst-case inputs (and where we care about the discrepancy of every prefix), the methods of Banaszczyk [Ban12] give a bound of $(s \log T)^{1/2}$.

Question 2. *Given an arbitrary distribution* p *supported over* s*-sparse vectors in* $[-1, 1]^n$*, is there an online algorithm that maintains discrepancy* $\mathsf{poly}(s, \log T, \log n)$ *on a sequence of* T *inputs sampled i.i.d. from* p*?*

Resolving the above question would imply polylogarithmic bounds for Tusnády's problem in d-dimensions (similar to that in Theorem 7.1.3) in the much more general setting where the points x_T are sampled from an arbitrary distribution over points in $[0, 1]^d$. Currently, Theorems 7.1.2 and 7.1.3 only hold when the points x_t are sampled from a product distribution on $[0, 1]^d$.

Prophet model. The last decade has seen several online problems being studied in the prophet model where the online inputs are sampled independently from known *non-identical* distributions (see, e.g., [Luc17]). The model clearly generalizes the i.i.d. model and for point mass distributions it captures the offline problem. This model becomes useful for online problems where the adversarial arrival guarantees are weak, which raises the following question.

Question 3. *Given arbitrary distributions* $\mathsf{p}_1, \ldots, \mathsf{p}_T$ *supported over vectors in* $[-1, 1]^n$*, is there an online algorithm that maintains discrepancy* $\mathsf{poly}(n, \log T, \log n)$ *on a sequence of*

T inputs where vector v_t is sampled independently from p_t?

The techniques in Theorem 7.1.4 do not work since the eigenbasis may change with each arrival. It will be also interesting to study this prophet model for distributions over s-sparse vectors.

Oblivious adversary model. A very interesting direction that is strictly harder than the above stochastic settings is to understand online vector balancing when the adversary is oblivious or non-adaptive, i.e., the adversary chooses the entire input sequence (without any stochastic assumptions) beforehand and is *not allowed* to change the inputs later based on the execution of the algorithm.

Recall that if the adversary is fully adaptive, then one cannot hope to prove a bound better than $\Theta(T^{1/2})$, but this might be possible for oblivious adversaries.

Question 4. *Is there an online algorithm that maintains discrepancy $\mathsf{poly}(n, \log T)$ on any sequence of T vectors in $[-1, 1]^n$ chosen by an oblivious adversary?*

One could also consider the same question in the Beck-Fiala setting, and ask if better bounds are possible when there is sparsity.

Question 5. *Is there an online algorithm that maintains discrepancy $\mathsf{poly}(s, \log T, \log n)$ on any sequence of T vectors in $[-1, 1]^n$ that are s-sparse and chosen by an oblivious adversary?*

Resolving Questions 4 and 5 would also have implications for both online geometric discrepancy and online envy minimization problems in the oblivious adversary setting.

7.9 Tight example for Anti-Concentration in the Original Basis for Interval Discrepancy

Let us briefly recall the setting. Consider the complete binary tree of height $\log T$ where the nodes are the dyadic intervals $I_{j,k}$ for $0 \leq j \leq \log T$ and $0 \leq k < 2^j$. Our objective was to

find the smallest β such that

$$\mathbb{E}_x\Big[\Big|\sum_{j,k}\sinh(\lambda d_{j,k})\cdot\mathbf{1}_{I_{j,k}}(x)\Big|\Big] \;\geq\; \frac{1}{\beta}\cdot\mathbb{E}_x\Big[\sum_{j,k}\cosh(\lambda d_{j,k})\cdot\mathbf{1}_{I_{j,k}}(x)\Big], \tag{7.10}$$

where x is a uniform point on the unit interval $[0,1]$, the function $\mathbf{1}_{I_{j,k}}$ is the indicator for the dyadic interval $I_{j,k}$, and $\lambda > 0$ and $d_{j,k} \in \mathbb{R}$. For simplicity, we set $\lambda = 1$ henceforth.

Observe that when a uniform random point x arrives at a leaf dyadic interval $I_{\log T,k}$, then only the variables along that root-leaf path contribute to both sides. Moreover, since x is uniform, the chosen leaf interval is also uniform among the leaves. Therefore, denoting by ℓ the random leaf and P_ℓ the corresponding root-leaf path, we want to ask for the smallest β satisfying

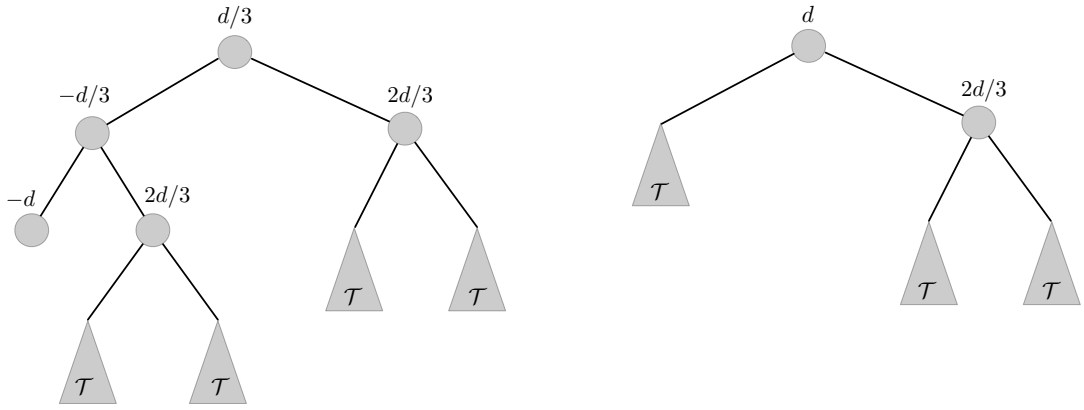
$$\mathbb{E}_{P_\ell}\Big[\Big|\sum_{I_{j,k}\in P_\ell} a_{j,k}\Big|\Big] \;\geq\; \frac{1}{\beta}\cdot\mathbb{E}_{P_\ell}\Big[\sum_{I_{j,k}\in P_\ell}|a_{j,k}|\Big], \tag{7.11}$$

where $a_{j,k} = \sinh(d_{j,k})$ for a node $I_{j,k}$ in the dyadic tree. Note that to get (7.11) from (7.10), we made the standard approximation that $\cosh(x) \approx |\sinh(x)|$ for $x \in \mathbb{R}$.

The following lemma shows that in general β could be exponentially large in the height of the tree, so in the above case since the height is $\log T$, the value of $\beta = \Omega(\mathsf{poly}(T))$. We remark that for non-binary trees, this was already shown by Jiang, Kulkarni, and Singla [JKS19].

Lemma 7.9.1. *There exists $d_{j,k}$ for $0 \leq j \leq h$ and $0 \leq k < 2^j$, such that $\beta = \exp(\Omega(h))$ in (7.11).*

Proof. Our construction has a fractal structure. Let $d > 0$ be a sufficiently large integer. Let $\mathcal{T}$ denote the tree structure shown in Figure 7.4(a) where the labels are the values that will be used for constructing $d_{j,k}$'s. We embed this structure in the complete binary tree of dyadic intervals and assign the $d_{j,k}$ values as follows: the root interval has value $d_{0,0} = d$ and its left children has the structure $\mathcal{T}$ with the values $d_{j,k}$ as assigned by the corresponding labels in $\mathcal{T}$, while the right child has value $d_{1,1} = 2d/3$ and has two child subtrees with



(a) Fractal structure $\mathcal{T}$ (b) Embedding of $\mathcal{T}$ in the dyadic tree

Figure 7.4: Construction of $d_{j,k}$'s satisfying (7.11)

structure $\mathcal{T}$ (see Figure 7.4(b)). The $d_{j,k}$ values for all the unassigned nodes (these lie in the subtree rooted at the nodes having values $d_{j,k} = -d$) are taken to be zero.

Note that $\mathcal{T}$ has the property that with probability $1/4$ it ends in a node $I_{j,k}$ with $a_{j,k} = \sinh(-d)$, and otherwise it enters another $\mathcal{T}$ (unless we already reached a leaf).

The proof now follows because if we take a random root-leaf path in our dyadic tree, with probability $1-\exp(-\Omega(h))$ it will end in a leaf with $\sinh(-d)$, which will cancel with $\sinh(d)$ at the root. Since every other entry on a root leaf path has magnitude at most $\sinh(2d/3)$, the left hand side in (7.11) will be

$$\begin{aligned}\mathbb{E}_{P_\ell}\Big[\Big|\sum_{I_{j,k}\in P_\ell} a_{j,k}\Big|\Big] &\le \Big(1-\exp(-\Omega(h))\Big)\cdot h\cdot|\sinh(2d/3)| + \exp(-\Omega(h))\cdot h\cdot|\sinh(d)| \\ &\le \frac{|\sinh(d)|}{\exp(\Omega(h))},\end{aligned}$$

while the right hand side is

$$\mathbb{E}_{P_\ell}\Big[\sum_{I_{j,k}\in P_\ell}|a_{j,k}|\Big] \quad\ge\quad |\sinh(d)|.$$

Therefore, $\beta = \exp(\Omega(h))$ in (7.11). □

7.10 Burkholder-Davis-Gundy Inequality

Let $Z_0, Z_1, \ldots, Z_t$ be a discrete martingale (with respect to $W_1, \ldots, W_t$) and let $\Delta Z_s = Z_s - Z_{s-1}$ denote the differences for all $s \in [t]$. Note that $Z_s = \Delta Z_1 + \Delta Z_2 + \ldots + \Delta Z_s$. Define $Z_t^* = \max_{0 \leq s \leq t} |Z_s|$ to be the maximum value of the martingale process till time t. Then, the well-known Burkholder-Davis-Gundy inequality says the following.

Theorem 7.10.1 ([BDG72])**.** *Let $1 \leq p < \infty$. Then, there exist positive constants c_p and C_p such that*

$$c_p \cdot \mathbb{E}\left[\left(\sum_{s=1}^{t} |\Delta Z_s|^2\right)^{p/2}\right] \quad \leq \quad \mathbb{E}[(Z_t^*)^p] \quad \leq \quad C_p \cdot \mathbb{E}\left[\left(\sum_{s=1}^{t} |\Delta Z_s|^2\right)^{p/2}\right].$$

Note that the inequality holds in much more general settings, but the above setting is sufficient for the purposes of this chapter.

Furthermore, for $p = 1$, which is the case we need for the purposes of this chapter, one can relate expected magnitude of Z_t^* and Z_t by the following inequality.

Lemma 7.10.2. $\mathbb{E}[Z_t^*] \leq (t+1) \cdot \mathbb{E}[|Z_t|]$.

Proof. First note that $f(Z_0), \ldots, f(Z_t)$ is a sub-martingale with respect to $W_1, \ldots, W_t$ for any convex function f. Choosing $f(z) = |z|$, it follows that the absolute value of the above martingale is a sub-martingale. Applying Doob's optional stopping theorem to this sub-martingale, one gets that $\mathbb{E}[|Z_t|] \geq \mathbb{E}[|Z_0|]$. Since, we could have started this sequence anywhere, it also follows for any $s < t$ that $\mathbb{E}[|Z_t|] \geq \mathbb{E}[|Z_s|]$.

Since $Z_t^* = \max_{s \leq t} |Z_s| \leq \sum_{s=0}^{t} |Z_s|$, using linearity of expectation, we get that

$$\mathbb{E}[Z_t^*] \quad \leq \quad \sum_{s=0}^{t} \mathbb{E}[|Z_s|] \quad \leq \quad (t+1)\mathbb{E}[|Z_t|].$$

□

Chapter 8

ONLINE DISCREPANCY II: A BETTER POTENTIAL FUNCTION

In this chapter, we continue our study of the online discrepancy problem. While the discrepancy bound in Theorem 7.1.4 from Chapter 7 depends poly-logarithmically on the number of vectors T, its dependence on the dimension n is large and sub-optimal. In this chapter, we present a better algorithm that achieves an optimal dependence on the dimension n for the online discrepancy problem in the stochastic setting. Our new algorithm is based on a better potential function. This chapter is based on joint work with Nikhil Bansal, Raghu Meka, Sahil Singla, and Makrand Sinha [BJM+21] that appeared in the *ACM-SIAM Symposium on Discrete Algorithms (SODA21)*.

8.1 Introduction

We consider the following online vector balancing question, originally proposed by Spencer [Sho77]: vectors $v_1, v_2, \ldots, v_T \in \mathbb{R}^n$ arrive online, and upon the arrival of v_t, a sign $\chi_t \in \{\pm 1\}$ must be chosen irrevocably, so that the ℓ_∞-norm of the *discrepancy vector* (signed sum) $d_t := \chi_1 v_1 + \ldots + \chi_t v_t$ remains as small as possible. That is, find the smallest B such that $\max_{t\in[T]} \|d_t\|_\infty \leq B$. More generally, one can consider the problem of minimizing $\max_{t\in T} \|d_t\|_K$ with respect to arbitrary norms given by a symmetric convex body K.

Offline setting. The offline version of the problem, where the vectors $v_1, \ldots, v_T$ are given in advance, has been extensively studied in discrepancy theory, and has various applications [Mat99, Cha00, CST+14]. Here we study three important problems in this vein:

Tusnády's problem. Given points $x_1, \ldots, x_T \in [0,1]^d$, we want to assign $\pm$ signs to the points, so that for every axis-parallel box, the difference between the number of points inside

the box that are assigned a plus sign and those assigned a minus sign is minimized.

Beck-Fiala and Komlós problem. Given $v_1, \ldots, v_T \in \mathbb{R}^n$ with Euclidean norm at most one, we want to minimize $\max_{t \in T} \|d_t\|_\infty$. After scaling, a special case of the Komlós problem is the Beck-Fiala setting where $v_1, \ldots, v_T \in [-1, 1]^n$ are s-sparse (with at most s non-zeros).

Banaszczyk's problem. Given $v_1, \ldots, v_T \in \mathbb{R}^n$ with Euclidean norm at most one, and a convex body $K \in \mathbb{R}^n$ with Gaussian measure[1] $\gamma_n(K) \geq 1 - 1/(2T)$, find the smallest B so that there exist signs such that $d_t \in B \cdot K$ for all $t \in [T]$.

One of the most general and powerful results here is due to Banaszczyk [Ban12]: there exist signs such that $d_t \in O(1) \cdot K$ for all $t \in [T]$ for any convex body $K \in \mathbb{R}^n$ with Gaussian measure[2] $\gamma_n(K) \geq 1 - 1/(2T)$. In particular, this gives the best known bounds of $O((\log T)^{1/2})$ for the Komlós problem; for the Beck-Fiala setting, when the vectors are s-sparse, the bound is $O((s \log T)^{1/2})$.

An extensively studied case, where sparsity plays a key role, is that of Tusnády's problem (see [Mat99] for a history), where the best known (non-algorithmic) results, building on a long line of work, are an $O(\log^{d-1/2} T)$ upper bound of [Nik17] and an almost matching $\Omega(\log^{d-1} T)$ lower bound of [MN15].

In general, several powerful techniques have been developed for offline discrepancy problems over the last several decades, starting with initial non-constructive approaches such as [Bec81, Spe85, Glu89, Gia97, Ban98, Ban12], and more recent algorithmic ones such as [Ban10, LM15a, Rot17, MNT14, BDG16, LRR17, ES18, BDGL18, DNTTJ18]. However, none of them applies to the online setting that we consider here.

Online setting. A naïve algorithm is to pick each sign χ_t randomly and independently, which by standard tail bounds gives $B = \Theta((T \log n)^{1/2})$ with high probability. In typical interesting settings, we have $T \geq \mathsf{poly}(n)$, and hence a natural question is whether the

[1] The Gaussian measure $\gamma_n(\mathcal{S})$ of a set $\mathcal{S} \subseteq \mathbb{R}^n$ is defined as $\mathbb{P}[G \in \mathcal{S}]$ where G is standard Gaussian in $\mathbb{R}^n$.

[2] We remark that if one only cares about the final discrepancy d_T, the condition in Banaszczyk's result can be improved to $\gamma_n(K) \geq 1/2$ (though, in all applications we are aware of, this makes no difference if $T = \mathrm{poly}(n)$ and makes a difference of at most $\sqrt{\log T}$) for general T).

dependence on T can be improved from $T^{1/2}$ to say, poly-logarithmic in T, and ideally to even match the known offline bounds.

Unfortunately, the $\Omega(T^{1/2})$ dependence is necessary if the adversary is adaptive[3]: at each time t, the adversary can choose the next input vector v_t to be *orthogonal* to d_{t-1}, causing $\|d_t\|_2$ to grow as $\Omega(T^{1/2})$ (see [Spe87] for an even stronger lower bound). Even for very special cases, such as for vectors in $\{-1,1\}^n$, strong $\Omega(2^n)$ lower bounds are known [Bár79]. Hence, we focus on a natural *stochastic* model where we relax the power of the adversary and assume that the arriving vectors are chosen in an i.i.d. manner from some—possibly adversarially chosen—distribution p. In this case, one could hope to exploit that $\langle d_{t-1}, v_t\rangle$ is not always zero, *e.g.*, due to anti-concentration properties of the input distribution, and beat the $\Omega(T^{1/2})$ bound.

Recently, Bansal and Spencer [BS20], considered the special case where p is the uniform distribution on all $\{-1,1\}^n$ vectors, and gave an almost optimal $O(n^{1/2}\log T)$ bound for the ℓ_∞ norm that holds with high probability for all $t \in [T]$. The setting of general distributions p turns out to be harder and was considered recently by [JKS19] and [BJSS20], motivated by *envy minimization* problems and an online version of Tusnády's problem. The latter was also considered independently by Dwivedi, Feldheim, Gurel-Gurevich, and Ramadas [DFGGR19] motivated by the problem of placing points uniformly in a grid.

For an arbitrary distribution p supported on vectors in $[-1,1]^n$, [BJSS20] give an algorithm achieving an $O(n^2\log T)$ bound for the ℓ_∞-norm. In contrast, the best offline bound is $O((n\log T)^{1/2})$, and hence $\widetilde{\Omega}(n^{3/2})$ factor worse, where $\widetilde{\Omega}(\cdot)$ ignores poly-logarithmic factors in n and T.

More significantly, the existing bounds for the online version are much worse than those of the offline version for the case of s-sparse vectors (*Beck-Fiala* setting) — [BJSS20] obtain a much weaker bound of $O(sn\log T)$ for the online setting while the offline bound of $O((s\log T)^{1/2})$ is independent of the ambient dimension n. These technical limitations also

[3]In the sense that the adversary can choose the next vector v_t based on the current discrepancy vector d_{t-1}.

carry over to the online Tusnády problem, where previous works [JKS19, DFGGR19, BJSS20] could only handle product distributions.

To this end, [BJSS20] propose two key problems in the i.i.d. setting. First, for a general distribution p on vectors in $[-1,1]^n$, can one get an optimal $\widetilde{O}(n^{1/2})$ or even $\widetilde{O}(n)$ dependence? Second, can one get $\mathrm{poly}(s, \log T)$ bounds when the vectors are s-sparse. In particular, as a special case, can one get $(\log T)^{O(d)}$ bounds for the Tusnády problem, when points arrive from an *arbitrary* non-product distribution on $[0,1]^d$.

8.1.1 Our Results

In this chapter we resolve both the above questions of [BJSS20], and prove much more general results that obtain bounds within poly-logarithmic factors of those achievable in the offline setting.

Online Komlós and Tusnády settings. We first consider Komlós' setting for online discrepancy minimization where the vectors have ℓ_2-norm at most 1. Recall, the best known offline bound in this setting is $O((\log T)^{1/2})$ [Ban12]. We achieve the same result, up to poly-logarithmic factors, in the online setting.

Theorem 8.1.1 (Online Komlós setting)**.** *Let p be a distribution in $\mathbb{R}^n$ supported on vectors with Euclidean norm at most 1. Then, for vectors $v_1, \ldots, v_T$ sampled i.i.d. from p, there is an online algorithm that with high probability maintains a discrepancy vector d_t such that $\|d_t\|_\infty = O(\log^4(nT))$ for all $t \in [T]$.*

In particular, for vectors in $[-1,1]^n$ this gives an $O(n^{1/2}\log^4(nT))$ bound, and for s-sparse vectors in $[-1,1]^n$, this gives an $O(s^{1/2}\log^4(nT))$ bound, both of which are optimal up to poly-logarithmic factors.

The above result implies significant savings for the online Tusnády problem. Call a set $B \subseteq [0,1]^n$ an axis-parallel box if $B = I_1 \times \cdots \times I_n$ for intervals $I_i \subseteq [0,1]$. In the online Tusnády problem, we see points $x_1, \ldots, x_T \in [0,1]^d$ and need to assign signs $\chi_1, \ldots, \chi_T$ in an online manner to minimize the discrepancy of every axis-parallel box at all times. More

precisely, for an axis-parallel box B, define[4]

$$\mathsf{disc}_t(B) := \left|\chi_1 \mathbf{1}_B(x_1) + \ldots + \chi_t \mathbf{1}_B(x_t)\right|.$$

Our goal is to assign the signs $\chi_1, \ldots, \chi_t$ so as to minimize $\max_{t \le T} \mathsf{disc}_t(B)$ for every axis-parallel box B.

There is a standard reduction (see Section 8.5.2) from the online Tusnády problem to the case of s-sparse vectors in $\mathbb{R}^N$ where $s = (\log T)^d$ but the ambient dimension N is $O_d(T^d)$. Using this reduction, along with Theorem 8.1.1, directly gives an $O(\log^{3d/2+4} T)$ bound for the online Tusnády's problem that works for any *arbitrary* distribution on points, instead of just product distributions as in [BJSS20]. In fact, we prove a more general result where we can choose arbitrary directions to test discrepancy and we use this flexibility (see Theorem 8.1.3 below) to improve the exponent of the bound further, and essentially match the best offline bound of $O((\log^{d-1/2} T)$ [Nik17].

Theorem 8.1.2 (Online Tusnády's problem for arbitrary p)**.** *Let p be an arbitrary distribution on $[0,1]^d$. For points $x_1, \ldots, x_T$ sampled i.i.d from p, there is an algorithm which selects signs $\chi_t \in \{\pm 1\}$ such that with high probability for every axis-parallel box B, we have $\max_{t \in [T]} \mathsf{disc}_t(B) = O_d(\log^{d+4} T)$.*

Theorem 8.1.1 and Theorem 8.1.2 follow from the more general result below.

Theorem 8.1.3 (Discrepancy for Arbitrary Test Directions)**.** *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a finite set of test vectors with Euclidean norm at most 1 and p be a distribution in $\mathbb{R}^n$ supported on vectors with Euclidean norm at most 1. Then, for vectors $v_1, \ldots, v_T$ sampled i.i.d. from p, there is an online algorithm that with high probability maintains a discrepancy vector d_t satisfying*

$$\max_{z \in \mathcal{S}} |d_t^\top z| = O((\log(|\mathcal{S}|) + \log T) \cdot \log^3(nT)) \quad \textit{for every } t \in [T].$$

[4]Here, and henceforth, for a set S, denote $\mathbf{1}_S(x)$ the indicator function that is 1 if $x \in S$ and 0 otherwise.

