Sketches and Traces

by

Frank Ban

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Mathematics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Prof. Christos Papadimitriou, Co-chair
Prof. Luca Trevisan, Co-chair
Prof. Sanjam Garg
Prof. Nikhil Srivastava

Summer 2019

Sketches and Traces

Abstract

Sketches and Traces

by

Frank Ban

Doctor of Philosophy in Mathematics

University of California, Berkeley

Prof. Christos Papadimitriou, Co-chair

Prof. Luca Trevisan, Co-chair

In this dissertation, we study two problems that have the theme of extracting information from lower dimensional samples.

A number of recent works have studied algorithms for entrywise ℓ_p-*low rank approximation*, namely algorithms which given an $n \times d$ matrix A (with $n \geq d$), output a rank-k matrix B minimizing $\|A - B\|_p^p = \sum_{i,j} |A_{i,j} - B_{i,j}|^p$ when $p > 0$; and $\|A - B\|_0 = \sum_{i,j} [A_{i,j} \neq B_{i,j}]$ for $p = 0$, where $\|A - B\|_0$ denotes the number of entries (i, j) for which $A_{i,j} \neq B_{i,j}$.

For $p = 1$, this is often considered more robust than the SVD, while for $p = 0$ this corresponds to minimizing the number of disagreements, or robust PCA. This problem is known to be NP-hard for $p \in \{0, 1\}$, already for $k = 1$, and while there are polynomial time approximation algorithms, their approximation factor is at best poly(k). It was left open if there was a polynomial-time approximation scheme (PTAS) for ℓ_p-approximation for any $p \geq 0$. We show the following:

1. On the algorithmic side, for $p \in (0, 2)$, we use a technique called *sketching* to give the first $n^{\text{poly}(k/\varepsilon)}$ time $(1 + \varepsilon)$-approximation algorithm. For $p = 0$, there are various problem formulations, a common one being the binary setting for which $A \in \{0, 1\}^{n \times d}$ and $B = U \cdot V$, where $U \in \{0, 1\}^{n \times k}$ and $V \in \{0, 1\}^{k \times d}$. For this setting, we obtain an algorithm with time $n \cdot d^{\text{poly}(k/\varepsilon)}$.

2. On the hardness front, for $p \in (1, 2)$, we show under the Small Set Expansion Hypothesis and Exponential Time Hypothesis (ETH), there is no constant factor approximation algorithm running in time 2^{k^δ} for a constant $\delta > 0$, showing an exponential dependence on k is necessary. We also show for finite fields of

constant size, under the ETH, that any fixed constant factor approximation algorithm requires 2^{k^δ} time for a constant $\delta > 0$.

Population recovery is the problem of learning an unknown distribution over an unknown set of n-bit strings, given access to independent *traces* from the distribution that have been independently corrupted according to some noise channel. Recent work has intensively studied such problems both for the bit-flip noise channel and for the erasure noise channel.

In this dissertation we initiate the study of population recovery under the *deletion channel*, in which each bit b is independently *deleted* with some fixed probability and the surviving bits are concatenated and transmitted. This is a far more challenging noise model than bit-flip noise or erasure noise; indeed, even the simplest case in which the population is of size 1 (corresponding to a trivial probability distribution supported on a single string) corresponds to the *trace reconstruction* problem, which is a challenging problem that has received much recent attention.

In this work we give algorithms and lower bounds for population recovery under the deletion channel when the population size is some value $\ell > 1$. As our main sample complexity upper bound, we show that for any population size $\ell = o(\log n / \log \log n)$, a population of ℓ strings from $\{0,1\}^n$ can be learned under deletion channel noise using $2^{n^{1/2+o(1)}}$ samples. On the lower bound side, we show that at least $n^{\Omega(\ell)}$ samples are required to perform population recovery under the deletion channel when the