In fact, the proof of the above theorem also shows that given any arbitrary distribution on unit test vectors z, one can maintain a bound on the exponential moment $\mathbb{E}_z[\exp(|\langle d_t, z\rangle|)]$ at all times.

The key idea involved in proving Theorem 8.1.3 above, is a novel potential function approach. In addition to controlling the discrepancy d_t in the test directions, we also control how the distribution of d_t relates to the input vector distribution p. This leads to better anti-concentration properties, which in turn gives better bounds on discrepancy in the test directions. We describe this idea in more detail in Sections 8.1.2 and 8.2.

Online Banaszczyk setting. Next, we consider discrepancy with respect to general norms given by an arbitrary convex body K. To recall, in the offline setting, Banaszczyk's seminal result [Ban12] shows that if K is any convex body with Gaussian measure $1 - 1/(2T)$, then for any vectors $v_1, \ldots, v_T$ of ℓ_2-norm at most 1, there exist signs $\chi_1, \ldots, \chi_T$ such that the discrepancy vectors $d_t \in O(1) \cdot K$ for all $t \in T$.

Here we study the online version when the input distribution $\mathsf{p} \in \mathbb{R}^n$ has sufficiently good tails. Specifically, we say a univariate random variable X has *sub-exponential tails* if for all $r > 0$, $\mathbb{P}\big[|X - \mathbb{E}[X]| > r\sigma(X)\big] \leq e^{-\Omega(r)}$, where $\sigma(X)$ denotes the standard-deviation of X. We say a multi-variate distribution $\mathsf{p} \in \mathbb{R}^n$ has sub-exponential tails if all its one-dimensional projections have sub-exponential tails. That is,

$$\mathbb{P}_{v\sim p}\left[\Big|\langle v, \theta\rangle - \mu_\theta]\Big| \geq \sigma_\theta \cdot r\right] \leq e^{-\Omega(r)} \quad \text{for every } \theta \in \mathbb{S}^{n-1} \text{ and every } r > 0,$$

where μ_θ and σ_θ are the mean and standard deviation[5] of the scalar random variable $X_\theta = \langle v, \theta\rangle$.

Many natural distributions, such as when v is chosen uniform over the vertices of the $\{\pm 1\}^n$ hypercube (scaled to have Euclidean norm one), uniform from a convex body, Gaussian distribution (scaled to have bounded norm with high probability), or uniform on the

[5]Note that when the input distribution p is α-isotropic, *i.e.* the covariance is αI_n, then $\sigma_\theta = \alpha$ for every direction θ, but the above definition is a natural generalization to handle an arbitrary covariance structure.

unit sphere, have a sub-exponential tail and in these cases our bounds match the offline bounds up to poly-logarithmic factors.

Theorem 8.1.4 (Online Banaszczyk Setting)**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body with $\gamma_n(K) \geq 1/2$ and p be a distribution with sub-exponential tails that is supported over vectors of Euclidean norm at most 1. Then, for vectors $v_1, \ldots, v_T$ sampled i.i.d. from p, there is an online algorithm that with high probability maintains a discrepancy vector d_t satisfying $d_t \in C \log^5(nT) \cdot K$ for all $t \in [T]$ and a universal constant C.*

The proof of the above theorem, while similar in spirit to Theorem 8.1.3, is much more delicate. In particular, we cannot use that theorem directly as capturing a general convex body as a polytope may require exponential number of constraints (the set $\mathcal{S}$ of test vectors).

Online Weighted Multi-Color Discrepancy. Finally we consider the setting of weighted multi-color discrepancy, where we are given vectors $v_1, \ldots, v_T \in \mathbb{R}^n$ sampled i.i.d. from a distribution p on vectors with ℓ_2-norm at most one, an integer R which is the number of colors available, positive weights $w_c \in [1, \eta]$ for each color $c \in [R]$, and a norm $\|\cdot\|_*$. At each time t, the algorithm has to choose a color $c \in [R]$ for the arriving vector, so that the discrepancy disc_t with respect to $\|\cdot\|_*$, defined below, is minimized for every $t \in [T]$:

$$\mathsf{disc}_t(\|\cdot\|_*) := \max_{c \neq c'} \mathsf{disc}_t(c, c') \quad \text{where} \quad \mathsf{disc}_t(c, c') := \left\| \frac{d_c(t)/w_c - d_{c'}(t)/w_{c'}}{1/w_c + 1/w_{c'}} \right\|_*,$$

with $d_c(t)$ being the sum of all the vectors that have been given the color c till time t. We note that (up to a factor of two) the case of unit weights and $R = 2$ is the same as assigning $\pm$ signs to the vectors $(v_i)_{i \leq T}$, and we will also refer to this setting as *signed discrepancy*.

We show that the bounds from the previous results also extend to the setting of multi-color discrepancy.

Theorem 8.1.5 (Weighted multi-color discrepancy)**.** *For any input distribution p and any set $\mathcal{S}$ of $\mathsf{poly}(nT)$ test vectors with Euclidean norm at most one, there is an online algorithm for the weighted multi-color discrepancy problem that maintains discrepancy $O(\log^2(R\eta) \cdot$*

$\log^4(nT))$ *with the norm* $\|\cdot\|_* = \max_{z\in\mathcal{S}} |\langle \cdot, z\rangle|$.

Further, if the input distribution p *has sub-exponential tails then one can maintain multi-color discrepancy* $O(\log^2(R\eta)\cdot\log^5(nT))$ *for any norm* $\|\cdot\|_*$ *given by a symmetric convex body* K *satisfying* $\gamma_n(K)\geq 1/2$.

As an application, the above theorem implies upper bounds for multi-player envy minimization in the online stochastic setting, as defined in [BKPP18], by reductions similar to those in [JKS19] and [BJSS20].

We remark that in the offline setting, such a statement with logarithmic dependence in R and η is easy to prove by identifying the various colors with leaves of a binary tree and recursively using the offline algorithm for signed discrepancy. It is not clear how to generalize such a strategy to the online stochastic setting, since the algorithm for signed discrepancy might use the stochasticity of the inputs quite strongly.

By exploiting the idea of working with the Haar basis, we show how to implement such a strategy in the online stochastic setting: we prove that if there is a greedy strategy for the signed discrepancy setting that uses a potential satisfying certain requirements, then it can be converted to the weighted multi-color discrepancy setting in a black-box manner.

8.1.2 High-Level Approach

Before describing our ideas, it is useful to discuss the bottlenecks in the previous approach. In particular, the quantitative bounds for the online Komlós problem, as well as for the case of sparse vectors obtained in [BJSS20] are the best possible using their approach, and improving them further required new ideas. We describe these ideas at a high-level here, and refer to Section 8.2 for a more technical overview.

Limitations of previous approach. For intuition, let us first consider the simpler setting, where we care about minimizing the Euclidean norm of the discrepancy vector d_t — this will already highlight the main issues. As mentioned before, if the adversary is adaptive in the online setting, then they can always choose the next input vector v_t to be orthogonal to d_{t-1}

(i.e., $\langle d_{t-1}, v_t \rangle = 0$) causing $\|d_t\|_2$ to grow as $T^{1/2}$. However, if $\langle d_{t-1}, v_t \rangle$ is typically large, then one can reduce $\|d_t\|_2$ by choosing $\chi_t = -\mathsf{sign}(\langle d_{t-1}, v_t \rangle)$, as the following shows:

$$\|d_t\|_2^2 - \|d_{t-1}\|_2^2 \;=\; 2\chi_t \cdot \langle d_{t-1}, v_t \rangle + \|v_t\|_2^2 \;\leq\; -2|\langle d_{t-1}, v_t \rangle| + 1. \tag{8.1}$$

The key idea in [BJSS20] was that if the vector v_t has uncorrelated coordinates (i.e., $\mathbb{E}_{v_t \sim \mathsf{p}}[v_t(i)v_t(j)] = 0$ for $i \neq j$), then one can exploit *anti-concentration* properties to essentially argue that $|\langle d_{t-1}, v_t \rangle|$ is typically large when $\|d_{t-1}\|_2$ is somewhat big, and the greedy choice above works, as it gives a *negative drift* for the ℓ_2-norm. However, uncorrelated vectors satisfy provably weaker anti-concentration properties, by up to a $n^{1/2}$ factor ($s^{1/2}$ for s-sparse vectors), compared to those with independent coordinates. This leads up to an extra $n^{1/2}$ loss in general.

Moreover, to ensure uncorrelation one has to work in the eigenbasis of the covariance matrix of p, which could destroy sparsity in the input vectors and give bounds that scale polynomially with n. [BJSS20] also show that one can combine the above high-level uncorrelation idea with a potential function that tracks a soft version of maximum discrepancy in any coordinate,

$$\Phi_{t-1} = \sum_{i=1}^{n} \exp(\lambda d_{t-1}(i)), \tag{8.2}$$

to even get bounds on the ℓ_∞-norm of d_t. However, this is also problematic as it might lead to another factor n loss, due to a change of basis (twice).

To achieve sparsity based bounds in the special case of online Tusnády's problem, previous approaches use the above ideas and exploit the special problem structure. In particular, when the input distribution p is a product distribution, [BJSS20] (and [DFGGR19]) observe that one can work with the natural Haar basis which also has a product structure in $[0,1]^d$ — this makes the input vectors uncorrelated, while simultaneously preserving the sparsity due to the recursive structure of the Haar basis. However, this severely restricts p to product

distributions and previously, it was unclear how to even handle a mixture of two product distributions.

New potential: anti-concentration from exponential moments. Our results are based on a new potential. Typical potential analyses for online problems show that no matter what the current state is, the potential does not rise much when the next input arrives. As discussed above, this is typically exploited in the online discrepancy setting using *anti-concentration* properties of the incoming vector $v_t \sim \mathsf{p}$ — one argues that no matter the current discrepancy vector d_{t-1}, the inner product $\langle d_{t-1}, v_t \rangle$ is typically large so that a sign can be chosen to decrease the potential (recall (8.1)).

However, as in [BJSS20], such a worst-case analysis is restrictive as it requires p to have additional desirable properties such as uncorrelated coordinates. A key conceptual idea in our work is that instead of just controlling a suitable proxy for the norm of the discrepancy vectors d_t, we also seek to control structural properties of the distribution d_t. Specifically, we also seek to evolve the distribution of d_t so that it has better anti-concentration properties with respect to the input distribution. In particular, one can get much better anti-concentration for a random variable if one also has control on the higher moments. For instance, if we can bound the fourth moment of the random variable $Y_t \equiv \langle d_{t-1}, v_t \rangle$, in terms of its variance, say $\mathbb{E}[Y_t^4] \ll \mathbb{E}[Y_t^2]^2$, then the Paley-Zygmund inequality implies that Y_t is far from zero. However, working with $\mathbb{E}[Y_t^4]$ itself is too weak as an invariant and necessitates looking at even higher moments.

A key idea is that these hurdles can be handled cleanly by looking at another potential that controls the *exponential moment* of Y_t. Specifically, all our results are based on an aggregate potential function based on combining a potential of the form (8.2), which enforces *discrepancy constraints*, together with variants of the following potential, for a suitable parameter λ, which enforces *anti-concentration constraints*:

$$\Phi_t \sim \mathbb{E}_v[\exp(\lambda|\langle d_t, v\rangle|)].$$

This clearly allows us to control higher moments of $\langle d_t, v\rangle$, in turn allowing us to show strong anti-concentration properties without any assumptions on p. We believe the above idea of controlling the space of possible states where the algorithm can be present in, could potentially be useful for other applications.

To illustrate the idea in the concrete setting of ℓ_2-discrepancy, let us consider the case when the input distribution p is mean-zero and $1/n$-*isotropic*, meaning the covariance $\mathbf{\Sigma} = \mathbb{E}_{v\sim\mathsf{p}}[vv^\top] = I_n/n$. Here, if we knew that the exponential moment $\mathbb{E}_{v\sim\mathsf{p}}[\exp(|\langle d_{t-1}, v\rangle|)] \le T$, then it implies that with high probability $|\langle d_{t-1}, v\rangle| \le \log T$ for $v \sim \mathsf{p}$. To avoid technicalities, let us assume that $|\langle d_{t-1}, v\rangle| \le \log T$ holds with probability one. Therefore, when v_t sampled independently from p arrives, then since $\mathbb{E}\big[|AB|\big] \ge \mathbb{E}[AB]/\|B\|_\infty$ for any coupled random variables A and B, taking $A = \langle d_{t-1}, v_t\rangle$ and $B = \langle d_{t-1}, v_t\rangle/\log T$, we get that

$$\mathbb{E}[|\langle d_{t-1}, v_t\rangle|] \;\ge\; \frac{1}{\log T}\cdot \mathbb{E}_{v_t}[d_{t-1}^\top v_t v_t^\top d_{t-1}] \;=\; \frac{1}{\log T}\cdot d_{t-1}^\top \mathbf{\Sigma} d_{t-1} \;=\; \frac{\|d_{t-1}\|_2^2}{n\log T}.$$

Therefore, whenever $\|d_{t-1}\|_2 \gg (n\log T)^{1/2}$, then the drift in ℓ_2-norm of the discrepancy vector d_t is negative. Thus, we can obtain the optimal ℓ_2-discrepancy bound of $O((n\log T)^{1/2})$.

Banaszczyk setting. In the Banaszczyk setting, the algorithm uses a carefully chosen set of test vectors at different scales that come from *generic chaining*. In particular, we use a potential function based on test vectors derived from the generic chaining decomposition of the polar K° of the body K.

However, as there can now be exponentially many such test vectors, more care is needed. First, we use that the Gaussian measure of K is large to control the number of test vectors at each scale in the generic chaining decomposition of K°. Second, to be able to perform a union bound over the test vectors at each scale, one needs substantially stronger tail bounds than in Theorem 8.1.3. To do this, we scale the test vectors to be quite large, but this becomes problematic with standard tools for potential analysis, such as Taylor approximation, as the

update to each term in the potential can be much larger than potential itself, and hard to control. Nevertheless, we show that if the distribution has sub-exponential tails, then such an approximation holds "on average" and the growth in the potential can be bounded.

8.2 Proof Overview

Recall the setting: the input vectors $(v_\tau)_{\tau \le T}$ are sampled i.i.d. from p and satisfy $\|v\|_2 \le 1$, and we need to assign signs $\chi_1, \ldots, \chi_T$ in an online manner so as to minimize some target norm of the discrepancy vectors $d_t = \sum_{\tau \le t} \chi_\tau v_\tau$. Moreover, we may also assume, without loss of generality that the distribution is mean-zero as the algorithm can toss a coin and work with either v or $-v$. This means that the covariance matrix $\mathbf{\Sigma} = \mathbb{E}_v[vv^\top]$ satisfying $0 \preccurlyeq \mathbf{\Sigma} \preccurlyeq I_n$.

8.2.1 Komlos Setting

Here our goal is to minimize $\|d_t\|_\infty$. First, consider the potential function $\mathbb{E}_{v \sim \mathsf{p}}[\cosh(\lambda\, d_t^\top v)]$ where $\cosh(a) = \frac{1}{2} \cdot (e^a + e^{-a})$. This however only puts anti-concentration constraints on the discrepancy vector and does not track the discrepancy in the coordinate directions. It is natural to add a potential term to enforce discrepancy constraints. In particular, let $\mathsf{p}_x = \frac{1}{2}\mathsf{p} + \frac{1}{2}\mathsf{p}_y$, where p_y is uniform over the standard basis vectors $(e_i)_{i \le n}$, then the potential

$$\Phi_t = \mathbb{E}_{x \sim \mathsf{p}_x}[\cosh(\lambda\, d_t^\top x)], \tag{8.3}$$

allows us to control the exponential moments of $\langle d_{t-1}, v_t \rangle$ as well as the discrepancy in the target test directions. In particular, if the above potential $\Phi_t \le \mathsf{poly}(T)$, then we get a bound of $O(\lambda^{-1} \log T)$ on $\|d_t\|_\infty$. Next we sketch a proof that for the greedy strategy using the above potential, one can take $\lambda = 1/\log T$, so that the potential remains bounded by $\mathsf{poly}(T)$ at all times.

Claim 8.2.1 (Informal: Bounded Drift)**.** *If* $\Phi_{t-1} \le T^2$*, then* $\mathbb{E}_{v_t}[\Delta\Phi_t] := \mathbb{E}_{v_t}[\Phi_t - \Phi_{t-1}] \le 2$.

The above implies using standard martingale arguments, that the potential remain bounded by T^2 with high probability and hence $\|d_t\|_\infty = \mathsf{polylog}(T)$ at all times $t \in [T]$.

To continue, let us first make a simplifying assumption that $\mathbf{\Sigma} = I_n/n$ and that at time t, the condition $\lambda|d_{t-1}^\top v_t| \le 2\log T$ holds with probability 1. We give an almost complete proof below under these conditions. The first condition can be dealt with by an appropriate decomposition of the covariance matrix as sketched below. The second condition only holds with high probability $(1 - 1/\mathsf{poly}(T))$, because we have a bound on the exponential moment, but the error event can be handled straightforwardly.

By Taylor expansion, we have that for all a,

$$\cosh(\lambda(a+\delta)) - \cosh(\lambda a) \;\le\; \lambda \sinh(\lambda a)\cdot\delta + \lambda^2|\sinh(\lambda a)|\cdot\delta^2 + \lambda^2 \qquad \text{for all } |\delta| \le 1, \tag{8.4}$$

where $\sinh(a) = \frac{1}{2}\cdot(e^a - e^{-a})$ and we used that $\cosh(a) \le |\sinh(a)| + 1$. Therefore, since $d_t = d_{t-1} + \chi_t v_t$, by the above inequality we have

$$\begin{aligned}\Delta\Phi_t &\le \chi_t \cdot \lambda\mathbb{E}_x\left[\sinh(\lambda d_{t-1}^\top x)\cdot x^\top v_t\right] + \lambda^2\mathbb{E}_x\left[|\sinh(\lambda d_{t-1}^\top x)|\cdot|x^\top v_t|^2\right]\\ &:= \chi_t\lambda L + \lambda^2 Q + \lambda^2.\end{aligned}$$

Since the algorithm chooses χ_t to minimize the potential, we have that $\mathbb{E}_{v_t}[\Delta\Phi_t] \le -\lambda\mathbb{E}_{v_t}[|L|] + \lambda^2\mathbb{E}_{v_t}[Q]$.

Upper bounding the quadratic term. Using that $\mathbf{\Sigma} = \mathbb{E}_{v_t}[v_t v_t^\top] = I_n/n$, we have

$$\begin{aligned}\mathbb{E}_{v_t}[Q] &= \mathbb{E}_{v_t,x}[|\sinh(\lambda d_{t-1}^\top x)|\cdot x^T v_t v_t^\top x] = \mathbb{E}_x[|\sinh(\lambda d_{t-1}^\top x)|\cdot x^T\mathbf{\Sigma}x]\\ &= \frac{1}{n}\cdot\mathbb{E}_x[|\sinh(\lambda d_{t-1}^\top x)|\cdot\|x\|^2] \le \frac{1}{n}\cdot\mathbb{E}_x[|\sinh(\lambda d_{t-1}^\top x)|],\end{aligned}$$

where the last inequality used that $\|x\|_2 \le 1$.

Lower bounding the linear term. For this we use the aforementioned coupling trick: $\mathbb{E}_{v_t}[|L|] \geq \mathbb{E}_{v_t}[LY]/\|Y\|_\infty$ for any coupled random variable Y [6]. Taking $Y = |d_{t-1}^\top v_t|$, we have that $\|Y\|_\infty \leq 2\lambda^{-1}\log T$. Therefore,

$$\begin{aligned}\mathbb{E}_{v_t}[|L|] &= \mathbb{E}_{v_t}\left|\mathbb{E}_x\left[\sinh(\lambda d_{t-1}^\top x)\cdot x^\top v_t\right]\right| \geq \frac{\lambda}{2\log T}\cdot\mathbb{E}_{v_t,x}\left[\sinh(\lambda d_{t-1}^\top x)|\cdot x^\top v_t v_t^\top d_{t-1}\right]\\ &= \frac{1}{2n\log T}\cdot\mathbb{E}_x[\sinh(\lambda d_{t-1}^\top x)\cdot\lambda d_{t-1}^\top x] \geq \frac{1}{2n\log T}\cdot\mathbb{E}_x[|\sinh(\lambda d_{t-1}^\top x)|]-2,\end{aligned}$$

using that $\sinh(a)a \geq |\sinh(a)| - 2$ for all $a \in \mathbb{R}$.

Therefore, if $\lambda = 1/(2\log T)$, we can bound the drift in the potential

$$\mathbb{E}_{v_t}[\Delta\Phi_t] \leq -\frac{\lambda}{2n\log T}\cdot\mathbb{E}_x[|\sinh(\lambda d_{t-1}^\top x)|] + \frac{\lambda^2}{n}\cdot\mathbb{E}_x[|\sinh(\lambda d_{t-1}^\top x)|] + 2 + \lambda^2 \leq 3.$$

Non-Isotropic Covariance. To handle the general case when the covariance $\mathbf{\Sigma}$ is not isotropic, let us assume that all the non-zero eigenvalues are of the form 2^{-k} for integers $k \geq 0$. One can always rescale the input vectors and any potential set of test vectors, so that the covariance satisfies the above, while the discrepancy is affected only by a constant factor. See Section 8.4 for details.

With the above assumption $\mathbf{\Sigma} = \sum_k 2^{-k}\Pi_k$ where Π_k is the orthogonal projection on to the subspace with eigenvalues 2^{-k}. Since, we only get T vectors, we can ignore the eigenvalues smaller than $(nT)^{-4}$ and only need to consider $O(\log(nT))$ different scales. Then, one can work with the following potential which imposes the alignment constraint in each such subspace:

$$\Phi_t = \sum_k \mathbb{E}_{x\sim\mathsf{p}_x}[\cosh(\lambda\, d_t^\top \Pi_k x)].$$

As we have $O(\log(nT))$ pairwise orthogonal subspaces, we can still choose $\lambda = 1/\mathsf{polylog}(nT)$ and with some care, the drift can be bounded using the aforementioned ideas. Once the potential is bounded, we can bound $\|d_t\|_\infty$ as before along with triangle inequality.

---

[6] Here $\|Y\|_\infty$ denotes the largest value of Y in its support.

8.2.2 Banaszczyk Setting

Recall that here we are given a convex body K with Gaussian volume at least $1/2$ and our goal is to bound K-norm of the discrepancy vector $\|d_t\|_K$. Here, $\|d\|_K$ intuitively is the minimum scaling γ of K so that $d \in \gamma K$. To this end, we will use the dual characterization of K: Let $K^\circ = \{y : \sup_{x \in K} |\langle x, y\rangle| \leq 1\}$, then $\|d\|_K = \sup_{y \in K^\circ} |\langle d, y\rangle|$.

To approach this first note that the arguments from previous section allow us not only to bound $\|d_t\|_\infty$ but also $\max_{z \in \mathcal{S}} \langle d_t, z\rangle$ for an arbitrary set of *test directions* $\mathcal{S}$ (of norm at most 1). As long as $|\mathcal{S}| \leq \mathsf{poly}(nT)$, we can bound $\max_{z \in \mathcal{S}} \langle d_t, z\rangle = \mathsf{poly}(\log(nT))$.

However, to handle a norm given by an arbitrary convex body K, one needs exponentially many test vectors, and the previous ideas are not enough. To design a suitable test distribution for an arbitrary convex body K, we use *generic chaining* to bound $\|d_t\|_K = \sup_{z \in K^\circ} \langle d_t, z\rangle$ by choosing epsilon-nets[7] of K° at geometrically decreasing scales. Again let us assume that the $\mathbf{\Sigma} = I_n/n$ for simplicity.

First, assuming Gaussian measure of K is at least $1/2$, it follows that $\mathsf{diam}(K^\circ) = O(1)$ (see Section 8.3.3). So, one can choose the coarsest epsilon-net at $O(1)$-scale while the finest epsilon-net can be taken at scale $\approx 1/\sqrt{n}$ since by adding the standard basis vectors to the test set, one can control $\|d_t\|_2 \leq \sqrt{n}$ (ignoring polylog factors) by using the previous ideas in the Komlös setting.

Now, one can use generic chaining as follows: define the directed layered graph $\mathcal{G}$ (see Figure 8.1) where the vertices $\mathcal{T}_\ell$ in layer ℓ are the elements of an optimal ε_ℓ-net of K° with $\varepsilon_\ell = 2^{-\ell}$. We add a directed edge from a vertex $u \in \mathcal{T}_\ell$ to vertex $v \in \mathcal{T}_{\ell+1}$ if $\|u - v\|_2 \leq \varepsilon_\ell$ and identify the corresponding edge with the vector $v - u$. The length of any such edge $v - u$, defined as $\|v - u\|_2$, is at most ε_ℓ.

Let us denote the set of edges between layer ℓ and $\ell + 1$ by $\mathcal{S}_\ell$. Now, one can express any $z \in K^\circ$ as $\sum_\ell w_\ell + w_{\mathsf{err}}$ where $w_\ell \in \mathcal{S}_\ell$ and $\|w_{\mathsf{err}}\|_2 \leq 1/\sqrt{n}$. Then, since we can control

[7] We remark that one can also work with admissible nets that come from Talagrand's majorizing measures theorem and probably save a logarithmic factor, but for simplicity we work with epsilon-nets at different scales.

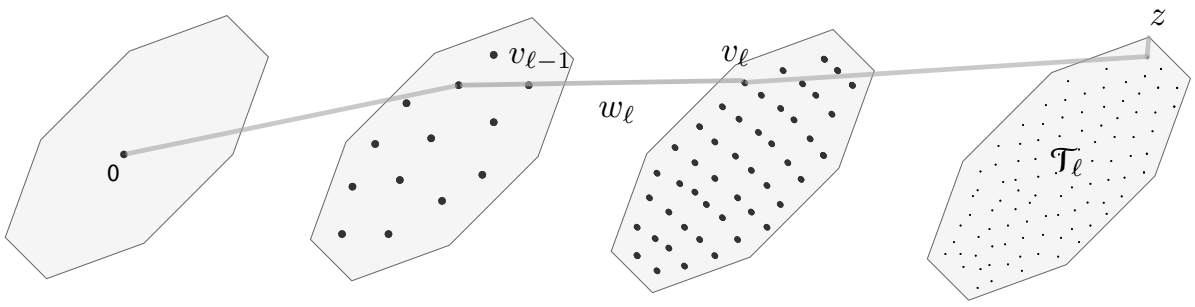


Figure 8.1: The chaining graph $\mathcal{G}$ showing epsilon-nets of the convex body at various scales. The edges connect near neighbors at two consecutive scales. Note that any point $z \in K^\circ$ can be expressed as the sum of the edge vectors w_ℓ where $w_\ell = v_\ell - v_{\ell-1}$, and $(v_{\ell-1}, v_\ell)$ is an edge between two points at scale $2^{-(\ell-1)}$ and $2^{-\ell}$.

$\|d_t\|_2 \le \sqrt{n}$, we have

$$\sup_{z \in K^\circ} \langle d_t, z \rangle \le \sum_\ell \max_{w \in \mathcal{S}_\ell} \langle d, w \rangle + \max_{\|w\|_2 \le n^{-1/2}} \langle d, w_{\mathsf{err}} \rangle = O(\log n) \cdot \max_\ell \max_{w \in \mathcal{S}_\ell} \langle d, w \rangle.$$

Thus, it suffices to control $\max_{w \in \mathcal{S}_\ell} \langle d, w \rangle$ for each scale using a suitable test distribution in the potential.

For example, suppose we knew that $\mathbb{E}_{\widetilde{w}}[\cosh(\lambda d^\top \widetilde{w})] \le T$ for $\widetilde{w}$ uniform in $r^2 \cdot \mathcal{S}_\ell$ for a scaling factor r^2. Then, it would follow that $\max_{w \in \mathcal{S}_\ell} \langle d, w \rangle = O(\lambda^{-1} r^{-2} \log |\mathcal{S}_\ell| \cdot \log T)$. Standard results in convex geometry (see Section 8.3.3) imply that $|\mathcal{S}_\ell| \le e^{O(1/\varepsilon_\ell^2)}$, so to obtain a $\mathsf{polylog}(nT)$ bound, one needs to scale the vectors $w \in \mathcal{S}_\ell$ by a factor of $r = 1/\varepsilon_\ell$. This implies that the ℓ_2-norm of scaled vector $r^2 \cdot w$ could be as large as $\sqrt{n}$.

This makes the drift analysis for the potential more challenging because now the Taylor expansion in (8.4) is not always valid as the update δ could be as large as $\sqrt{n}$. This is where the sub-exponential tail of the input distribution is useful for us. Since the input distribution is $1/n$-isotropic and sub-exponential tailed, we know that if $\|w\|_2 \le \sqrt{n}$, then for a typical

choice of $v \sim \mathsf{p}$, the following holds

$$\langle v_t, w\rangle \approx \mathbb{E}_{v_t}[\langle v_t, w\rangle^2] = \mathbb{E}_{v_t}[w^\top v_t v_t^\top w] = \frac{\|w\|_2^2}{n} \leq 1.$$

Thus, with some work one can show that, the previous Taylor expansion essentially holds "on average" and the drift can be bounded. The case of general covariances can be handled by doing a decomposition as before. Although the full analysis becomes somewhat technical, all the main ideas are presented above.

8.2.3 Multi-color Discrepancy

For the multi-color discrepancy setting, we show that if there is an online algorithm that uses a greedy strategy with respect to a certain kind of potential Φ, then one can adapt the same potential to the multi-color setting in a black-box manner.

In particular, let the number of colors $R = 2^h$ for an integer h and all weights be unit. Let us identify the leaves of a complete binary tree $\mathcal{T}$ of height h with a color. Our goal is then to assign the incoming vector to one of the leaves. In the offline setting, this is easy to do with a logarithmic dependence of R — we start at the root and use the algorithm for the signed discrepancy setting to decide to which sub-tree the vector be assigned and then we recurse until the vector is assigned to one of the leaves. Such a strategy in the online stochastic setting is not obvious, as the distribution of the incoming vector might change as one decides which sub-tree it belongs to.

By exploiting the idea used in [BJSS20] and [DFGGR19] of working with the Haar basis, we can implement such a strategy if the potential Φ satisfies certain requirements. Let us define $d_\ell(t)$ to be the sum of all the input vectors assigned to that leaf at time t. In the same way, for an internal node u of $\mathcal{T}$, we can define $d_u(t)$ to be the sum of the vectors $d_\ell(t)$ for all the leaves ℓ in the sub-tree rooted at u. The crucial insight is then, one can track the difference of the discrepancy vectors of the two children $d_u^-(t)$ for every internal node u of

the tree $\mathcal{T}$. In particular, one can work with the potential

$$\Psi_t = \sum_{u \in \mathcal{T}} \Phi\big(\beta\, d_u^-(t)\big),$$

for some parameter β, and assign the incoming vector to the leaf that minimizes the increase in Ψ_t. Then, essentially we show that the analysis for the potential Φ translates to the setting of the potential Ψ_t if Φ satisfies certain requirements (see Section 8.7).

8.3 Preliminaries

8.3.1 Notation

Throughout this chapter, log denotes the natural logarithm unless the base is explicitly mentioned. We use $[k]$ to denote the set $\{1, 2, \ldots, k\}$. Sets will be denoted by script letters (e.g. $\mathcal{T}$).

Random variables are denoted by capital letters (e.g. A) and values they attain are denoted by lower-case letters possibly with subscripts and superscripts (e.g. a, a_1, a', etc.). Events in a probability space will be denoted by calligraphic letters (e.g. $\mathcal{E}$). We also use $\mathbf{1}_{\mathcal{E}}$ to denote the indicator random variable for the event $\mathcal{E}$. We write $\lambda \mathsf{p} + (1-\lambda)\mathsf{p}'$ to denote the convex combination of the two distributions.

Given a distribution p, we use the notation $x \sim \mathsf{p}$ to denote an element x sampled from the distribution p. For a real function f, we will write $\mathbb{E}_{x \sim \mathsf{p}}[f(x)]$ to denote the expected value of $f(x)$ under x sampled from p. If the distribution is clear from the context, then we will abbreviate the above as $\mathbb{E}_x[f(x)]$.

For a symmetric matrix M, we use M^+ to denote the Moore-Penrose pseudo-inverse, $\|M\|_{\mathsf{op}}$ for the operator norm of M and $\mathrm{Tr}(M)$ for the trace of M.

8.3.2 Sub-exponential Tails

Recall that a subexponential distribution p on $\mathbb{R}$ satisfies the following for every $r > 0$, $\mathbb{P}_{x\sim\mathsf{p}}[|x-\mu| \geq \sigma r] \leq e^{-\Omega(r)}$ where $\mu = \mathbb{E}_x[x]$ and $\sigma^2 = \mathbb{E}_x[(x-\mu)^2]$. A standard property of a distribution with a sub-exponential tail is *hypercontractivity* and a bound on the exponential moment (c.f. §2.7 in [Ver18]).

Proposition 8.3.1. *Let p be a distribution on $\mathbb{R}$ that has a sub-exponential tail with mean zero and variance σ^2. Then, for a constant $C > 0$, we have that $\mathbb{E}_{x\sim\mathsf{p}}[e^{s|x|}] \leq C$ for all $|s| \leq 1/2\sigma$. Moreover, for every $k > 0$, we have $\mathbb{E}_{x\sim\mathsf{p}}[|x|^k]^{1/k} \leq C \cdot k\sigma$.*

8.3.3 Convex Geometry

Given a convex body $K \subseteq \mathbb{R}^n$, its *polar* convex body is defined as $K^\circ = \{y \mid \sup_{x\in K} |\langle x, y\rangle| \leq 1\}$. If K is symmetric, then it defines a norm $\|\cdot\|_K$ which is defined as $\|\cdot\|_K = \sup_{y\in K^\circ} \langle \cdot, y\rangle$.

For a linear subspace $H \subseteq \mathbb{R}^n$, we have that $(K \cap H)^\circ = \Pi_H(K^\circ)$ where Π_H is the orthogonal projection on to the subspace H.

Gaussian Measure. We denote by γ_n the n-dimensional standard Gaussian measure on $\mathbb{R}^n$. More precisely, for any measurable set $\mathcal{A} \subseteq \mathbb{R}^n$, we have

$$\gamma_n(\mathcal{A}) = \frac{1}{(\sqrt{2\pi})^n} \int_{\mathcal{A}} e^{-\|x\|_2^2/2} dx.$$

For a k-dimensional linear subspace H of $\mathbb{R}^n$ and a set $\mathcal{A} \subseteq H$, we denote by $\gamma_k(\mathcal{A})$ the Gaussian measure of the set $\mathcal{A}$ where H is taken to be the whole space. For convenience, we will sometimes write $\gamma_H(\mathcal{A})$ to denote $\gamma_{\dim(H)}(\mathcal{A} \cap H)$.

The following is a standard inequality for the Gaussian measure of slices of a convex body. For a proof, see Lemma 14 in [DGLN16].

Proposition 8.3.2. *Let $K \subseteq \mathbb{R}^n$ with $\gamma_n(K) \geq 1/2$ and $H \subseteq \mathbb{R}^n$ be a linear subspace of dimension k. Then, $\gamma_k(K \cap H) \geq \gamma_n(K)$.*

Gaussian Width. For a set $\mathcal{T} \subseteq \mathbb{R}^n$, let $w(\mathcal{T}) = \mathbb{E}_g[\sup_{x\in\mathcal{T}}\langle g, x\rangle]$ denote the *Gaussian width* of $\mathcal{T}$ where $g \in \mathbb{R}^n$ is sampled from the standard normal distribution. Let $\mathsf{diam}(\mathcal{T}) = \sup_{x,y\in\mathcal{T}} \|x - y\|_2$ denote the diameter of the set $\mathcal{T}$.

The following lemma is standard up to the exact constants. For a proof, see Lemmas 26 and 27 in [DGLN16].

Proposition 8.3.3. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body with $\gamma_n(K) \geq 1/2$. Then, $w(K^\circ) \leq \frac{3}{2}$ and $\mathsf{diam}(K^\circ) \leq 4$.*

To prevent confusion, we remark that the Gaussian width is $\Theta(\sqrt{n})$ factor larger than the *spherical width* defined as $\mathbb{E}_\theta[\sup_{x\in\mathcal{T}}\langle\theta, x\rangle]$ for a randomly chosen θ from the unit sphere $\mathbb{S}^{n-1}$. So the above proposition implies that the spherical width of K° is $O(1/\sqrt{n})$.

For a linear subspace $H \subseteq \mathbb{R}^n$ and a subset $\mathcal{T} \subseteq H$, we will use the notation $w_H(\mathcal{T}) = \mathbb{E}_g[\sup_{x\in\mathcal{T}}\langle g, x\rangle]$ to denote the Gaussian width of $\mathcal{T}$ in the subspace H, where g is sampled from the standard normal distribution on the subspace H. Proposition 8.3.2 and Proposition 8.3.3 also imply that $w_H(\mathcal{T}) \leq 3/2$.

Covering Numbers. For a set $\mathcal{T} \subseteq \mathbb{R}^n$, let $N(\mathcal{T}, \varepsilon)$ denote the size of the smallest ε-net of $\mathcal{T}$ in the Euclidean metric, *i.e.*, the smallest number of closed Euclidean balls of radius ε whose union covers $\mathcal{T}$. Then, we have the following inequality (c.f. [Wai19], §5.5).

Proposition 8.3.4 (Sudakov minoration)**.** *For any set $\mathcal{T} \subseteq \mathbb{R}^n$ and any $\varepsilon > 0$*

$$w(\mathcal{T}) \geq \frac{\varepsilon}{2}\sqrt{\log N(\mathcal{T}, \varepsilon)}, \quad \textit{or equivalently,} \quad N(\mathcal{T}, \varepsilon) \leq e^{4w(\mathcal{T})^2/\varepsilon^2}.$$

Analogously, for a linear subspace $H \subseteq \mathbb{R}^n$ and a subset $\mathcal{T} \subseteq H$, we also have $w_H(\mathcal{T}) \geq \frac{\varepsilon}{2}\sqrt{\log N_H(\mathcal{T}, \varepsilon)}$, where $N_H(\mathcal{T}, \varepsilon)$ denote the covering numbering of $\mathcal{T}$ when H is considered the whole space.

8.4 Reduction to Dyadic Covariance

For all our problems, we may assume without loss of generality that the distribution p has zero mean, i.e. $\mathbb{E}_{v\sim\mathsf{p}}[v] = 0$, since our algorithm can toss an unbiased random coin and work with either v or $-v$. Now the covariance matrix $\mathbf{\Sigma}$ of the input distribution p is given by $\mathbf{\Sigma} = \mathbb{E}_{v\sim\mathsf{p}}[vv^\top]$. Since $\|v\|_2 \leq 1$, we have that $0 \preccurlyeq \mathbf{\Sigma} \preccurlyeq I$ and $\mathrm{Tr}(\mathbf{\Sigma}) \leq 1$.

However, it will be more convenient for the proof to assume that all the non-zero eigenvalues of the covariance matrix $\mathbf{\Sigma}$ are of the form 2^{-k} for an integer k. In this section, by slightly rescaling the input distribution and the test vectors, we show that one can assume this without any loss of generality.

Consider the spectral decomposition of $\mathbf{\Sigma} = \sum_{i=1}^{n} \sigma_i u_i u_i^\top$, where $0 \leq \sigma_n \leq \ldots \leq \sigma_1 \leq 1$ and $u_1, \ldots, u_n$ form an orthonormal basis of $\mathbb{R}^n$. Moreover, since we only get T vectors, we can essentially ignore all eigenvalues smaller than, say $(nT)^{-8}$, as this error will not affect the discrepancy too much.

For a positive integer κ denoting the number of different scales, we say that $\mathbf{\Sigma}$ is κ-dyadic if every non-zero eigenvalue σ is 2^{-k} for some $k \in [\kappa]$.

Lemma 8.4.1. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be an arbitrary set of test vectors with Euclidean norm at most nT and $v \sim \mathsf{p}$ with covariance $\mathbf{\Sigma} = \sum_i \sigma_i u_i u_i^\top$. Then, there exists a positive-semi definite matrix M with $\|M\|_{\mathsf{op}} \leq 1$ such that the covariance of Mv is κ-dyadic for $\kappa = \lceil 8\log(nT)\rceil$. Moreover, there exists a test set $\mathcal{S}'$ consisting of vectors with Euclidean norm at most $\max_{y\in\mathcal{S}} \|y\|$, such that for any signs $(\chi_t)_{t\in T}$, the discrepancy vector $d_t = \sum_{\tau=1}^{t} \chi_\tau v_\tau$ satisfies the following with probability $1 - (nT)^{-4}$,*

$$\max_{y\in\mathcal{S}} |d_t^\top y| = 2 \cdot \max_{z\in\mathcal{S}'} |(Md_t)^\top z| + O(1).$$

Proof. For notational simplicity, we use d to denote d_t. We construct matrix M to be postive semi-definite with eigenvectors $u_1, \ldots, u_n$. For any $i \in [n]$ such that $\sigma_i \in (2^{-k}, 2^{-k+1}]$ for some $k \in [\kappa]$, we set $Mu_i = (2^k\sigma_i)^{-1/2} \cdot u_i$, and for every $i \in [n]$ such that $\sigma_i \leq 2^{-\kappa}$, we set

$Mu_i = 0$. It is easy to check that the covariance of Mv for $v \sim \mathsf{p}$ is κ-dyadic.

We define the new test set to be $\mathcal{S}' = \{\frac{1}{2}M^+ y \mid y \in \mathcal{S}\}$ where M^+ is the pseudo-inverse of M. Note that $\|M^+\|_{\mathsf{op}} \leq 2$, so every $z \in \mathcal{S}'$ satisfies $\|z\|_2 \leq \max_{y \in \mathcal{S}} \|y\| \leq nT$. To upper bound the discrepancy with respect to the test set, let Π_{err} be the projector onto the span of eigenvectors u_i with $\sigma_i \leq 2^{-\kappa}$ and let Π be the projector onto its orthogonal subspace. Then, for any $y \in \mathcal{S}$, we have

$$|d^\top y| \leq |d^\top \Pi y| + |d^\top \Pi_{\mathsf{err}} y| \leq |(Md)^\top (M^+ y)| + nT \cdot \|\Pi_{\mathsf{err}} d\|_2.$$

Note that $\mathbb{E}\|\Pi_{\mathsf{err}} d\|_2^2 \leq (nT)^{-8} \cdot (nT)^2$, so by Markov's inequality, with probability at least $1 - (nT)^{-4}$, we have that $\|\Pi_{\mathsf{err}} d\|_2 \leq (nT)^{-1}$ and hence, $|d^\top \Pi_{\mathsf{err}} y| = O(1)$ for every $y \in \mathcal{S}$. It follows that

$$\max_{y \in \mathcal{S}} |d^\top y| \leq 2 \cdot \max_{z \in \mathcal{S}'} |(Md)^\top z| + O(1).$$

□

For all applications in this chapter, the test vectors will always have Euclidean norm at most nT, so we can always assume without loss of generality that the input distribution p, which is supported over vectors with Euclidean norm at most one, has mean $\mathbb{E}_{v \sim \mathsf{p}}[v] = 0$, and its covariance $\mathbf{\Sigma} = \mathbb{E}_v[vv^\top]$ is κ-dyadic for $\kappa = 8\lceil \log(nT) \rceil$. We will make this assumption in the rest of this chapter without stating it explicitly sometimes.

8.5 Discrepancy for Arbitrary Test Vectors

In this section, we consider discrepancy minimization with respect to an arbitrary set of test vectors with Euclidean length at most 1.

Theorem 8.1.3 (Discrepancy for Arbitrary Test Directions)**.** *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a finite set of test vectors with Euclidean norm at most* 1 *and p be a distribution in $\mathbb{R}^n$ supported on vectors with Euclidean norm at most* 1*. Then, for vectors $v_1, \ldots, v_T$ sampled i.i.d. from p, there is*

an online algorithm that with high probability maintains a discrepancy vector d_t satisfying

$$\max_{z\in\mathcal{S}} |d_t^\top z| = O((\log(|\mathcal{S}|) + \log T)\cdot \log^3(nT)) \quad \textit{for every } t \in [T].$$

Before getting into the details of the proof, we first give two important applications of Theorem 8.1.3 to the Komlós problem in Section 8.5.1 and to the Tusnady's problem in Section 8.5.2. The proof of Theorem 8.1.3 will be discussed in Section 8.5.3.

8.5.1 Discrepancy for Online Komlós Setting

Theorem 8.1.1 (Online Komlós setting)**.** *Let p be a distribution in $\mathbb{R}^n$ supported on vectors with Euclidean norm at most 1. Then, for vectors $v_1, \ldots, v_T$ sampled i.i.d. from p, there is an online algorithm that with high probability maintains a discrepancy vector d_t such that $\|d_t\|_\infty = O(\log^4(nT))$ for all $t \in [T]$.*

Proof of Theorem 8.1.1. Taking the set of test vectors $\mathcal{S} = \{e_1, \cdots, e_n\}$ where e_i's are the standard basis vectors in $\mathbb{R}^n$, Theorem 8.1.3 implies an algorithm that w.h.p. maintains a discrepancy vector d_t such that $\|d_t\|_\infty = O(\log^4(nT))$ for all $t \in [T]$. □

8.5.2 An Application to Online Tusnady's Problem

Theorem 8.1.2 (Online Tusnády's problem for arbitrary p)**.** *Let p be an arbitrary distribution on $[0,1]^d$. For points $x_1, \ldots, x_T$ sampled i.i.d from p, there is an algorithm which selects signs $\chi_t \in \{\pm 1\}$ such that with high probability for every axis-parallel box B, we have $\max_{t\in[T]} \mathsf{disc}_t(B) = O_d(\log^{d+4} T)$.*

Firstly, using the probability integral transformation along each dimension, we may assume without loss of generality that the marginal of p along each dimension $i \in [d]$, denoted as p_i, is the uniform distribution on $[0,1]$. More specifically, we replace each incoming point $x \in [0,1]^d$ by $(F_1(x_1), \cdots, F_d(x_d))$, where F_i is the cumulative density function for p_i. Note

that $F_i(x_i)$ is uniform on $[0,1]$ when $x_i \sim \mathsf{p}_i$. We make such an assumption throughout this subsection.

A standard approach in tackling Tusnády's problem is to decompose the unit cube $[0,1]^d$ into a canonical set of boxes known as dyadic boxes (see [Mat99]). Define dyadic intervals $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ for $j \in \mathbb{Z}_{\geq 0}$ and $0 \leq k < 2^j$. A dyadic box is one of the form

$$B_{\boldsymbol{j},\boldsymbol{k}} := I_{\boldsymbol{j}(1),\boldsymbol{k}(1)} \times \ldots \times I_{\boldsymbol{j}(d),\boldsymbol{k}(d)},$$

with $\boldsymbol{j}, \boldsymbol{k} \in \mathbb{Z}^d$ such that $0 \leq \boldsymbol{j}$ and $0 \leq \boldsymbol{k} < 2^{\boldsymbol{j}}$, and each side has length at least $1/T$. One can handle the error from the smaller dyadic boxes separately since few points will land in each such box. Denoting the set of dyadic boxes as $\mathcal{D} = \{B_{\boldsymbol{j},\boldsymbol{k}} \mid 0 \leq \boldsymbol{j} \leq (\log T)\mathbf{1}\ ,\ 0 \leq \boldsymbol{k} < 2^{\boldsymbol{j}}\}$, where $\mathbf{1} \in \mathbb{R}^d$ is the all ones vector, we note that $|\mathcal{D}| = O_d(T^d)$.

Usually, one proves a discrepancy upper bound on the set of dyadic boxes, which implies a discrepancy upper bound on all axis-parallel boxes since each axis-parallel box can be expressed roughly as the disjoint union of $O_d(\log^d T)$ dyadic boxes. This was precisely the approach used for the online Tusnády's problem in [BJSS20]. However, such an argument has a fundamental barrier. Since each arrival lands in approximately $O_d(\log^d T)$ boxes in $\mathcal{D}$, one can at best obtain a discrepancy upper bound of $O_d(\log^{d/2} T)$ for the set of dyadic boxes, which leads to $O_d(\log^{3d/2} T)$ discrepancy for all boxes.

Using the idea of test vectors in Theorem 8.1.3, we can save a factor of $O_d(\log^{d/2} T)$ over the approach above. Roughly, this saving comes from the discrepancy of dyadic boxes accumulates in an ℓ_2 manner as opposed to directly adding up. A similar idea was previously exploited by [BG17] for the offline Tusnády's problem.

Proof of Theorem 8.1.2. We view Tusnády's problem as a vector balancing problem in $|\mathcal{D}|$-dimensions with coordinates indexed by dyadic boxes, where we define $v_t(B) = \mathbf{1}_B(x_t)$ for each arrival $t \in [T]$ and every dyadic box $B \in \mathcal{D}$. Each coordinate B of the discrepancy

vector $d_t = \sum_{i=1}^{t} \chi_i v_i$ is exactly $\mathsf{disc}_t(B)$. Notice that $\|v_t\|_2 \leq O_d(\log^{d/2} T)$ since v_t is $O_d(\log^d T)$-sparse. Note that v_t's are the input vectors for the vector balancing problem.

Now we define the set of test vectors $\mathcal{S}$ that will allow us to bound the discrepancy of any axis-parallel box. For every box B that can be exactly expressed as the disjoint union of several dyadic boxes, i.e. $B = \cup_{B' \in \mathcal{D}'} B'$ for some subset $\mathcal{D}' \subseteq \mathcal{D}$ of disjoint dyadic boxes, we create a test vector $z_B \in \{0,1\}^{|\mathcal{D}|}$ with $z_B(B') = 1$ if and only if $B' \in \mathcal{D}'$. We call such box B a *dyadic-generated* box. Since there are multiple choices of $\mathcal{D}'$ that give the same dyadic-generated box B, we only take $\mathcal{D}'$ to be the one that contains the smallest number of dyadic boxes. $\mathcal{S}$ will be the set of all such dyadic-generated boxes.

Recalling that $|\mathcal{D}| = O_d(T^d)$, it follows that $|\mathcal{S}| \leq 2|\mathcal{D}| = O_d(T^d)$, as each coordinate of a box in $\mathcal{S}$ corresponds to an endpoint of one of the dyadic intervals in $\mathcal{D}$. Moreover, every test vector $z_B \in \mathcal{S}$ is $O_d(\log^d T)$-sparse and thus $\|z_B\|_2 \leq O_d(\log^{d/2} T)$. Using Theorem 8.1.3 with both the input and test vectors scaled down by $O_d(\log^{d/2} T)$, we obtain an algorithm that w.h.p. maintains discrepancy vector d_t such that for all $t \in [T]$,

$$\max_{z_B \in \mathcal{S}} |d_t^\top z_B| \leq O_d(\log^{d+4} T).$$

Since $d_t^\top z_B = \mathsf{disc}_t(B)$ which follows from B being a disjoint union of dyadic boxes, we have $\mathsf{disc}_t(B) \leq O_d(\log^{d+4} T)$ for any dyadic-generated box B.

To upper bound the discrepancy of arbitrary axis-parallel boxes, we first introduce the notion of *stripes*. A stripe in $[0,1]^d$ is an axis-parallel box that is of the form $I_1 \times \cdots \times I_d$ where exactly one of the intervals I_i is allowed to be a proper sub-interval $[a,b] \subseteq [0,1]$. The width of such a stripe is defined to be $b - a$. Stripes whose projection is $[a,b]$ in dimension i satisfying $b - a = 1/T$ correspond to the smallest dyadic interval in dimension i. We call such stripes *minimum dyadic* stripes. There are exactly T minimum dyadic stripes for each dimension $i \in [d]$. Since minimum dyadic stripes have width $1/T$ and the marginal of p along any dimension is the uniform distribution over $[0,1]$, a standard application of Chernoff bound implies that w.h.p. the total number of points in all the minimum dyadic

stripes is at most $O_d(\log(T))$ points.

For a general axis-parallel box $\widetilde{B}$, it is well-known that $\widetilde{B}$ can be expressed as the disjoint union of a dyadic-generated box B together with at most $k \leq 2d$ boxes $B_1, \ldots, B_k$ where each $B_i \subseteq S_i$ is a subset of a minimum dyadic stripe. We can thus upper bound

$$\mathsf{disc}_t(\widetilde{B}) \leq \mathsf{disc}_t(B) + \sum_{i=1}^{k} \mathsf{disc}_t(B_i) \leq \mathsf{disc}_t(B) + \sum_{i=1}^{k} r_i.$$

where r_i is the total number of points in the stripe S_i. As mentioned, w.h.p. we can upper bound $\sum_{i=1}^{k} r_i = O_d(\log(T))$ and thus one obtains $\mathsf{disc}_t(\widetilde{B}) = O_d(\log^{d+4} T)$ for any axis-parallel box $\widetilde{B}$. This proves the theorem. □

8.5.3 Proof of Theorem 8.1.3

Potential Function and Algorithm. By Lemma 8.4.1, it is without loss of generality to assume that p is κ-dyadic, where $\kappa = 8\lceil \log(nT) \rceil$. For any $k \in [\kappa]$, we use Π_k to denote the projection matrix onto the eigenspace of $\mathbf{\Sigma}$ corresponding to the eigenvalue 2^{-k} and define $\Pi = \sum_{k=1}^{\kappa} \Pi_k$ to be the sum of these projection matrices.

The algorithm for Theorem 8.1.3 will use a greedy strategy that chooses the next sign so that a certain potential function is minimized. To define the potential, we first define a distribution where some noise is added to the input distribution p to account for the test vectors. Let p_z be the uniform distribution over the set of test vectors $\mathcal{S}$. We define the noisy distribution p_x to be $\mathsf{p}_x := \mathsf{p}/2 + \mathsf{p}_z/2$, i.e., a random sample from p_x is drawn with probability $1/2$ each from p or p_z. Note that any vector x in the support of p_x satisfies $\|x\|_2 \leq 1$ since both the input distribution p and the set of test vectors $\mathcal{S}$ lie inside the unit Euclidean ball.

At any time step t, let $d_t = \chi_1 v_1 + \ldots + \chi_t v_t$ denote the current discrepancy vector after the signs $\chi_1, \ldots, \chi_t \in \{\pm 1\}$ have been chosen. Set $\lambda^{-1} = 100\kappa \log(nT)$ and define the

potential

$$\Phi_t \quad = \quad \Phi(d_t) \quad := \quad \sum_{k=1}^{\kappa} \mathbb{E}_{x\sim \mathsf{p}_x} \left[\cosh\left(\lambda d_t^\top \Pi_k x\right)\right].$$

When the vector v_t arrives, the algorithm greedily chooses the sign χ_t that minimizes the increase $\Phi_t - \Phi_{t-1}$.

Analysis. The above potential is useful because it allows us to give tail bounds on the length of the discrepancy vectors in most directions given by the distribution p while simultaneously controlling the discrepancy in the test directions. In particular, let $\mathcal{G}_t$ denote the set of *good* vectors v in the support of p that satisfy $\lambda|d_t^\top \Pi v| \leq \kappa \cdot \log(4\Phi_t/\delta)$. Then, we have the following lemma.

Lemma 8.5.1. *For any $\delta > 0$ and any time t, we have*

(a) $\mathbb{P}_{v\sim\mathsf{p}}(v \notin \mathcal{G}_t) \leq \delta$.

(b) $|d_t^\top \Pi_k z| \leq \lambda^{-1} \log(4|\mathcal{S}|\Phi_t)$ *for all* $z \in \mathcal{S}$ *and* $k \in [\kappa]$.

Proof. (a) Recall that with probability $1/2$ a sample from p_x is drawn from the input distribution p. Using this and the fact that $0 \leq \exp(x) \leq 2\cosh(x)$ for any $x \in \mathbb{R}$, we have $\sum_{k\in[\kappa]} \mathbb{E}_{v\sim\mathsf{p}} \left[\exp(\lambda|d_t^\top \Pi_k v|)\right] \leq 4\Phi_t$. Note that for any $v \notin \mathcal{G}_t$, we have $\lambda|d_t^\top \Pi v| > \kappa \cdot \log(4\Phi_t/\delta)$ by definition, so it follows that $\lambda|d_t^\top \Pi_k v| > \log(4\Phi_t/\delta)$ for at least one $k \in [\kappa]$. Thus, applying Markov's inequality we get that $\mathbb{P}_{v\sim\mathsf{p}}(v \notin \mathcal{G}_t) \leq \delta$.

(b) Similarly, a random sample from p_x is drawn from the uniform distribution over $\mathcal{S}$ with probability $1/2$, so $\exp\left(\lambda|d^\top \Pi_k z|\right) \leq 4|\mathcal{S}|\Phi_t$ for every $z \in \mathcal{S}$ and $k \in [\kappa]$. This implies that $|d^\top \Pi_k z| \leq \lambda^{-1} \log(4|\mathcal{S}|\Phi_t)$.

□

The next lemma shows that the expected increase in the potential is small on average.

Lemma 8.5.2 (Bounded positive drift)**.** *At any time step $t \in [T]$, if $\Phi_{t-1} \leq 3T^5$, then $\mathbb{E}_{v_t}[\Phi_t] - \Phi_{t-1} \leq 2$.*

Using Lemma 8.5.2, we first finish the proof of Theorem 8.1.3.

Proof of Theorem 8.1.3. We first use Lemma 8.5.2 to prove that with probability at least $1 - T^{-4}$, the potential $\Phi_t \leq 3T^5$ for every $t \in [T]$. Such an argument is standard and has previously appeared in [JKS19, BJSS20]. In particular, we consider a truncated random process $\widetilde{\Phi}_t$ which is the same as Φ_t until $\Phi_{t_0} > 3T^5$ for some time step t_0; for any t from time t_0 to T, we define $\widetilde{\Phi}_t = 3T^5$. It follows that $\mathbb{P}[\widetilde{\Phi}_t \geq 3T^5] = \mathbb{P}[\Phi_t \geq 3T^5]$. Lemma 8.5.2 implies that for any time $t \in [T]$, the expected value of the truncated process $\widetilde{\Phi}_t$ over the input sequence $v_1, \ldots, v_T$ is at most $3T$. By Markov's inequality, with probability at least $1 - T^{-4}$, the potential $\Phi_t \leq 3T^5$ for every $t \in [T]$.

When the potential $\Phi_t \leq 3T^5$, part (b) of Lemma 8.5.1 implies that $|d^\top \Pi_k z| = O(\lambda^{-1} \cdot (\log(|\mathcal{S}|) + \log T))$ for any $z \in \mathcal{S}$ and $k \in [\kappa]$. Thus, it follows that for every $z \in \mathcal{S}$,

$$|d^\top z| \;\leq\; \sum_{k \in [\kappa]} |d^\top \Pi_k z| = O(\kappa \lambda^{-1} (\log(|\mathcal{S}|) + \log T)) = O((\log(|\mathcal{S}|) + \log T) \cdot \log^3(nT)),$$

which completes the proof of the theorem.

□

To finish the proof, we prove the remaining Lemma 8.5.2 next.

Proof of Lemma 8.5.2. Let us fix a time t. To simplify the notation, let $\Phi = \Phi_{t-1}$ and $\Delta\Phi = \Phi_t - \Phi$, and let $d = d_{t-1}$ and $v = v_t$. To bound the change $\Delta\Phi$, we use Taylor expansion. Since $\cosh'(a) = \sinh(a)$ and $\sinh'(a) = \cosh(a)$, for any $a, b \in \mathbb{R}$ satisfying

$|a-b| \le 1$, we have

$$\begin{aligned}\cosh(\lambda a) - \cosh(\lambda b) &= \lambda \sinh(\lambda b)\cdot(a-b) + \frac{\lambda^2}{2!}\cosh(\lambda b)\cdot(a-b)^2 + \cdots \\ &\le \lambda \sinh(\lambda b)\cdot(a-b) + \lambda^2 \cosh(\lambda b)\cdot(a-b)^2 \\ &\le \lambda \sinh(\lambda b)\cdot(a-b) + \lambda^2 |\sinh(\lambda b)|\cdot(a-b)^2 + \lambda^2(a-b)^2,\end{aligned}$$

where the first inequality follows since $|\sinh(a)| \le \cosh(a)$ for all $a \in \mathbb{R}$, and since $|a-b| \le 1$ and $\lambda < 1$, so the higher order terms in the Taylor expansion are dominated by the first and second order terms. The second inequality uses that $\cosh(a) \le |\sinh(a)| + 1$ for $a \in \mathbb{R}$.

After choosing the sign χ_t, the discrepancy vector $d_t = d + \chi_t v$. Defining $s_k(x) = \sinh(\lambda \cdot d^\top \Pi_k x)$ and noting that $|v^\top \Pi_k x| \le 1$, the above upper bound on the Taylor expansion gives us that

$$\begin{aligned}\Delta\Phi &= \sum_{k\in[\kappa]} \mathbb{E}_x\left[\cosh\left(\lambda(d+\chi_t v)^\top \Pi_k x\right)\right] - \sum_{k\in[\kappa]} \mathbb{E}_x\left[\cosh\left(\lambda d^\top \Pi_k x\right)\right] \\ &\le \underbrace{\chi_t\left(\sum_{k\in[\kappa]} \lambda\ \mathbb{E}_x\left[s_k(x) v^\top \Pi_k x\right]\right)}_{:=\ \chi_t L} + \underbrace{\sum_{k\in[\kappa]} \lambda^2\ \mathbb{E}_x\left[|s_k(x)|\cdot x^\top \Pi_k v v^\top \Pi_k x\right]}_{:=\ Q} \\ &\quad + \underbrace{\sum_{k\in[\kappa]} \lambda^2\ \mathbb{E}_x\left[x^\top \Pi_k v v^\top \Pi_k x\right]}_{:=\ Q_*},\end{aligned}$$

where $\chi_t L, Q$, and Q_* denote the first, second, and third terms respectively. Recall that our algorithm uses the greedy strategy by choosing χ_t to be the sign that minimizes the potential. Taking expectation over the random incoming vector $v \sim \mathsf{p}$, we get

$$\mathbb{E}_v[\Delta\Phi] \le -\mathbb{E}_v[|L|] + \mathbb{E}_v[Q] + \mathbb{E}_v[Q_*].$$

We will prove the following upper bounds on the quadratic (in λ) terms Q and Q_*.

Claim 8.5.3. $\mathbb{E}_v[Q] \leq 2\lambda^2 \sum_{k\in[\kappa]} 2^{-k}\ \mathbb{E}_x[|s_k(x)|]$ *and* $\mathbb{E}_v[Q_*] \leq 4\lambda^2$.

On the other hand, we will show that the linear (in λ) term L is also large in expectation.

Claim 8.5.4. $\mathbb{E}_v[|L|] \geq \lambda B^{-1} \sum_{k\in[\kappa]} 2^{-k}\ \mathbb{E}_x[|s_k(x)|] - 1$ *for some value* $B \leq 2\kappa \cdot \log(\Phi^2 \kappa n)$.

By our assumption that $\Phi \leq 3T^5$, we have that $2\lambda \leq B^{-1}$. Therefore, combining the above two claims, we get that

$$\mathbb{E}_v[\Delta\Phi] \leq (2\lambda^2 - \lambda B^{-1}) \left(\sum_{k\in[\kappa]} 2^{-k}\ \mathbb{E}_x[|s_k(x)|] \right) + 1 + 4\lambda^2 \leq 2.$$

This finishes the proof of Lemma 8.5.2 assuming the claims which we prove next. □

Proof of Claim 8.5.3. Recall that $\mathbb{E}_v[vv^\top] = \mathbf{\Sigma}$ and that $\Pi_k \mathbf{\Sigma} \Pi_k = 2^{-k}\Pi_k$. Using linearity of expectation,

$$\begin{aligned} \mathbb{E}_v[Q] &= \sum_{k\in[\kappa]} \lambda^2\ \mathbb{E}_x[|s_{k(x)}| \cdot x^\top \Pi_k \mathbf{\Sigma} \Pi_k x] = \lambda^2 \sum_{k\in[\kappa]} 2^{-k}\ \mathbb{E}_x[|s_k(x)| \cdot x^\top \Pi_k x] \\ &\leq 2\lambda^2 \sum_{k\in[\kappa]} 2^{-k}\ \mathbb{E}_x[|s_k(x)|], \end{aligned}$$

where the last inequality uses that $\|x\|_2 \leq 1$. Similarly,

$$\mathbb{E}_v[Q_*] = \sum_{k\in[\kappa]} \lambda^2\ \mathbb{E}_x\left[x^\top \Pi_k \mathbf{\Sigma} \Pi_k x\right] \leq 2\lambda^2 \sum_{k\in[\kappa]} 2^{-k} \leq 4\lambda^2.$$

□

Proof of Claim 8.5.4. To lower bound the linear term, we use the fact that $|L(v)| \geq \|f\|_\infty^{-1} \cdot f(v) \cdot L(v)$ for any real-valued non-zero function f. We will choose the function $f(v) = d^\top \Pi v \cdot \mathbf{1}_{\mathcal{G}}(v)$ where $\mathcal{G}$ will be the event that $|d^\top \Pi v|$ is small, which we know is true because

of Lemma 8.5.1. In particular, set $\delta^{-1} = \lambda\Phi T$ and let $\mathcal{G}$ denote the set of vectors v in the support of p such that $\lambda|d^\top \Pi v| \le \kappa \cdot \log(4\Phi/\delta) := B$. Then, $f(v) = d^\top \Pi v \cdot \mathbf{1}_{\mathcal{G}}(v)$ satisfies $\|f\|_\infty \le \lambda^{-1} B$, and we can lower bound,

$$\begin{aligned}\mathbb{E}_v[|L|] &\ge \frac{\lambda}{\lambda^{-1}B} \sum_{k\in[\kappa]} \mathbb{E}_{v,x}[s_k(x) \cdot d^\top \Pi v \cdot v^\top \Pi_k x \cdot \mathbf{1}_{\mathcal{G}}(v)] \\ &= \frac{\lambda^2}{B} \sum_{k\in[\kappa]} \mathbb{E}_x[s_k(x) \cdot d^\top \Pi \mathbf{\Sigma} \Pi_k x] - \frac{\lambda^2}{B} \sum_{k\in[\kappa]} \mathbb{E}_x[s_k(x) \cdot d^\top \Pi \mathbf{\Sigma}_{\mathsf{err}} \Pi_k x], \end{aligned} \tag{8.5}$$

where $\mathbf{\Sigma}_{\mathsf{err}} = \mathbb{E}_v[vv^\top(1 - \mathbf{1}_{\mathcal{G}}(v))]$ satisfies $\|\mathbf{\Sigma}_{\mathsf{err}}\|_{\mathsf{op}} \le \mathbb{P}_{v\sim\mathsf{p}}(v \notin \mathcal{G}) \le \delta$ using Lemma 8.5.1. To bound the first term in (8.5), recall that $s_k(x) = \sinh(\lambda d^\top \Pi_k x)$. Using $\Pi \mathbf{\Sigma} \Pi_k = 2^{-k}\Pi_k$ and the fact that $\sinh(a)a \ge |\sinh(a)| - 2$ for any $a \in \mathbb{R}$, we have

$$\lambda\, \mathbb{E}_x[s_k(x) \cdot d^\top \Pi \mathbf{\Sigma} \Pi_k x] \;=\; 2^{-k}\, \mathbb{E}_x[s_k(x) \cdot \lambda d^\top \Pi_k x] \;\ge\; 2^{-k}\, \left(\mathbb{E}_x[|s_k(x)|] - 2\right).$$

For the second term, we use the bound $\|\mathbf{\Sigma}_{\mathsf{err}}\|_{\mathsf{op}} \le \delta$ to obtain

$$|d^\top \Pi \mathbf{\Sigma}_{\mathsf{err}} \Pi_k x| \;\le\; \|\mathbf{\Sigma}_{\mathsf{err}}\|_{\mathsf{op}} \cdot \|d\|_2 \cdot \|x\|_2 \;\le\; \delta \|d\|_2.$$

Since $\|d\|_2 \le T$ always holds, by our choice of δ,

$$\lambda |d^\top \Pi \mathbf{\Sigma}_{\mathsf{err}} \Pi_k x| \le \Phi^{-1}.$$

Plugging the above bounds in (8.5),

$$\begin{aligned}\mathbb{E}_v[|L|] &\geq \frac{\lambda}{B}\sum_{k\in[\kappa]} 2^{-k}\ \left(\mathbb{E}_x[|s_k(x)|]-2\right) - \frac{\lambda}{B}\cdot\Phi^{-1}\left(\sum_{k\in[\kappa]}\mathbb{E}_x[|s_k(x)|]\right)\\ &\geq \frac{\lambda}{B}\sum_{k\in[\kappa]} 2^{-k}\ \mathbb{E}_x[|s_k(x)|] - \frac{\lambda}{B}\sum_{k\in[\kappa]} 2^{-k+1} - \frac{\lambda}{B}\\ &\geq \frac{\lambda}{B}\sum_{k\in[\kappa]} 2^{-k}\ \mathbb{E}_x[|s_k(x)|] - 1,\end{aligned}$$

where the second inequality follows since $\sum_{k\in[\kappa]}\mathbb{E}_x[|s_k(x)|]\leq\Phi$. □

8.6 Discrepancy with Respect to Arbitrary Convex Bodies

Our main result of this section is the following theorem.

Theorem 8.1.4 (Online Banaszczyk Setting)**.** *Let $K\subseteq\mathbb{R}^n$ be a symmetric convex body with $\gamma_n(K)\geq 1/2$ and p be a distribution with sub-exponential tails that is supported over vectors of Euclidean norm at most 1. Then, for vectors $v_1,\ldots,v_T$ sampled i.i.d. from p, there is an online algorithm that with high probability maintains a discrepancy vector d_t satisfying $d_t\in C\log^5(nT)\cdot K$ for all $t\in[T]$ and a universal constant C.*

8.6.1 Potential Function and Algorithm

As in the previous section, it is without loss of generality to assume that p is κ-dyadic, where $\kappa=8\lceil\log(nT)\rceil$. For any $k\in[\kappa]$, recall that Π_k denotes the projection matrix onto the eigenspace of $\mathbf{\Sigma}$ corresponding to the eigenvalue 2^{-k} and $\Pi=\sum_{k=1}^{\kappa}\Pi_k$. Further, let us also recall that Π_{err} is the projection matrix onto the subspace spanned by eigenvectors corresponding to eigenvalues of $\mathbf{\Sigma}$ that are at most $2^{-\kappa}$. We also note that $\dim(\mathrm{im}(\Pi_k))\leq\min\{2^k,n\}$ since $\mathrm{Tr}(\mathbf{\Sigma})\leq 1$.

Our algorithm to bound the discrepancy with respect to an arbitrary symmetric convex body $K\subseteq\mathbb{R}^n$ with $\gamma_n(K)\geq 1/2$ will use a greedy strategy with a similar potential function as in §8.5. Let p_z be a distribution on *test vectors* in $\mathbb{R}^n$ that will be specified later. Define

the noisy distribution $\mathsf{p}_x = \mathsf{p}/2 + \mathsf{p}_z/2$, *i.e*, a random sample from p_x is drawn from p or p_z with probability $1/2$ each.

At any time step t, let $d_t = \chi_1 v_1 + \ldots + \chi_t v_t$ denote the current discrepancy vector after the signs $\chi_1, \ldots, \chi_t \in \{\pm 1\}$ have been chosen. Set $\lambda^{-1} = 100\kappa \log(nT)$, and define the potential

$$\Phi_t = \Phi(d_t) := \sum_{k \in [\kappa]} \mathbb{E}_{x \sim \mathsf{p}_x} \left[\exp\left(\lambda \, d_t^\top \Pi_k x \right) \right].$$

When the vector v_t arrives, the algorithm chooses the sign χ_t that minimizes the increase $\Phi_t - \Phi_{t-1}$.

Test Distribution. To complete the description of the algorithm, we need to choose a suitable distribution p_z on test vectors to give us control on the norm $\|\cdot\|_K = \sup_{y \in K^\circ} \langle \cdot, y \rangle$. For this, we will use generic chaining.

First let us denote by $H_k = \mathrm{im}(\Pi_k)$ the linear subspace that is the image of the projection matrix Π_k where the subspaces $\{H_k\}_{k \in [\kappa]}$ are orthogonal and span $\mathbb{R}^n$. Moreover, recall that $\dim(H_k) \leq \min\{2^k, n\}$.

Let us denote by $K_k = K \cap H_k$ the slice of the convex body K with the subspace H_k. Proposition 8.3.2 implies that $\gamma_{H_k}(K) \geq 1/2$ for each $k \in [\kappa]$ and combined with Proposition 8.3.3 this implies that $K_k^\circ := (K_k)^\circ = \Pi_k(K^\circ)$ satisfies $\mathsf{diam}(K_k^\circ) \leq 4$ and $w_{H_k}(K_k^\circ) \leq 3/2$ for every k.

Consider ε-nets of the polar bodies K_k° at geometrically decreasing dyadic scales. Let

$$\varepsilon_{\mathsf{min}}(k) = 2^{-\left\lceil \log_2\left(\frac{1}{10\lambda}\sqrt{\dim(H_k)}\right)\right\rceil} \text{ and } \varepsilon_{\mathsf{max}}(k) = 2^{-\log_2 \lceil 1/\mathsf{diam}(K_k^\circ)\rceil},$$

be the finest and the coarsest scales for a fixed k, and for integers $\ell \in [\log_2(1/\varepsilon_{\mathsf{max}}(k)), \log_2(1/\varepsilon_{\mathsf{min}}(k))]$, define the scale $\varepsilon(\ell, k) = 2^{-\ell}$. We call these *admissible* scales for any fixed k.

Note that for a fixed $k \in [\kappa]$, the number of admissible scales is at most $2\log_2(nT)$ since $\mathsf{diam}(K_k^\circ) \leq 4$. The smallest scale is chosen because with high probability we can always control the Euclidean norm of the discrepancy vector in the subspace H_k to be

$\lambda^{-1}\log(nT)\sqrt{\dim(H_k)}$ using a test distribution as used in Komlos's setting.

Let $\mathcal{T}(\ell,k)$ be an optimal $\varepsilon(\ell,k)$-net of K_k°. For each k, define the following directed layered graph $\mathcal{G}_k$ (recall Figure 8.1) where the vertices in layer ℓ are the elements of $\mathcal{T}(\ell,k)$. Note that the first layer indexed by $\log_2(1/\varepsilon_{\mathsf{max}}(k))$ consists of a single vertex, the origin. We add a directed edge from $u \in \mathcal{T}(\ell,k)$ to $v \in \mathcal{T}(\ell+1,k)$ if $\|v-u\|_2 \le \varepsilon(\ell,k)$. We identify an edge (u,v) with the vector $v-u$ and define its length as $\|v-u\|_2$. Let $\mathcal{S}(\ell,k)$ denote the set of edges between layer ℓ and $\ell+1$. Note that any edge $(u,v) \in \mathcal{S}(\ell,k)$ has length at most $\varepsilon(\ell,k)$ and since $w_{H_k}(K_k^\circ) \le 3/2$, Proposition 8.3.4 implies that,

$$|\mathcal{S}(\ell,k)| \quad \le \quad |\mathcal{T}(\ell+1,k)|^2 \quad \le \quad 2^{16/\varepsilon(\ell,k)^2}. \tag{8.6}$$

Pick the final test distribution as $\mathsf{p}_z = \mathsf{p}_{\mathbf{\Sigma}}/2 + \mathsf{p}_y/2$ where $\mathsf{p}_{\mathbf{\Sigma}}$ and p_y denote the distributions given in Figure 8.2.

(a) $\mathsf{p}_{\mathbf{\Sigma}}$ is uniform over the eigenvectors $u_1,\ldots,u_n$ of the covariance matrix $\mathbf{\Sigma}$.

(b) p_y samples a random vector as follows: pick an integer k uniformly from $[\kappa]$ and an admissible scale $\varepsilon(\ell,k)$ with probability $\dfrac{2^{-2/\varepsilon(\ell,k)^2}}{\sum_\ell 2^{-2/\varepsilon(\ell,k)^2}}$. Choose a uniform vector from $r(\ell,k)^2 \cdot \mathcal{S}(\ell,k)$, where the scaling factor $r(\ell,k) := 1/\varepsilon(\ell,k)$.

Figure 8.2: Test distributions $\mathsf{p}_{\mathbf{\Sigma}}$ and p_y

The above test distribution completes the description of the algorithm. Note that adding the eigenvectors will allow us to control the Euclidean length of the discrepancy vectors in the subspaces H_k as they form an orthonormal basis for these subspaces. Also observe that, as opposed to the previous section, the test vectors chosen above may have large Euclidean length as we scaled them. For future reference, we note that the entire probability mass assigned to length r vectors in the support of p_y is at most 2^{-2r^2} where $r \ge 1/4$.

8.6.2 Potential Implies Low Discrepancy

The test distribution p_z is useful because of the following lemma. In particular, a $\mathsf{poly}(n, T)$ upper bound on the potential function implies a polylogarithmic discrepancy upper bound on $\|d_t\|_K$.

Lemma 8.6.1. *At any time t, we have that*

$$\|\Pi_k d_t\|_2 \leq \lambda^{-1} \log(4n\Phi_t)\sqrt{\dim(H_k)} \quad \textit{and} \quad \|d_t\|_K \leq O(\kappa \cdot \lambda^{-1} \cdot \log(nT) \cdot \log(\Phi_t)).$$

Proof. To derive a bound on the Euclidean length of $\Pi_k d_t$, we note that a random sample from p_x is drawn from the uniform distribution over $\{u_i\}_{i\leq n}$ with probability $1/4$, so $\exp\left(\lambda |d_t^\top \Pi_k u_i|\right) \leq 4n\Phi_t$ for every $k \in [\kappa]$ and every $i \in [n]$. Since $\{u_i\}_{i\leq n}$ also form an eigenbasis for Π, we get that $|d_t^\top \Pi_k u_i| \leq \lambda^{-1}\log(4n\Phi_t)$ which implies that $\|\Pi_k d_t\|_2 \leq \lambda^{-1}\log(4n\Phi_t)\sqrt{\dim(H_k)}$.

To see the bound on $\|d_t\|_K$, we note that

$$\|d_t\|_K = \sup_{y\in K^\circ} \langle d_t, y\rangle \;\leq\; \sum_{k\in[\kappa]} \sup_{y\in K_k^\circ} \langle \Pi_k d_t, y\rangle \;\leq\; \sum_{k\in[\kappa]} \left(\sup_{z\in\mathcal{T}(\ell,k)} |d_t^\top \Pi_k z| + \varepsilon_{\mathsf{min}}(k)\|\Pi_k d_t\|_2 \right), \tag{8.7}$$

where the last inequality holds since $\mathcal{T}(\ell, k)$ is an $\varepsilon_{\mathsf{min}}(k)$-net of K_k°. By our choice of $\varepsilon_{\mathsf{min}}(k)$ and the bound on $\|\Pi_k d_t\|_2$ from the first part of the Lemma, we have that $\varepsilon_{\mathsf{min}}(k)\|\Pi_k d_t\|_2 \leq 10\log(4n\Phi_t)$.

To upper bound $\sup_{z\in\mathcal{T}(\ell,k)}\langle \Pi_k d_t, z\rangle$, we pick any arbitrary $z \in \mathcal{T}(\ell, k)$ and consider any path from the origin to z in the graph $\mathcal{G}_k$. Let $(u_\ell, u_{\ell+1})$ be the edges of this path for $\ell \in [\log_2(1/\varepsilon_{\mathsf{min}}), \log_2(1/\varepsilon_{\mathsf{max}})]$ where $u_\ell = 0$ for $\ell = \log_2(1/\varepsilon_{\mathsf{max}})$ and $u_\ell = z$ for $\ell = \log_2(1/\varepsilon_{\mathsf{min}})$. Then $z = \sum_\ell w_\ell$ where $w_\ell = (u_{\ell+1} - u_\ell)$. By our choice of the test distribution, the bound

on the potential implies the following for any edge $w \in \mathcal{S}(\ell, k)$,

$$\exp\left(\lambda \cdot r(\ell,k)^2 \cdot |d_t^\top \Pi_k w|\right) \;\leq\; 2^{2/\varepsilon(\ell,k)^2} \cdot |\mathcal{S}(\ell,k)| \cdot 4\Phi_t \;\leq\; 2^{18/\varepsilon(\ell,k)^2} \cdot 4\Phi_t,$$

where the second inequality follows from $|\mathcal{S}(\ell,k)| \leq 2^{16/\varepsilon(\ell,k)^2}$ in (8.6). This implies that for any edge $w \in \mathcal{S}(\ell,k)$,

$$|d_t^\top \Pi_k w| \;\leq\; \lambda^{-1} \log(4\Phi_t).$$

Since $z = \sum_\ell w_\ell$ and there are at most $\log(n)$ different scales ℓ, we get that $|d_t^\top \Pi_k z| \leq \lambda^{-1} \cdot \log(n) \cdot \log(4\Phi_t)$. Since z was arbitrary in $\mathcal{T}(\ell,k)$, plugging the above bound in (8.7) completes the proof. □

The next lemma shows that the expected increase (or drift) in the potential is small on average.

Lemma 8.6.2 (Bounded Positive Drift)**.** *Let* p *be supported on the unit Euclidean ball in* $\mathbb{R}^n$ *and has a sub-exponential tail. There exist an absolute constant* $C > 0$ *such that if* $\Phi_{t-1} \leq T^5$ *for any* t*, then* $\mathbb{E}_{v_t \sim \mathsf{p}}[\Phi_t] - \Phi_{t-1} \leq C$*.*

Analogous to the proof of Theorem 8.1.3, Lemma 8.6.2 implies that w.h.p. the potential $\Phi_t \leq T^5$ for every $t \in [T]$. Combined with Lemma 8.6.1, and recalling that $\kappa = O(\log nT)$ and $\lambda^{-1} = O(\kappa \log(nT))$, this proves Theorem 8.1.4. To finish the proof, we prove Lemma 8.6.2 in the next section.

8.6.3 Drift Analysis: Proof of Lemma 8.6.2

The proof is quite similar to the analysis for Komlos's setting. In particular, we have the following tail bound analogous to Lemma 8.5.1. Let $\mathcal{G}_t$ denote the set of *good* vectors v in the support of p that satisfy $\lambda |d_t^\top \Pi v| \leq \kappa \cdot \log(4\Phi_t/\delta)$.

Lemma 8.6.3. *For any* $\delta > 0$ *and any time* t*, we have* $\mathbb{P}_{v \sim \mathsf{p}}(v \notin \mathcal{G}_t) \leq \delta$*.*

We omit the proof of the above lemma as it is the same as that of Lemma 8.5.1.

Proof of Lemma 8.6.2. Recall that our potential function is defined to be

$$\Phi_t := \sum_{k\in[\kappa]} \mathbb{E}_{x\sim \mathsf{p}_x}\left[\exp\left(\lambda\ d_t^\top \Pi_k x\right)\right],$$

where $\mathsf{p}_x = \mathsf{p}/2 + \mathsf{p}_{\mathbf{\Sigma}}/4 + \mathsf{p}_y/4$ is a combination of the input distribution p and test distributions $\mathsf{p}_{\mathbf{\Sigma}}$ and p_y, each constituting a constant mass.

Let us fix a time t. To simplify the notation, we denote $\Phi = \Phi_{t-1}$ and $\Delta\Phi = \Phi_t - \Phi$, and denote $d = d_{t-1}$ and $v = v_t$. To bound the potential change $\Delta\Phi$, we use the following inequality, which follows from a modification of the Taylor series expansion of $\cosh(r)$ and holds for any $a, b \in \mathbb{R}$,

$$\cosh(\lambda a) - \cosh(\lambda b) \le \lambda \sinh(\lambda b)\cdot(a-b) + \frac{\lambda^2}{2}\cosh(\lambda b)\cdot e^{|a-b|}(a-b)^2. \tag{8.8}$$

Note that when $|a-b| \ll 1$, then $e^{|a-b|} \le 2$, so one gets the first two terms of the Taylor expansion as an upper bound, but here we will also need it when $|a-b| \gg 1$.

Note that every vector in the support of p and $\mathsf{p}_{\mathbf{\Sigma}}$ has Euclidean length at most 1, while $y \sim \mathsf{p}_y$ may have large Euclidean length due to the scaling factor of $r(\ell,k)^2$. Therefore, we decompose the distribution p_x appearing in the potential as $\mathsf{p}_x = \frac{3}{4}\mathsf{p}_w + \frac{1}{4}\mathsf{p}_y$, where the distribution $\mathsf{p}_w = \frac{2}{3}\mathsf{p} + \frac{1}{3}\mathsf{p}_{\mathbf{\Sigma}}$ is supported on vectors with Euclidean length at most 1.

After choosing the sign χ_t for v, the discrepancy vector d_t becomes $d + \chi_t v$. For ease of notation, define $s_k(x) = \sinh(\lambda \cdot d^\top \Pi_k x)$ and $c_k(x) = \cosh(\lambda \cdot d^\top \Pi_k x)$ for any $x \in \mathbb{R}^n$. Now (8.8) implies that $\Delta\Phi := \Delta\Phi_1 + \Delta\Phi_2$ where

$$\begin{aligned}\Delta\Phi_1 &\le \chi_t \cdot \frac{3}{4}\left(\sum_{k\in[\kappa]} \lambda\ \mathbb{E}_w\left[s_k(w) v^\top \Pi_k w\right]\right) + \frac{3}{4}\sum_{k\in[\kappa]} \lambda^2\ \mathbb{E}_w\left[c_k(w)\cdot w^\top \Pi_k v v^\top \Pi_k w\right] \\ &:= \chi_t L_1 + Q_1,\end{aligned}$$

and

$$\Delta\Phi_2 \le \chi_t \cdot \frac{1}{4}\left(\sum_{k\in[\kappa]} \lambda\ \mathbb{E}_y\left[s_k(y)v^\top \Pi_k y\right]\right) + \frac{1}{4}\sum_{k\in[\kappa]} \lambda^2\ \mathbb{E}_y\left[c_k(y)\cdot e^{\lambda|v^\top \Pi_k y|} y^\top \Pi_k v v^\top \Pi_k y\right]$$
$$:= \chi_t L_2 + Q_2.$$

Since our algorithm chooses sign χ_t to minimize the potential increase, taking expectation over the incoming vector v, we get

$$\mathbb{E}_v[\Delta\Phi] \le -\mathbb{E}_v[|L_1 + L_2|] + \mathbb{E}_v[Q_1 + Q_2].$$

We will prove the following upper bounds on the quadratic terms (in λ) Q_1 and Q_2.

Claim 8.6.4. $\mathbb{E}_v[Q_1+Q_2] \le C\cdot\lambda^2 \sum_{k\in[\kappa]} 2^{-k}\ \mathbb{E}_x[c_k(x)\|x\|_2^2]$ *for an absolute constant* $C > 0$.

On the other hand, we will show that the linear (in λ) terms $L_1 + L_2$ is also large in expectation.

Claim 8.6.5. $\mathbb{E}_v[|L_1+L_2|] \ge \lambda B^{-1}\sum_{k\in[\kappa]} 2^{-k}\ \ \mathbb{E}_x[c_k(x)\|x\|_2^2] - O(1)$ *for some* $B \le 4\kappa\log(\Phi^2 n\kappa)$.

By our assumption of $\Phi \le T^5$, so it follows that $2\lambda \le B^{-1}$. Therefore, combining the above two claims,

$$\mathbb{E}_v[\Delta\Phi] \quad\le\quad (2\lambda^2 - \lambda B^{-1})\left(\sum_{k\in[\kappa]} 2^{-k}\ \mathbb{E}_x\left[c_k(x)\|x\|_2^2\right]\right) + C \quad\le\quad C,$$

which finishes the proof of Lemma 8.6.2 assuming the claims. □

To prove the missing claims, we need the following property that follows from the sub-exponential tail of the input distribution p.

Lemma 8.6.6. *There exists a constant $C > 0$, such that for every integer $k \in [\kappa]$, and any $y \in \mathrm{im}(\Pi_k)$ satisfying $\|y\|_2 \leq \frac{1}{4}\sqrt{\min\{2^k, n\}}$, the following holds*

$$\mathbb{E}_{v \sim \mathsf{p}}\left[e^{\lambda |v^\top y|} \cdot |v^\top y|^2\right] \leq C \cdot 2^{-k} \cdot \|y\|_2^2 \text{ for all } \lambda \leq 1.$$

We remark that this is the only step in the proof which requires the sub-exponential tail, as otherwise the exponential term above may be quite large. It may however be possible to exploit some more structure from the test vectors y and the discrepancy vector to prove the above lemma without any sub-exponential tail requirements from the input distribution.

Proof. As $y \in \mathrm{im}(\Pi_k)$, we have that $v^\top y = v^\top \Pi_k y$ which is a scalar sub-exponential random variable with zero mean and variance at most

$$\sigma_y^2 \;:=\; \mathbb{E}_v[|v^\top \Pi_k y|^2] \;\leq\; \|\Pi_k \mathbf{\Sigma} \Pi_k\|_{\mathsf{op}} \|y\|_2^2 \;\leq\; 2^{-k}\|y\|_2^2 \;\leq\; 1/16.$$

Using Cauchy-Schwarz and Proposition 8.3.1, we get that

$$\begin{aligned}\mathbb{E}_v\left[e^{\lambda |v^\top y|} \cdot |v^\top y|^2\right] &\leq \sqrt{\mathbb{E}_v\left[e^{2\lambda |v^\top y|}\right]} \cdot \sqrt{\mathbb{E}_v\left[|v^\top y|^4\right]} \\ &\leq C \cdot \mathbb{E}_v\left[|v^\top \Pi_k y|^2\right] \;\leq\; C \cdot 2^{-k}\, \|y\|_2^2,\end{aligned}$$

where the exponential term is bounded since $\sigma_y \leq 1/4$. □

Proof of Claim 8.6.4. Recall that $\mathbb{E}_v[vv^\top] = \mathbf{\Sigma}$ which satisfies $\Pi_k \mathbf{\Sigma} \Pi_k = 2^{-k}\Pi_k$. Therefore, using linearity of expectation,

$$\begin{aligned}\mathbb{E}_v[Q_1] \;=\; \frac{3}{4}\sum_{k \in [\kappa]} \lambda^2\, \mathbb{E}_w[c_k(w) \cdot w^\top \Pi_k \mathbf{\Sigma} \Pi_k w] \;&=\; \lambda^2 \cdot \frac{3}{4}\sum_{k \in [\kappa]} 2^{-k}\, \mathbb{E}_w[c_k(w) \cdot w^\top \Pi_k w] \\ &\leq\; 2\lambda^2 \cdot \frac{3}{4}\sum_{k \in [\kappa]} 2^{-k}\, \mathbb{E}_w[c_k(w)\|w\|_2^2]. \qquad (8.9)\end{aligned}$$

We next use Lemma 8.6.6 to bound the second quadratic term

$$\mathbb{E}_v[Q_2] \;=\; \frac{1}{4}\sum_{k\in[\kappa]} \lambda^2\, \mathbb{E}_y\left[c_k(y)\cdot e^{\lambda|v^\top \Pi_k y|} y^\top \Pi_k v v^\top \Pi_k y\right].$$

For any $k \in [\kappa]$ and any $y \in \mathrm{im}(\Pi_k)$ that is in the support of p_y, we have that

$$\lambda\|\Pi_k y\|_2 \le \lambda\cdot\|y\|_2 \;\le\; \lambda/\varepsilon_{\mathsf{min}}(k) \;\le\; \lambda\cdot\frac{1}{10\lambda}\cdot\sqrt{\dim(H_k)} \;\le\; \frac{1}{4}\sqrt{\min\{n,2^k\}}.$$

On the other hand, if $y \in \mathrm{im}(\Pi_{k'})$ for $k' \neq k$, then the above quantity is zero. Lemma 8.6.6 then implies that for any y in the support of p_y,

$$\mathbb{E}_v[e^{|\lambda v^\top \Pi_k y|}\cdot|\lambda v^\top \Pi_k y|^2] \;\le\; C_1\cdot 2^{-k}\|\lambda\Pi_k y\|_2^2 \;\le\; C_1\lambda^2\cdot 2^{-k}\|y\|_2^2,$$

where C_1 is some absolute constant. Therefore, we obtain the following bound

$$\mathbb{E}_v[Q_2] \;\le\; C_1\cdot\lambda^2\cdot\sum_{k\in[\kappa]} 2^{-k}\;\mathbb{E}_y[c_k(y)\|y\|_2^2]. \tag{8.10}$$

Summing up (8.9) and (8.10) finishes the proof of the claim. □

Proof of Claim 8.6.5. Let $L = L_1 + L_2$. To lower bound the linear term, we proceed similarly as in the proof of Claim 8.5.4 and use the fact that $|L(v)| \ge \|f\|_\infty^{-1}\cdot f(v)\cdot L(v)$ for any real-valued non-zero function f. We will choose the function $f(v) = d^\top \Pi v\cdot \mathbf{1}_{\mathcal{G}}(v)$ where $\mathcal{G}$ will be the event that $|d^\top\Pi v|$ is small which we know is true because of Lemma 8.6.3.

In particular, set $\delta^{-1} = \lambda^{-2} n\cdot\Phi\cdot\log(4n\Phi)$ and let $\mathcal{G}$ denote the set of vectors v in the support of p such that $\lambda|d^\top\Pi v| \le \kappa\cdot\log(4\Phi/\delta) := B$. Then, $f(v) = d^\top\Pi v\cdot\mathbf{1}_{\mathcal{G}}(v)$ satisfies

$\|f\|_\infty \le \lambda^{-1}B$, and we can lower bound,

$$\begin{aligned}\mathbb{E}_v[|L|] &\ge \frac{\lambda}{\lambda^{-1}B}\cdot\frac{3}{4}\sum_{k\in[\kappa]}\mathbb{E}_{vw}[s_k(w)\cdot d^\top\Pi v\cdot v^\top\Pi_k w\cdot\mathbf{1}_{\mathcal{G}}(v)]\\ &\quad+\frac{\lambda}{\lambda^{-1}B}\cdot\frac{1}{4}\sum_{k\in[\kappa]}\mathbb{E}_{vy}[s_k(y)\cdot d^\top\Pi v\cdot v^\top\Pi_k y\cdot\mathbf{1}_{\mathcal{G}}(v)]\\ &=\frac{\lambda^2}{B}\cdot\frac{3}{4}\sum_{k\in[\kappa]}\mathbb{E}_w[s_k(w)\cdot d^\top\Pi\mathbf{\Sigma}\Pi_k w]\;-\;\frac{\lambda^2}{B}\cdot\frac{3}{4}\sum_{k\in[\kappa]}\mathbb{E}_w[s_k(w)\cdot d^\top\Pi\mathbf{\Sigma}_{\mathsf{err}}\Pi_k w]\\ &\quad+\frac{\lambda^2}{B}\cdot\frac{1}{4}\sum_{k\in[\kappa]}\mathbb{E}_y[s_k(y)\cdot d^\top\Pi\mathbf{\Sigma}\Pi_k y]\;-\;\frac{\lambda^2}{B}\cdot\frac{1}{4}\sum_{k\in[\kappa]}\mathbb{E}_y[s_k(y)\cdot d^\top\Pi\mathbf{\Sigma}_{\mathsf{err}}\Pi_k y],\end{aligned}\tag{8.11}$$

where $\mathbf{\Sigma}_{\mathsf{err}} = \mathbb{E}_v[vv^\top(1-\mathbf{1}_{\mathcal{G}}(v))]$ satisfies $\|\mathbf{\Sigma}_{\mathsf{err}}\|_{\mathsf{op}} \le \mathbb{P}_{v\sim\mathsf{p}}(v\notin\mathcal{G}) \le \delta$ using Lemma 8.5.1.

To bound the terms involving $\mathbf{\Sigma}$ in (8.11), we recall that $s_k(x) = \sinh(\lambda d^\top\Pi_k x)$ and $c_k(x) = \cosh(\lambda d^\top\Pi_k x)$. Using $\Pi\mathbf{\Sigma}\Pi_k = 2^{-k}\Pi_k$ and the fact that $\sinh(a)a \ge \cosh(a)|a| - 2$ for any $a\in\mathbb{R}$, we have

$$\lambda\,\mathbb{E}_w[s_k(w)\cdot d^\top\Pi\mathbf{\Sigma}\Pi_k w] \;=\; 2^{-k}\,\mathbb{E}_w[s_k(w)\cdot\lambda d^\top\Pi_k w] \;\ge\; 2^{-k}\,\left(\mathbb{E}_w[c_k(w)|\lambda d^\top\Pi_k w|] - 2\right),$$

and similarly for y.

The terms with $\mathbf{\Sigma}_{\mathsf{err}}$ can be upper bounded using $\|\mathbf{\Sigma}_{\mathsf{err}}\|_{\mathsf{op}} \le \delta$. In particular, we have

$$|d^\top\Pi\mathbf{\Sigma}_{\mathsf{err}}\Pi_k x| \le \|\Pi d\|_2\|\mathbf{\Sigma}_{\mathsf{err}}\|_{\mathsf{op}}\|x\|_2 \le \delta\|\Pi d\|_2\|x\|_2.$$

Since $\Pi = \sum_{k\in[\kappa]}\Pi_k$ and $(\Pi_k)_{k\in[\kappa]}$ are orthogonal projectors, Lemma 8.6.1 implies that $\|\Pi d\|_2 \le \lambda^{-1}\log(4n\Phi)\sqrt{n}$. Moreover, we have $\|w\|_2 \le 1$ and $\|y\|_2 \le \min_k\{1/\varepsilon_{\mathsf{min}}(k)\} \le \frac{1}{10\lambda}\cdot\sqrt{n}$. Then, by our choice of $\delta^{-1} = \lambda^{-2}n\Phi\cdot\log(4n\Phi)$, we have

$$\lambda\,|d^\top\Pi\mathbf{\Sigma}_{\mathsf{err}}\Pi_k x| \;\le\; \delta\lambda^{-1}n\log(4n\Phi) \;=\; \Phi^{-1}.$$

Plugging the above bounds in (8.11), we obtain

$$\mathbb{E}_v[|L|] \;\geq\; \frac{\lambda}{B}\cdot\frac{3}{4}\sum_{k\in[\kappa]} 2^{-k}\;\mathbb{E}_w[c_k(w)|\lambda d^\top \Pi_k w|] + \frac{\lambda}{B}\cdot\frac{1}{4}\sum_{k\in[\kappa]} 2^{-k}\;\mathbb{E}_y[c_k(y)|\lambda d^\top \Pi_k y|] - 4 \tag{8.12}$$

where we used the upper bound $\sum_{k\in[\kappa]}\mathbb{E}_x[|s_k(x)|] \leq \Phi$ to control the error term involving $\mathbf{\Sigma}_{\mathsf{err}}$.

To finish the proof, we bound the two terms in (8.12) separately. We first use the inequality that $\cosh(a)a \geq \cosh(a) - 2$ for all $a \in \mathbb{R}$ and the fact that $\|w\|_2 \leq 1$ for every w in the support of p_w to get that

$$\mathbb{E}_w[c_k(w)|\lambda d^\top \Pi_k w|] \;\geq\; \mathbb{E}_w[c_k(w)] - 2 \;\geq\; \mathbb{E}_w[c_k(w)\|w\|_2^2] - 2. \tag{8.13}$$

To bound the second term in (8.12), we recall that the entire probability mass assigned to length r vectors (i.e. $\epsilon(\ell, k) = 1/r$) in the support of p_y is at most 2^{-2r^2}, where $r \geq 1/4$. Let $\mathcal{E}$ be the event that $|\lambda d^\top \Pi_k y| \leq \|y\|_2^2$. Note that $c_k(y)\|y\|_2^2 \leq 2^{r^2} r^2$ if $\|y\|_2 = r$. This implies that

$$\begin{aligned}\mathbb{E}_y[c_k(y)|\lambda d^\top \Pi_k y|] \;&\geq\; \mathbb{E}_y[c_k(y)\|y\|_2^2] - \mathbb{E}_y[c_k(y)\|y\|_2^2\cdot \mathbf{1}_{\mathcal{E}}(y)] \\ &\geq\; \mathbb{E}_y[c_k(y)\|y\|_2^2] - \int_{1/4}^{\infty} 2^{-2r^2} 2^{r^2} r^2 \;\geq\; \mathbb{E}_y[c_k(y)\|y\|_2^2] - 1. \end{aligned} \tag{8.14}$$

Since $\mathsf{p}_x = \frac{3}{4}\mathsf{p}_w + \frac{1}{4}\mathsf{p}_y$, plugging (8.13) and (8.14) into (8.12) give that

$$\mathbb{E}_v[|L|] \geq \lambda B^{-1}\sum_{k\in[\kappa]} 2^{-k}\;\;\mathbb{E}_x[c_k(x)\|x\|_2^2] - C,$$

for some constant $C > 0$. This completes the proof of the claim. □

8.7 Generalization to Weighted Multi-Color Discrepancy

In this section, we prove Theorem 8.1.5 which we restate below for convenience.

Theorem 8.1.5 (Weighted multi-color discrepancy)**.** *For any input distribution* p *and any set* $\mathcal{S}$ *of* $\mathsf{poly}(nT)$ *test vectors with Euclidean norm at most one, there is an online algorithm for the weighted multi-color discrepancy problem that maintains discrepancy* $O(\log^2(R\eta) \cdot \log^4(nT))$ *with the norm* $\|\cdot\|_* = \max_{z\in\mathcal{S}} |\langle \cdot, z\rangle|$.

Further, if the input distribution p *has sub-exponential tails then one can maintain multi-color discrepancy* $O(\log^2(R\eta) \cdot \log^5(nT))$ *for any norm* $\|\cdot\|_*$ *given by a symmetric convex body* K *satisfying* $\gamma_n(K) \geq 1/2$.

Theorem 8.1.5 follows from a black-box way of converting an algorithm for the signed discrepancy setting to the multi-color setting.

In particular, for a parameter $0 \leq \lambda \leq 1$, let $\Phi : \mathbb{R}^n \to \mathbb{R}_+$ be a potential function satisfying

$$\begin{aligned} &\Phi(d + \alpha v) \leq \Phi(d) + \lambda\alpha L_d(v) + \lambda^2\alpha^2 Q_d(v) \qquad \text{for every } d, v \in \mathbb{R}^n \text{ and } |\alpha| \leq 1, \text{ and,} \\ &-\lambda \cdot \mathbb{E}_{v\sim\mathsf{p}}[|L_d(v)|] + \lambda^2 \cdot \mathbb{E}_{v\sim\mathsf{p}}[Q_d(v)] = O(1) \quad \text{for any } d \text{ such that } \Phi(d) \leq 3T^5, \end{aligned} \tag{8.15}$$

where $L_d : \mathbb{R}^n \to \mathbb{R}$ and $Q_d : \mathbb{R}^n \to \mathbb{R}_+$ are arbitrary functions of v that depend on d.

One can verify that the first condition is always satisfied for the potential functions used for proving Theorem 8.1.3 and Theorem 8.1.4, while the second condition holds for $\lambda = O(1/\log^2(nT))$ because of Lemma 8.5.2 and Lemma 8.6.2.

Moreover, for parameters n and T, let $B_{\|\cdot\|_*}$ be such that if the potential $\Phi(d) = \Phi$, then the corresponding norm $\|d\|_* \leq B_{\|\cdot\|_*} \log(nT\Phi)$. Part (b) of Lemma 8.5.1 implies that for any test set $\mathcal{S}$ of $\mathsf{poly}(nT)$ vectors contained in the unit Euclidean ball, if the norm $\|\cdot\|_* = \max_{z\in\mathcal{S}} |\langle \cdot, z\rangle|$, then $B_{\|\cdot\|_*} = O(\log^3(nT))$. Similarly, if $\|\cdot\|_*$ is given by a symmetric convex body with Gaussian measure at least $1/2$, then Lemma 8.6.1 implies that $B_{\|\cdot\|_*} = O(\log^4(nT))$.

We will use the above properties of the potential Φ to give a greedy algorithm for the multi-color discrepancy setting.

8.7.1 Weighted Binary Tree Embedding

We first show how to embed the weighted multi-color discrepancy problem into a binary tree $\mathcal{T}$ of height $O(\log(R\eta)$. For each color c, we create $\lfloor w_c \rfloor$ nodes with weight $w_c/\lfloor w_c \rfloor \in [1, 2]$ each. The total number of nodes is thus $M_\ell = \sum_{c\in[R]} \lfloor w_c \rfloor = O(R\eta)$. In the following, we place these nodes as the leaves of an (incomplete) binary tree.

Take the height $h = O(\log(R\eta))$ to be the smallest exponent of 2 such that $2^h \geq M_\ell$. We first remove $2^h - M_\ell < 2^{h-1}$ leaves from the complete binary tree of height h such that none of the removed leaves are siblings. Denote the set of remaining leaves as $\mathcal{L}(\mathcal{T})$. Then from left to right, assign the leaves in $\mathcal{L}(\mathcal{T})$ to the R colors so that leaves corresponding to the same color are consecutive. For each leaf node $\ell \in \mathcal{L}(\mathcal{T})$ that is assigned the color $c \in [R]$, we assign it the weight $w_\ell = w_c/\lfloor w_c \rfloor$.

We index the internal nodes of the tree as follows: for integers $0 \leq j \leq h-1$ and $0 \leq k \leq 2^j$, we use (j, k) to denote the 2^k-th node at depth j. Note that the left and right children of a node (j, k) are the nodes $(j + 1, 2k)$ and $(j + 1, 2k + 1)$. The weight $w_{j,k}$ of an internal node (j, k) is defined to be sum of weights of all the leaves in the sub-tree rooted at (j, k). This way of embedding satisfies certain desirable properties which we give in the following lemma.

Lemma 8.7.1 (Balanced tree embedding)**.** *For the weighted (incomplete) binary tree $\mathcal{T}$ defined above, for any two nodes (j, k) and (j, k') in the same level,*

$$1/4 \leq w_{j,k}/w_{j,k'} \leq 4.$$

Proof. Observe that each leaf node $\ell \in \mathcal{L}(\mathcal{T})$ has weight $w_\ell \in [1, 2]$. Moreover, for each internal node $(h - 1, k)$ in the level just above the leaves, at least one of its children is not removed in the construction of $\mathcal{T}$. Therefore, it follows that $w_{j,k} = a_{j,k}2^{h-j}$ for some

$a_{j,k} \in [1/2, 2]$ and similarly for (j, k'). The lemma now immediately follows from these observations. □

Induced random walk on the weighted tree. Randomly choosing a leaf with probability proportional to its weight induces a natural random walk on the tree $\mathcal{T}$: the walk starts from the root and moves down the tree until it reaches one of the leaves. Conditioned on the event that the walk is at some node (j, k) in the j-th level, it goes to left child $(j + 1, 2k)$ with probability $q^l_{j,k} = w_{j+1,2k}/w_{j,k}$ and to the right child $(j + 1, 2k + 1)$ with probability $q^r_{j,k} = w_{j+1,2k+1}/w_{j,k}$. Note that by Lemma 8.7.1 above, we have that both $q^l_{j,k}, q^r_{j,k} \in [1/5, 4/5]$ for each internal node (j, k) in the tree. Note that $w_{j,k}/w_{0,0}$ denotes the probability that the random walk passes through the vertex j, k.

8.7.2 Algorithm and Analysis

Recall that each leaf $\ell \in \mathcal{L}(\mathcal{T})$ of the tree $\mathcal{T}$ is associated with a color. Our online algorithm will assign each arriving vector v_t to one of the leaves $\ell \in \mathcal{L}(\mathcal{T})$ and its color will then be the color of the corresponding leaf.

For a leaf $\ell \in \mathcal{L}(\mathcal{T})$, let $d_\ell(t)$ denote the sum of all the input vectors that are associated with the leaf ℓ at time t. For an internal node (j, k), we define $d_{j,k}(t)$ to be the sum $\sum_{\ell \in \mathcal{L}(\mathcal{T}_{j,k})} d_\ell(t)$ where $\mathcal{L}(\mathcal{T}_{j,k})$ is the set of all the leaves in the sub-tree rooted at (j, k). Also, let $d^l_{j,k}(t) = d_{j+1,2k}(t)$ and $d^r_{j,k}(t) = d_{j+1,2k+1}(t)$ be the vectors associated with the left and right child of the node (j, k).

Finally let,

$$d^-_{j,k}(t) = \frac{d^l_{j,k}(t)/q^l_{j,k} - d^r_{j,k}(t)/q^r_{j,k}}{1/q^l_{j,k} + 1/q^r_{j,k}} = q^r_{j,k} d^l_{j,k}(t) - q^l_{j,k} d^r_{j,k},$$

denote the weighted difference between the two children vectors for the (j, k)-th node of the tree.

Algorithm. For $\beta = 1/(400h)$, consider the following potential function

$$\Psi_t = \sum_{j,k \in \mathcal{T}} \Phi(\beta\ d^-_{j,k}(t)),$$

where the sum is over all the internal nodes (j,k) of $\mathcal{T}$.

The algorithm assigns the incoming vector v_t to the leaf $\ell \in \mathcal{L}(\mathcal{T})$, so that the increase in the potential $\Psi_t - \Psi_{t-1}$ is minimized. The color assigned to the vector v_t is then the color of the corresponding leaf ℓ.

We show that if the potential Φ satisfies (8.15), then the drift for the potential Ψ can be bounded.

Lemma 8.7.2. *If at any time t, if $\Psi_{t-1} \le T^5$, then the following holds*

$$\mathbb{E}_{v_t \sim \mathsf{p}}[\Delta\Psi_t] := \mathbb{E}_{v_t \sim \mathsf{p}}[\Psi_t - \Psi_{t-1}] = O(1).$$

Using standard arguments as used in the proof of Theorem 8.1.3, this implies that with high probability $\Psi_t \le T^5$ at all times t.

Moreover, the above potential also gives a bound on the discrepancy because of the following lemma.

Lemma 8.7.3. *If $\Psi_t \le T^5$, then $\mathsf{disc}_t = O(\beta^{-1}h \cdot B_{\|\cdot\|_*} \cdot \log(nT\Psi_t)) = O(h^2 \cdot B_{\|\cdot\|_*} \cdot \log(nT))$.*

Combined with part (b) of Lemma 8.5.1 and Lemma 8.6.1, the above implies Theorem 8.1.5. Next we prove Lemma 8.7.3 and Lemma 8.7.2 in that order.

Bounded Potential Implies Low Discrepancy. For notational simplicity, we fix a time t and drop the time index below.

Proof of Lemma 8.7.3. First note that $\Phi(\beta \cdot d^-_{j,k}) \le \Psi$, and therefore, $\left\|d^-_{j,k}\right\|_* \le \beta^{-1}B(\|\cdot\|_*) := U$ for every internal node (j,k).

We next claim by induction that the above implies the following for every internal node (j,k),

$$\left\| d_{j,k} - d_{0,0} \cdot \frac{w_{j,k}}{w_{0,0}} \right\|_* \leq \beta_j U, \tag{8.16}$$

where $\beta_j = 1 + 4/5 + \cdots + (4/5)^j$.

The claim is trivially true for the root. For an arbitrary node $(j+1, 2k)$ at depth j that is the left child of some node (j,k), we have that

$$\begin{aligned}\left\| d_{j+1,2k} - d_{0,0} \cdot \frac{w_{j+1,2k}}{w_{0,0}} \right\|_* &\leq \left\| d_{j+1,2k} - d_{j,k} \cdot \frac{w_{j+1,2k}}{w_{j,k}} \right\|_* + q^l_{j,k} \cdot \left\| d_{j,k} - d_{0,0} \cdot \frac{w_{j,k}}{w_{0,0}} \right\|_* \\ &\leq \left\| d^l_{j,k} - d_{j,k} \cdot q^l_{j,k} \right\|_* + q^l_{j,k} \beta_j U,\end{aligned}$$

since $w_{j+1,2k}/w_{j,k} = q^l_{j,k}$ and $q^l_{j,k}, q^r_{j,k} \in [1/5, 4/5]$. Note that $d_{j,k} = d^l_{j,k} + d^r_{j,k}$, so the first term above equals $\left\| d^-_{j,k} \right\|_*$. Therefore, it follows that $\left\| d_{j+1,2k} - d_{0,0} \cdot (w_{j+1,2k}/w_{0,0}) \right\|_* \leq \beta_{j+1} U$. The claim follows analogously for all nodes that are the right children of its parent.

To see the statement of the lemma, consider any color $c \in [R]$. We say that an internal node has color c if all its leaves are assigned color c. A maximal color-c node is a node that has color c but its ancestor doesn't have color c. We denote the set of maximal c-color node to be $\mathcal{M}_c$. Notice that $|\mathcal{M}_c| \leq 2h$ since c-color leaves are consecutive. Also, note that $\sum_{(j,k)\in\mathcal{M}_c} w_{j,k} = w_c$ and that $\sum_{(j,k)\in\mathcal{M}_c} d_{j,k} = d_c$ is exactly the sum of vectors with color c. Therefore, we have

$$\left\| d_c/w_c - d_{0,0}/w_{0,0} \right\|_* \leq \left\| d_c - d_{0,0} \cdot \frac{w_c}{w_{0,0}} \right\|_* \leq \sum_{(j,k)\in\mathcal{M}_c} \left\| d_{j,k} - d_{0,0} \cdot \frac{w_{j,k}}{w_{0,0}} \right\|_* = O(h \cdot U),$$

where the first inequality follows since $w_c \geq 1$ and the last follows from (8.16).

Thus, for any two colors $c \neq c'$, we have

$$\mathsf{disc}_t(c,c') = \left\| \frac{d_c/w_c - d_{c'}/w_{c'}}{1/w_c + 1/w_{c'}} \right\|_* \leq \left\| \frac{d_c/w_c - d_{0,0}/w_{0,0}}{1/w_c + 1/w_{c'}} \right\|_* + \left\| \frac{d_{c'}/w_{c'} - d_{0,0}/w_{0,0}}{1/w_c + 1/w_{c'}} \right\|_* = O(h \cdot U).$$

This finishes the proof of the lemma. □

Bounding the Drift. Now we give the proof of Lemma 8.7.2.

Proof of Lemma 8.7.2. We fix the time t and write $d^-_{j,k} = d^-_{j,k}(t-1)$. Let $X_{j,k}(\ell) \cdot v_t$ denote the change of $d^-_{j,k}$ when the leaf chosen for v_t is ℓ. More specifically, $X_{j,k}(\ell)$ is $q^r_{j,k}$ if the leaf ℓ belongs to the left sub-tree of node (j,k), is $-q^l_{j,k}$ if it belongs to the right sub-tree, and is 0 otherwise. Then, $d^-_{j,k}(t) = d^-_{j,k} + X_{j,k}(\ell) \cdot v_t$ if the leaf ℓ is chosen.

By our assumption on the potential, we have that $\Delta\Psi_t \leq \beta\lambda L + \beta^2\lambda^2 Q$ where

$$L = \sum_{(j,k)\in\mathcal{P}(\ell)} X_{j,k}(\ell) \cdot L_{j,k}(v_t)$$
$$Q = \sum_{(j,k)\in\mathcal{P}(\ell)} X_{j,k}(\ell)^2 \cdot Q_{j,k}(v_t),$$

and $\mathcal{P}(\ell)$ is the root-leaf path to the leaf ℓ.

Consider choosing leaf ℓ (and hence the root-leaf path $\mathcal{P}(\ell)$) randomly in the following way: First pick a uniformly random layer $j^* \in \{0, 1, \cdots, h-1\}$ (i.e., level of the tree), then starting from the root randomly choose a child according to the random walk probability for all layers except j^*; for layer j^*, suppose we arrive at node (j^*, k), we pick the left child if $L_{j^*,k}(v_t) \leq 0$, and the right child otherwise. Note that conditioned on a fixed value of j_*, this ensures that $\mathbb{E}_\ell[X_{j,k}L_{j,k}(v_t)]$ is always negative if $j = j_*$ and is zero otherwise.

Since we follow the random walk before layer j^*, for a fixed choice of j^* we get a node in its layer proportional to their weights. Let us write $\mathcal{N}_j$ for the set of all nodes at depth j. In expectation over the randomness of the input vector v_t and our random choice of leaf ℓ, we have

$$\mathbb{E}_{v_t,\ell}[L] \leq -\frac{1}{h} \cdot \sum_{j=0}^{h-1} \sum_{k\in\mathcal{N}_j} \frac{w_{j,k}}{\sum_{j\in\mathcal{N}_j} w_{j,k}} \cdot \min\{q^l_{j,k}, q^r_{j,k}\} \cdot \mathbb{E}_{v_t}[|L_{j,k}|].$$

For the Q term, recall that one is randomly picking a child until layer j^*, in which one

picks a child depending on $L_{j^*,k}$, and then we continue randomly for the remaining layers. Note that since Q is always positive, this can be at most 20 times a process that always picks a random root-leaf path, since we have $q^l_{j,k}, q^r_{j,k} \in [1/5, 4/5]$. Therefore, we have

$$\mathbb{E}_{v_t,\ell}[Q] \leq 20 \cdot \sum_{j=0}^{h-1} \sum_{k \in \mathcal{N}_j} \frac{w_{j,k}}{\sum_{j \in \mathcal{N}_j} w_{j,k}} \cdot \mathbb{E}_{v_t}[Q_{j,k}].$$

By our choice of $\beta = 1/(400h)$, the above implies that

$$\begin{aligned}
\mathbb{E}_{v_t}[\Delta\Psi_t] &\leq -\sum_{j=0}^{h-1} \sum_{k \in \mathcal{N}_j} \frac{w_{j,k}}{\sum_{j \in \mathcal{N}_j} w_{j,k}} \cdot \left(-\frac{\beta\lambda}{20h} \mathbb{E}_{v_t}[|L_{j,k}|] + 20\beta^2\lambda^2 \mathbb{E}_{v_t}[Q_{j,k}] \right) \\
&\leq -\sum_{j=0}^{h-1} \sum_{k \in \mathcal{N}_j} \frac{w_{j,k}}{\sum_{j \in \mathcal{N}_j} w_{j,k}} \cdot \frac{1}{8000h^2} \cdot \left(-\lambda \mathbb{E}_{v_t}[|L_{j,k}|] + \lambda^2 \mathbb{E}_{v_t}[Q_{j,k}] \right) = O(1).
\end{aligned}$$

Since the algorithm is greedy, the leaf ℓ it assigns to the incoming vector v produces an even smaller drift, so this completes the proof. □

Chapter 9

ONLINE DISCREPANCY III: A POTENTIAL FUNCTION BASED ANALYSIS OF THE SELF-BALANCING WALK

In this chapter, we study the online discrepancy problem in the oblivious adversary setting, which is more difficult than the stochastic setting studied in the previous two chapters. It turns out that $\mathsf{poly}(\log T)$ discrepancy bound can also be achieved in this more general setting. This result was first proved by Alweiss, Liu, and Sawhney [ALS21] using a simple but powerful algorithm known as the self-balancing walk. However, their original analysis of this algorithm was based on the notion of mean-preserving spread and is less explicit. In this chapter, we present a more direct folklore proof of their result.

9.1 Introduction

We revisit the Online Komlós problem studied in Chapter 8: vectors $v_1, v_2, \ldots, v_T \in \mathbb{R}^n$ with Euclidean norm at most 1 arrive online, and upon the arrival of v_t, a sign $x_t \in \{\pm 1\}$ must be chosen irrevocably, so that the ℓ_∞-norm of the *discrepancy vector* (signed sum) $d_t := x_1 v_1 + \ldots + x_t v_t$ remains as small as possible. That is, find the smallest B such that $\max_{t\in[T]} \|d_t\|_\infty \leq B$.

In Chapter 8, we proved the following Theorem 8.1.1 for the Online Komlós problem under the assumption that vectors $v_1, \ldots, v_T$ are sampled i.i.d. from some distribution p.

Theorem 8.1.1 (Online Komlós setting)**.** *Let* p *be a distribution in* $\mathbb{R}^n$ *supported on vectors with Euclidean norm at most* 1*. Then, for vectors* $v_1, \ldots, v_T$ *sampled i.i.d. from* p*, there is an online algorithm that with high probability maintains a discrepancy vector* d_t *such that* $\|d_t\|_\infty = O(\log^4(nT))$ *for all* $t \in [T]$*.*

While Theorem 8.1.1 matches the best known offline discrepancy bound of $O((\log T)^{1/2})$

due to Banaszczyk [Ban12] up to poly-logarithmic factors, the stochasticity assumption that the vectors are sampled i.i.d. from a distribution is not known to be necessary. It is natural to ask if Theorem 8.1.1 can be extended to the stronger *oblivious adversary* setting. Here, an adversary fixes vectors $v_1, \cdots, v_T \in \mathbb{R}^n$ ahead of time, and then the player is presented with the vectors $v_1, \cdots, v_T$ in an online manner. It was first proved by Alweiss, Liu, and Sawhney [ALS21] that poly-logarithmic discrepancy bound can indeed be achieved even against an oblivious adversary.

Theorem 9.1.1 (Self-Balancing Walk, [ALS21])**.** *For any vectors $v_1, \ldots, v_T \in \mathbb{R}^n$ with Euclidean norm at most 1, there is an online algorithm that with high probability maintains a discrepancy vector d_t such that $\|d_t\|_\infty = O(\log(nT))$ for all $t \in [T]$.*

Alweiss, Liu, and Sawhney designed a simple but powerful algorithm for Theorem 9.1.1 known as the *self-balancing walk*, which we present in Section 9.2. Their original analysis of this algorithm is rather implicit, relying on the notion of mean-preserving spread. In the rest of this chapter, we present a more direct analysis of their algorithm using a potential function similar in spirit to the one we used in Chapter 8. The analysis presented in this chapter is folklore[1], which also appeared later more formally in [DM21].

9.2 Self-Balancing Walk

In this section, we present the algorithm given in [ALS21], known as the self-balancing walk. At the start of the algorithm, it picks a target discrepancy bound of $c > 0$. The algorithm starts with $d_0 = 0$ and maintains the discrepancy vector d_t until the current time step $t \in [T]$. In the t-th step, the algorithm checks the conditions (1) $|\langle d_{t-1}, v_t\rangle| > c$, and (2) the discrepancy $\|d_{t-1}\|_\infty > c$. If any of these two conditions hold, then the algorithm aborts and outputs "Fail"; otherwise, it chooses the signs probabilistically according to the inner product $|\langle d_{t-1}, v_t\rangle|$, with bias towards -1 if $\langle d_{t-1}, v_t\rangle > 0$. A formal description of the self-balancing walk algorithm is given in Algorithm 5.

[1]To the best of our knowledge, this potential-based analysis of the self-balancing walk has been independently reconstructed multiple times.

Algorithm 5 Self-Balancing Walk

procedure SELF-BALANCING($v_1, \cdots, v_T$) ▷ $v_1, \cdots, v_T \in \mathbb{R}^n$ and each $\|v_i\|_2 \leq 1$
  $d_0 \leftarrow 0$
  $c \leftarrow 100 \log(nT)$
  for $t = 1, \cdots, T$ **do**
    if $|\langle d_{t-1}, v_t \rangle| > c$ or $\|d_{t-1}\|_\infty > c$ **then**
      Fail
    end if
    $p_t \leftarrow \frac{1}{2} - \frac{\langle d_{t-1}, v_t \rangle}{2c}$
    $x_t \leftarrow 1$ with probability p_t, and $x_t \leftarrow -1$ with probability $1 - p_t$
    $d_t \leftarrow d_{t-1} + x_t v_t$
  end for
end procedure

9.3 Potential Function Analysis for ALS

In this section, we give a folklore analysis of the self-balancing walk (Algorithm 5) and prove Theorem 9.1.1, which we restate below for convenience.

Theorem 9.1.1 (Self-Balancing Walk, [ALS21])**.** *For any vectors $v_1, \ldots, v_T \in \mathbb{R}^n$ with Euclidean norm at most 1, there is an online algorithm that with high probability maintains a discrepancy vector d_t such that $\|d_t\|_\infty = O(\log(nT))$ for all $t \in [T]$.*

Proof of Theorem 9.1.1. Let d_t be the discrepancy vector maintained at time step $t \in [T]$. Note that Algorithm 5 might output "Fail" at some time step t_0. If this happens, we let $d_t = 0$ for all steps $t \geq t_0$, i.e., we move the discrepancy vector back to the origin. Note that under this definition, we always have $\|d_t\|_\infty \leq c$, where $c = 100 \log(nT)$ as in Algorithm 5.

Our goal is to use induction to prove that for all time steps $t \in [T]$,

$$\mathbb{E}[\exp(\langle d_t, \theta \rangle)] \leq \exp(c\|\theta\|_2^2), \quad \text{for all vector } \theta \in \mathbb{R}^n, \qquad \text{(Induction Hypothesis)}$$

where the randomness is over the outcome of d_t, i.e., the algorithm's choice of $x_1, \cdots, x_t$. Note that (Induction Hypothesis) is equivalent to saying that the random discrepancy vector d_t is $O(\sqrt{c})$-subgaussian.

Before proving (Induction Hypothesis), we first show how it implies the theorem. If (Induction Hypothesis) holds for all time step $t \in [T]$, we have via Markov's inequality that $\mathbb{P}[|\langle d_{t-1}, v_t\rangle| > c] \leq 1/\mathsf{poly}(n,T)$ and $\mathbb{P}[\|d_{t-1}\|_\infty > c] \leq 1/\mathsf{poly}(n,T)$ for large enough $\mathsf{poly}(n,T)$. This implies that the probability that $t \in [T]$ is the first step where Algorithm 5 outputs "Fail" is at most $1/\mathsf{poly}(n,T)$. Union bound over all $t \in [T]$ shows that Algorithm 5 never outputs "Fail" with high probability. The discrepancy bound of c then follows immediately as the algorithm always outputs "Fail" in step t if $\|d_{t-1}\|_\infty > c$. We are therefore left with proving (Induction Hypothesis).

Induction Basis. Since $d_0 = 0$, (Induction Hypothesis) is clearly satisfied at $t = 0$.

Induction Step. Assuming (Induction Hypothesis) holds for the current step t, we now prove that it holds for the next step $t+1$. For notational simplicity, we drop the subscripts and let d be the current discrepancy vector, v be the incoming vector, x be the (random) sign chosen for v, and $y = d + xv$ be the (random) discrepancy vector in the next step. Let $\theta \in \mathbb{R}^n$ be an arbitrary test vector for which we want to establish (Induction Hypothesis) in the next step.

For now, let us assume that the condition $|\langle d, v\rangle| \leq c$ holds for every outcome of d, and the algorithm proceeds by picking a $\pm$-sign in the current step. It's easy to handle the case $|\langle d, v\rangle| > c$ (where the algorithm moves d to 0) using symmetry and we do that towards the end of this proof. Conditioning on an outcome of d, we note that

$$\mathbb{E}[y|d] = \left(\frac{1}{2} - \frac{\langle d, v\rangle}{2c}\right) \cdot (d+v) + \left(\frac{1}{2} + \frac{\langle d, v\rangle}{2c}\right)(d-v) = \left(I - \frac{vv^\top}{c}\right) d.$$

We can write $y = \mathbb{E}[y|d] + r$, where r is a random vector in the direction of v such that $\|r\|_2 \leq 2\|v\|_2 \leq 2$. Using these notations, we can bound

$$\begin{aligned}\mathbb{E}[\exp(\langle y, \theta\rangle)] &= \mathbb{E}\left[\exp\left(\left\langle \left(I - \frac{vv^\top}{c}\right) d, \theta \right\rangle\right) \cdot \exp(\langle r, \theta\rangle)\right] \\ &\leq \exp(\langle v, \theta\rangle^2) \cdot \mathbb{E}\left[\exp\left(\left\langle \left(I - \frac{vv^\top}{c}\right) d, \theta \right\rangle\right)\right],\end{aligned}$$

where we condition on d and take the expectation only with respect to the randomness over r. Now for the second term above, we can write

$$\left\langle \left(I - \frac{vv^\top}{c} \right) d, \theta \right\rangle = \langle d, \theta \rangle - \frac{\langle v, d \rangle \cdot \langle v, \theta \rangle}{c} = \left\langle d, \theta - \frac{\langle v, \theta \rangle}{c} v \right\rangle.$$

Applying (Induction Hypothesis) on d using the test vector $\theta - \frac{\langle v,\theta \rangle}{c} \cdot v \in \mathbb{R}^n$ gives

$$\begin{aligned}
\mathbb{E}[\exp(\langle y, \theta \rangle)] &\leq \exp(\langle v, \theta \rangle^2) \cdot \exp\left(c \cdot \left\| \theta - \frac{\langle v, \theta \rangle}{c} v \right\|_2^2 \right) \\
&\leq \exp(\langle v, \theta \rangle^2) \cdot \exp\left(c \cdot \|\theta\|_2^2 - 2\langle v, \theta \rangle^2 + \frac{\|v\|_2^2}{c} \cdot \langle v, \theta \rangle^2 \right) \leq \exp(c \cdot \|\theta\|_2^2).
\end{aligned}$$

Since the test direction $\theta \in \mathbb{R}^n$ is arbitrary, this completes the induction proof assuming that we always have $|\langle d, v \rangle| \leq c$.

Finally we handle the issue of possibly having some outcome of $d = d^*$ which satisfies $|\langle d^*, v \rangle| > c$. The point mass at such an outcome d^* will be moved to $y = 0$ in the next step. The main observation is that due to symmetry, the outcome $d = d^*$ and $d = -d^*$ has the same probability mass. Using this symmetry, we note that the inequality

$$\mathbb{E}[\exp(\langle y, \theta \rangle) | d^* \text{ or } -d^*] \leq \exp(\langle v, \theta \rangle^2) \cdot \mathbb{E}\left[\exp\left(\left\langle \left(I - \frac{vv^\top}{c} \right) d, \theta \right\rangle \right) \middle| d^* \text{ or } -d^* \right] \quad (9.1)$$

is still correct when we average over the two symmetric point masses of d at d^* and $-d^*$. This is because the LHS of (9.1) is 1 (since $y = 0$) and the RHS of (9.1) is at least 1 by symmetry. We can then proceed exactly as in the main case where there's no violation to $|\langle x, v \rangle| \leq c$. This completes the proof of (Induction Hypothesis) and by our earlier argument, the proof of the theorem. □

BIBLIOGRAPHY

[AAGM15] Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D Milman. *Asymptotic geometric analysis, Part I*, volume 202. American Mathematical Soc., 2015.

[AC91] Ilan Adler and Steven Cosares. A strongly polynomial algorithm for a special class of linear programs. *Operations Research*, 39(6):955–960, 1991.

[ADRSD15] Divesh Aggarwal, Daniel Dadush, Oded Regev, and Noah Stephens-Davidowitz. Solving the shortest vector problem in 2n time using discrete gaussian sampling. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 733–742, 2015.

[AEG+22] Simon Apers, Yuval Efron, Paweł Gawrychowski, Troy Lee, Sagnik Mukhopadhyay, and Danupon Nanongkai. Cut query algorithms with star contraction. *arXiv preprint arXiv:2201.05674*, 2022.

[AKS01] Miklós Ajtai, Ravi Kumar, and Dandapani Sivakumar. A sieve algorithm for the shortest lattice vector problem. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 601–610, 2001.

[ALPTJ10] Radosław Adamczak, Alexander Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *Journal of the American Mathematical Society*, 23(2):535–561, 2010.

[ALS20] Brian Axelrod, Yang P Liu, and Aaron Sidford. Near-optimal approximate discrete and continuous submodular function minimization. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 837–853. SIAM, 2020.

[ALS21] Ryan Alweiss, Yang P Liu, and Mehtaab Sawhney. Discrepancy minimization via a self-balancing walk. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 14–20, 2021.

[ANTSV02] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma, and Umesh Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM (JACM)*, 49(4):496–511, 2002.

[AW02] Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[B+13] Francis Bach et al. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends® in Machine Learning*, 6(2-3):145–373, 2013.

[Ban98] Wojciech Banaszczyk. Balancing vectors and gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12(4):351–360, 1998.

[Ban10] Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 3–10. IEEE, 2010.

[Ban12] Wojciech Banaszczyk. On series of signed vectors and their rearrangements. *Random Structures & Algorithms*, 40(3):301–316, 2012.

[Ban19] Nikhil Bansal. On a generalization of iterated and randomized rounding. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2019.

[Bár79] Imre Bárány. On a Class of Balancing Games. *J. Comb. Theory, Ser. A*, 26(2):115–126, 1979.

[Bár08] Imre Bárány. On the power of linear dependencies. In *Building bridges*, pages 31–45. Springer, 2008.

[BBvH21] Afonso S Bandeira, March T Boedihardjo, and Ramon van Handel. Matrix concentration inequalities and free probability. *arXiv preprint arXiv:2108.06312*, 2021.

[BCKL14a] Nikhil Bansal, Moses Charikar, Ravishankar Krishnaswamy, and Shi Li. Better algorithms and hardness for broadcast scheduling via a discrepancy approach. In *Proceedings of the Twenty-fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 55–71. SIAM, 2014.

[BCKL14b] Nikhil Bansal, Moses Charikar, Ravishankar Krishnaswamy, and Shi Li. Better algorithms and hardness for broadcast scheduling via a discrepancy approach. In *Symposium on Discrete Algorithms*, pages 55–71, 2014.

[BCL94] Keith Ball, Eric A Carlen, and Elliott H Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Inventiones Mathematicae*, 115(1):463–482, 1994.

[BDG72] D. L. Burkholder, B. J. Davis, and R. F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proceedings of BSMSP*, volume 2, pages 223–240, 1972.

[BDG16] Nikhil Bansal, Daniel Dadush, and Shashwat Garg. An algorithm for komlós conjecture matching banaszczyk's bound. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 788–799. IEEE, 2016.

[BDG19] Nikhil Bansal, Daniel Dadush, and Shashwat Garg. An algorithm for komlós conjecture matching banaszczyk's bound. *SIAM Journal on Computing*, 48(2):534–553, 2019.

[BDGL18] Nikhil Bansal, Daniel Dadush, Shashwat Garg, and Shachar Lovett. The gram-schmidt walk: a cure for the banaszczyk blues. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 587–597, 2018.

[Bec81] József Beck. Roth's estimate of the discrepancy of integer sequences is nearly sharp. *Combinatorica*, 1(4):319–325, 1981.

[BF81] József Beck and Tibor Fiala. "integer-making" theorems. *Discrete Applied Mathematics*, 3(1):1–8, 1981.

[BG81] Imre Bárány and Victor S Grinberg. On some combinatorial questions in finite-dimensional spaces. *Linear Algebra and its Applications*, 41:1–9, 1981.

[BG17] Nikhil Bansal and Shashwat Garg. Algorithmic discrepancy beyond partial coloring. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2017.

[BJL+19] Sébastien Bubeck, Qijia Jiang, Yin Tat Lee, Yuanzhi Li, and Aaron Sidford. Complexity of highly parallel non-smooth convex optimization. *Advances in Neural Information Processing Systems*, 2019.

[BJM+21] Nikhil Bansal, Haotian Jiang, Raghu Meka, Sahil Singla, and Makrand Sinha. Online discrepancy minimization for stochastic arrivals. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2842–2861. SIAM, 2021.

[BJM22a] Nikhil Bansal, Haotian Jiang, and Raghu Meka. An elementary proof of a more general matrix discrepancy bound. *Unpublished manuscript*, 2022.

[BJM22b] Nikhil Bansal, Haotian Jiang, and Raghu Meka. Resolving matrix spencer conjecture up to poly-logarithmic rank. *arXiv preprint arXiv:2208.11286*, 2022.

[BJSS20] Nikhil Bansal, Haotian Jiang, Sahil Singla, and Makrand Sinha. Online vector balancing and geometric discrepancy. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1139–1152, 2020.

[BKPP18] Gerdus Benade, Aleksandr M. Kazachkov, Ariel D. Procaccia, and Christos-Alexandros Psomas. How to Make Envy Vanish Over Time. In *Proceedings of EC 2018*, pages 593–610, 2018.

[BLV22] Nikhil Bansal, Aditi Laddha, and Santosh S Vempala. A unified approach to discrepancy minimization. *arXiv preprint arXiv:2205.01023*, 2022.

[BN17] Nikhil Bansal and Viswanath Nagarajan. Approximation-friendly discrepancy rounding. In *A Journey Through Discrete Mathematics*, pages 89–114. Springer, 2017.

[BRS22] Nikhil Bansal, Lars Rohwedder, and Ola Svensson. Flow time scheduling and prefix beck-fiala. In *Symposium on Theory of Computing (STOC)*, 2022.

[BS13] Nikhil Bansal and Joel Spencer. Deterministic discrepancy minimization. *Algorithmica*, 67(4):451–471, 2013.

[BS18] Eric Balkanski and Yaron Singer. Parallelization does not accelerate convex optimization: Adaptivity lower bounds for non-smooth convex minimization. *arXiv preprint arXiv:1808.03880*, 2018.

[BS20] Eric Balkanski and Yaron Singer. A lower bound for parallel submodular minimization. In *Proceedings of the 52nd annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 130–139, 2020.

[BSS12] Joshua Batson, Daniel A Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM Journal on Computing*, 41(6):1704–1721, 2012.

[Bub15] Sébastien Bubeck. Convex optimization: Algorithms and complexity, 2015.

[Bud11] Eric Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011.

[BV04] Dimitris Bertsimas and Santosh Vempala. Solving convex programs by random walks. *Journal of the ACM (JACM)*, 51(4):540–556, 2004.

[BVZ01] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11):1222–1239, 2001.

[Car16] Eric A. Carlen. A remainder term for hölder's inequality for matrices and quantum entropy inequalities, 2016.

[Cas71] John William Scott Cassels. *An introduction to the theory of numbers*. Springer-Verlag, 1971.

[CCK21] Deeparnab Chakrabarty, Yu Chen, and Sanjeev Khanna. A polynomial lower bound on the number of rounds for parallel submodular function minimization and matroid intersection. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2021.

[CGJS22] Deeparnab Chakrabarty, Andrei Graur, Haotian Jiang, and Aaron Sidford. Improved lower bounds for submodular function minimization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2022.

[Cha00] Bernard Chazelle. *The discrepancy method: randomness and complexity*. Cambridge University Press, 2000.

[Cho94] Sergej Chobanyan. Convergence as of rearranged random series in Banach space and associated inequalities. In *Probability in Banach Spaces, 9*, pages 3–29. Springer, 1994.

[Chu12] Sergei Chubanov. A strongly polynomial algorithm for linear systems having a binary solution. *Mathematical programming*, 134(2):533–570, 2012.

[Chu15] Sergei Chubanov. A polynomial algorithm for linear optimization which is strongly polynomial under certain conditions on optimal solutions, 2015.

[CLSW17] Deeparnab Chakrabarty, Yin Tat Lee, Aaron Sidford, and Sam Chiu-wai Wong. Subquadratic submodular function minimization. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1220–1231, 2017.

[CM94] Edith Cohen and Nimrod Megiddo. Improved algorithms for linear inequalities with two variables per inequality. *SIAM Journal on Computing*, 23(6):1313–1347, 1994.

[CST+14] William Chen, Anand Srivastav, Giancarlo Travaglini, et al. *A panorama of discrepancy theory*, volume 2107. Springer, 2014.

[Cun85] William H Cunningham. On submodular function minimization. *Combinatorica*, 5(3):185–192, 1985.

[Dad12] Daniel Dadush. *Integer programming, lattice algorithms, and deterministic volume estimation*. PhD thesis, Georgia Institute of Technology, 2012.

[DBW12] John C Duchi, Peter L Bartlett, and Martin J Wainwright. Randomized smoothing for stochastic optimization. *SIAM Journal on Optimization*, 22(2):674–701, 2012.

[DFGGR19] Raaz Dwivedi, Ohad N. Feldheim, Ori Gurel-Gurevich, and Aaditya Ramdas. The power of online thinning in reducing discrepancy. *Probability Theory and Related Fields*, 174:103–131, 2019.

[DG19] Jelena Diakonikolas and Cristóbal Guzmán. Lower bounds for parallel and randomized convex optimization. In *Conference on Learning Theory*, pages 1132–1157. PMLR, 2019.

[DGK+14] John P. Dickerson, Jonathan R. Goldman, Jeremy Karp, Ariel D. Procaccia, and Tuomas Sandholm. The computational rise and fall of fairness. In *Proceedings of AAAI*, pages 1405–1411, 2014.

[DGLN16] Daniel Dadush, Shashwat Garg, Shachar Lovett, and Aleksandar Nikolov. Towards a constructive version of banaszczyk's vector balancing theorem. *arXiv preprint arXiv:1612.04304*, 2016.

[DHNV20] Daniel Dadush, Sophie Huiberts, Bento Natura, and László A Végh. A scaling-invariant algorithm for linear programming whose running time depends only on the constraint matrix. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 761–774, 2020.

[DJR22] Daniel Dadush, Haotian Jiang, and Victor Reis. A new framework for matrix discrepancy: partial coloring bounds via mirror descent. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 649–658, 2022.

[DM13] Daniel Dadush and Daniele Micciancio. Algorithms for the densest sub-lattice problem. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1103–1122. SIAM, 2013.

[DM21] Raaz Dwivedi and Lester Mackey. Kernel thinning. *arXiv preprint arXiv:2105.05842*, 2021.

[DNTTJ18] Daniel Dadush, Aleksandar Nikolov, Kunal Talwar, and Nicole Tomczak-Jaegermann. Balancing vectors in any norm. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1–10. IEEE, 2018.

[DVZ18] Daniel Dadush, László A Végh, and Giacomo Zambelli. Geometric rescaling algorithms for submodular function minimization. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 832–848. SIAM, 2018.

[DVZ20] Daniel Dadush, László A Végh, and Giacomo Zambelli. Rescaling algorithms for linear conic feasibility. *Mathematics of Operations Research*, 45(2):732–754, 2020.

[DVZ21] Daniel Dadush, László A. Végh, and Giacomo Zambelli. Geometric rescaling algorithms for submodular function minimization. *Mathematics of Operations Research*, 46(3):1081–1108, 2021.

[Edm70] Jack Edmonds. Submodular functions, matroids, and certain polyhedra. *Edited by G. Goos, J. Hartmanis, and J. van Leeuwen*, page 11, 1970.

[Eld13] Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013.

[EPR13] Friedrich Eisenbrand, Dömötör Pálvölgyi, and Thomas Rothvoß. Bin packing via discrepancy of permutations. *ACM Transactions on Algorithms (TALG)*, 9(3):1–15, 2013.

[ES18] Ronen Eldan and Mohit Singh. Efficient algorithms for discrepancy minimization in convex sets. *Random Structures & Algorithms*, 53(2):289–307, 2018.

[FI00] Lisa Fleischer and Satoru Iwata. Improved algorithms for submodular function minimization and submodular flow. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 107–116, 2000.

[FI03] Lisa Fleischer and Satoru Iwata. A push-relabel framework for submodular function minimization and applications to parametric optimization. *Discrete Applied Mathematics*, 131(2):311–322, 2003.

[Fol66] Duncan Karl Foley. *Resource allocation and the public sector*. Yale University, 1966.

[Fra21] Cole Franks. A simplified disproof of beck's three permutations conjecture and an application to root-mean-squared discrepancy. *Combinatorics, Probability and Computing*, 30(3):398–411, 2021.

[Gia97] Apostolos A Giannopoulos. On some vector balancing problems. *Studia Mathematica*, 122(3):225–234, 1997.

[GLS81] Martin Grötschel, László Lovász, and Alexander Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

[GLS84] Martin Grötschel, László Lovász, and Alexander Schrijver. Geometric methods in combinatorial optimization. In *Progress in combinatorial optimization*, pages 167–183. Elsevier, 1984.

[GLS88] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*. Springer, 1988.

[Glu89] Efim Davydovich Gluskin. Extremal properties of orthogonal parallelepipeds and their applications to the geometry of banach spaces. *Mathematics of the USSR-Sbornik*, 64(1):85, 1989.

[GPRW20] Andrei Graur, Tristan Pollner, Vidhya Ramaswamy, and S. Matthew Weinberg. New query lower bounds for submodular function minimization. *11th Innovations in Theoretical Computer Science Conference, ITCS*, pages 64:1–64:16, 2020.

[Har08] Nicholas James Alexander Harvey. *Matchings, matroids and submodular functions*. PhD thesis, Massachusetts Institute of Technology, 2008.

[HPV17] Aicke Hinrichs, Joscha Prochno, and Jan Vybiral. Entropy numbers of embeddings of schatten classes. *Journal of Functional Analysis*, 273(10):3241–3261, 2017.

[HRS22] Samuel B Hopkins, Prasad Raghavendra, and Abhishek Shetty. Matrix discrepancy from quantum communication. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 637–648, 2022.

[HSSZ19] Christopher Harshaw, Fredrik Sävje, Daniel Spielman, and Peng Zhang. Balancing covariates in randomized experiments with the gram–schmidt walk design. *arXiv preprint arXiv:1911.03071*, 2019.

[IFF01] Satoru Iwata, Lisa Fleischer, and Satoru Fujishige. A combinatorial strongly polynomial algorithm for minimizing submodular functions. *Journal of the ACM (JACM)*, 48(4):761–777, 2001.

[IO09] Satoru Iwata and James B Orlin. A simple combinatorial algorithm for submodular function minimization. In *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*, pages 1230–1237. SIAM, 2009.

[Iwa03] Satoru Iwata. A faster scaling algorithm for minimizing submodular functions. *SIAM Journal on Computing*, 32(4):833–840, 2003.

[Iwa08] Satoru Iwata. Submodular function minimization. *Mathematical Programming*, 112(1):45, 2008.

[Jia21] Haotian Jiang. Minimizing convex functions with integral minimizers. In *SODA*. https://arxiv.org/pdf/2007.01445.pdf, 2021.

[Jia22] Haotian Jiang. Minimizing convex functions with rational minimizers. *ACM Journal of the ACM (JACM)*, 2022.

[JKS19] Haotian Jiang, Janardhan Kulkarni, and Sahil Singla. Online geometric discrepancy for stochastic arrivals with applications to envy minimization. *arXiv preprint arXiv:1910.01073*, 2019.

[JLLV21] He Jia, Aditi Laddha, Yin Tat Lee, and Santosh Vempala. Reducing isotropy and volume to kls: an o*(n 3 ψ 2) volume algorithm. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 961–974, 2021.

[JLSW20] Haotian Jiang, Yin Tat Lee, Zhao Song, and Sam Chiu-wai Wong. An improved cutting plane method for convex optimization, convex-concave games, and its applications. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 944–953. https://arxiv.org/pdf/2004.04250, 2020.

[Kha80] Leonid G Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72, 1980.

[KKT08] Pushmeet Kohli, M Pawan Kumar, and Philip HS Torr. p^3 & beyond: Move making algorithms for solving higher order functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1645–1656, 2008.

[KL19] Zohar Karnin and Edo Liberty. Discrepancy, coresets, and sketches in machine learning. In *Conference on Learning Theory*, pages 1975–1993. PMLR, 2019.

[KLS95] Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13(3-4):541–559, 1995.

[KLS97] Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an o*(n5) volume algorithm for convex bodies. *Random Structures & Algorithms*, 11(1):1–50, 1997.

[KLS20] Rasmus Kyng, Kyle Luh, and Zhao Song. Four deviations suffice for rank 1 matrices. *Advances in Mathematics*, 375:107366, 2020.

[KM87] Hermann König and Vitali D Milman. On the covering numbers of convex bodies. In *Geometrical Aspects of Functional Analysis*, pages 82–95. Springer, 1987.

[KT10] Pushmeet Kohli and Philip HS Torr. Dynamic graph cuts and their applications in computer vision. In *Computer Vision*, pages 51–108. Springer, 2010.

[KTE88] Leonid G Khachiyan, Sergei Pavlovich Tarasov, and I. I. Erlikh. The method of inscribed ellipsoids. *Soviet Math. Dokl*, 37(1):226–230, 1988.

[LB11] Hui Lin and Jeff Bilmes. Optimal selection of limited vocabulary speech corpora. In *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[Lev65] Anatoly Yur'evich Levin. An algorithm for minimizing convex functions. *Doklady Akademii Nauk*, 160(6):1244–1247, 1965.

[LLL82] Arjen Lenstra, Hendrik Lenstra, and László Lovász. Factoring polynomials with rational coefficients. *Math. Ann*, 261:515–534, 1982.

[LLSZ21] Troy Lee, Tongyang Li, Miklos Santha, and Shengyu Zhang. On the cut dimension of a graph. In *36th Computational Complexity Conference (CCC 2021)*, pages 15:1–15:35, 2021.

[LM15a] Shachar Lovett and Raghu Meka. Constructive discrepancy minimization by walking on the edges. *SIAM Journal on Computing*, 44(5):1573–1582, 2015.

[LM15b] Shachar Lovett and Raghu Meka. Constructive Discrepancy Minimization by Walking on the Edges. *SIAM J. Comput.*, 44(5):1573–1582, 2015.

[LMMS04] Richard J Lipton, Evangelos Markakis, Elchanan Mossel, and Amin Saberi. On approximately fair allocations of indivisible goods. In *Proceedings of the 5th ACM Conference on Electronic Commerce*, pages 125–131, 2004.

[LP86] Françoise Lust-Piquard. Inégalités de khintchine dans $c_p(1 < p < \infty)$. *C. R. Math. Acad. Sci. Paris*, 303(7):289–292, 1986.

[LPP91] Françoise Lust-Piquard and Gilles Pisier. Non commutative khintchine and paley inequalities. *Arkiv för matematik*, 29(1):241–260, 1991.

[LRR17] Avi Levy, Harishchandra Ramadas, and Thomas Rothvoss. Deterministic discrepancy minimization via the multiplicative weight update method. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 380–391. Springer, 2017.

[LRS11] Lap Chi Lau, Ramamoorthi Ravi, and Mohit Singh. *Iterative methods in combinatorial optimization*, volume 46. Cambridge University Press, 2011.

[LSV86] László Lovász, Joel Spencer, and Katalin Vesztergombi. Discrepancy of set-systems and matrices. *European Journal of Combinatorics*, 7(2):151–160, 1986.

[LSW15] Yin Tat Lee, Aaron Sidford, and Sam Chiu-wai Wong. A faster cutting plane method and its implications for combinatorial and convex optimization. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1049–1065. IEEE, 2015.

[Luc17] Brendan Lucier. An economic view of prophet inequalities. *SIGecom Exchanges*, 16(1):24–47, 2017.

[LV07] László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.

[Mat99] Jiri Matousek. *Geometric discrepancy: An illustrated guide*, volume 18. Springer Science & Business Media, 1999.

[McC05] S Thomas McCormick. Submodular function minimization. *Discrete Optimization*, 12:321–391, 2005.

[Mec07] Mark Meckes. On the spectral norm of a random Toeplitz matrix. *Electronic Communications in Probability*, 12(none):315 – 325, 2007.

[Meg83] Nimrod Megiddo. Towards a genuinely polynomial algorithm for linear programming. *SIAM Journal on Computing*, 12(2):347–353, 1983.

[Mek14] Raghu Meka. Discrepancy and beating the union bound. In *Windows On Theory, A Research Blog*, 2014.

[Min53] Hermann Minkowski. Geometrie der zahlen. *Chelsea, reprint*, 1953.

[MN12] Shanmugavelayutham Muthukrishnan and Aleksandar Nikolov. Optimal private halfspace counting via discrepancy. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1285–1292, 2012.

[MN15] Jiří Matoušek and Aleksandar Nikolov. Combinatorial discrepancy for boxes via the γ_2 norm. In *Proceedings of SoCG 2015*, pages 1–15, 2015.

[MNT14] Jiří Matoušek, Aleksandar Nikolov, and Kunal Talwar. Factorization norms and hereditary discrepancy. *CoRR*, abs/1408.1376, 2014.

[MSS15] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison—singer problem. *Annals of Mathematics*, pages 327–350, 2015.

[MV13] Daniele Micciancio and Panagiotis Voulgaris. A deterministic single exponential time algorithm for most lattice problems based on voronoi cell computations. *SIAM Journal on Computing*, 42(3):1364–1391, 2013.

[Nem94] Arkadi Nemirovski. On parallel complexity of nonsmooth convex optimization. *Journal of Complexity*, 10(4):451–463, 1994.

[New65] Donald J Newman. Location of the maximum on unimodal surfaces. *Journal of the ACM (JACM)*, 12(3):395–398, 1965.

[Nik17] Aleksandar Nikolov. Tighter bounds for the discrepancy of boxes and polytopes. *Mathematika*, 63(3):1091–1113, 2017.

[NN89] YE Nesterov and AS Nemirovskii. Self-concordant functions and polynomial time methods in convex programming. preprint, central economic & mathematical institute, ussr acad. *Sci. Moscow, USSR*, 1989.

[NNN12] Alantha Newman, Ofer Neiman, and Aleksandar Nikolov. Beck's three permutations conjecture: A counterexample and some consequences. In *Proceedings of FOCS 2012*, pages 253–262, 2012.

[NTZ13] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the approximate and sparse cases. In *Symposium on Theory of Computing (STOC)*, 2013.

[NV13] Hoi H Nguyen and Van H Vu. Small ball probability, inverse theorems, and applications. In *Erdős centennial*, pages 409–463. Springer, 2013.

[NY83] Arkadi Semenovič Nemirovski and David Borisovich Yudin. Problem complexity and method efficiency in optimization. 1983.

[Oli10] Roberto Oliveira. Sums of random hermitian matrices and an inequality by rudelson. *Electron. Commun. Probab.*, 15:no. 19, 203–212, 2010.

[Orl09] James B Orlin. A faster strongly polynomial time algorithm for submodular function minimization. *Mathematical Programming*, 118(2):237–251, 2009.

[OV20] Neil Olver and László A Végh. A simpler and faster strongly polynomial algorithm for generalized flow maximization. *Journal of the ACM (JACM)*, 67(2):1–26, 2020.

[Phi09] Jeff M Phillips. *Small and stable descriptors of distributions for geometric statistical problems*. PhD thesis, Duke University, 2009.

[Pis03] Gilles Pisier. *Introduction to operator space theory*. Cambridge University Press, 2003.

[Rot13] Thomas Rothvoss. Approximating bin packing within o (log opt* log log opt) bins. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 20–29. IEEE, 2013.

[Rot17] Thomas Rothvoss. Constructive discrepancy minimization for convex sets. *SIAM Journal on Computing*, 46(1):224–234, 2017.

[Roy14] Thomas Royen. A simple proof of the gaussian correlation conjecture extended to multivariate gamma distributions. *arXiv preprint arXiv:1408.1028*, 2014.

[RR20a] Victor Reis and Thomas Rothvoss. Linear size sparsifier and the geometry of the operator norm ball. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2337–2348. SIAM, 2020.

[RR20b] Victor Reis and Thomas Rothvoss. Linear size sparsifier and the geometry of the operator norm ball. In *Symposium on Discrete Algorithms (SODA)*, pages 2337–2348. SIAM, 2020.

[RSW18] Aviad Rubinstein, Tselil Schramm, and S. Matthew Weinberg. Computing Exact Minimum Cuts Without Knowing the Graph. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, pages 39:1–39:16, 2018.

[Sch87] Claus-Peter Schnorr. A hierarchy of polynomial time lattice basis reduction algorithms. *Theoretical computer science*, 53(2-3):201–224, 1987.

[Sch98] Alexander Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, 1998.

[Sch00] Alexander Schrijver. A combinatorial algorithm minimizing submodular functions in strongly polynomial time. *Journal of Combinatorial Theory, Series B*, 80(2):346–355, 2000.

[Sch03] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003.

[Sho77] Naum Z Shor. Cut-off method with space extension in convex programming problems. *Cybernetics*, 13(1):94–96, 1977.

[Šid67] Zbyněk Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967.

[Sma98] Steve Smale. Mathematical problems for the next century. *The mathematical intelligencer*, 20(2):7–15, 1998.

[Spe77] Joel Spencer. Balancing games. *Journal of Combinatorial Theory, Series B*, 23(1):68–74, 1977.

[Spe85] Joel Spencer. Six standard deviations suffice. *Transactions of the American mathematical society*, 289(2):679–706, 1985.

[Spe87] Joel H. Spencer. *Ten lectures on the probabilistic method*, volume 52. Society for Industrial and Applied Mathematics Philadelphia, 1987.

[SST97] Joel H. Spencer, Aravind Srinivasan, and Prasad Tetali. The discrepancy of permutation families. In *Proceedings of SODA*, 1997.

[SV13] Nikhil Srivastava and Roman Vershynin. Covariance estimation for distributions with $2 + \varepsilon$ moments. *The Annals of Probability*, 41(5):3081 – 3111, 2013.

[SZ20] Zhao Song and Ruizhe Zhang. Hyperbolic concentration, anti-concentration, and discrepancy. *arXiv preprint arXiv:2008.09593*, 2020.

[Tao12] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[Tar86] Eva Tardos. A strongly polynomial algorithm to solve combinatorial linear programs. *Operations Research*, 34(2):250–256, 1986.

[Tro15] Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1-2):1–230, 2015.

[Tro18] Joel A. Tropp. Second-order matrix concentration inequalities. *Applied and Computational Harmonic Analysis*, 44(3):700–736, 2018.

[TV85] William Thomson and Hal Varian. Theories of justice based on symmetry. *Social goals and social organizations: essays in memory of Elisha Pazner*, 126, 1985.

[Vai89] Pravin M Vaidya. A new algorithm for minimizing convex functions over convex sets. In *30th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 338–343, 1989.

[Vég17] László A Végh. A strongly polynomial algorithm for generalized flow maximization. *Mathematics of Operations Research*, 42(1):179–211, 2017.

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[Vu14] Van H Vu. *Modern aspects of random matrix theory*, volume 72. American Mathematical Society, 2014.

[VY96] Stephen A Vavasis and Yinyu Ye. A primal-dual interior point method whose running time depends only on the constraint matrix. *Mathematical Programming*, 74(1):79–120, 1996.

[Vyg03] Jens Vygen. A note on schrijver's submodular function minimization algorithm. *Journal of Combinatorial Theory, Series B*, 88(2):399–402, 2003.

[Wai19] Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

[Wal04] David Walnut. *An Introduction to Wavelet Analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Basel, 1 edition, 1 2004.

[YN76] David B Yudin and Arkadii S Nemirovski. Evaluation of the information complexity of mathematical programming problems. *Ekonomika i Matematicheskie Metody*, 12:128–142, 1976.

[Zou12] Anastasios Zouzias. A matrix hyperbolic cosine algorithm and applications. In *International Colloquium on Automata, Languages, and Programming*, pages 846–858. Springer, 2012.

[ZP19] David Zeng and Alexandros Psomas. Fairness-efficiency tradeoffs in dynamic fair division. *CoRR*, abs/1907.11672, 2019.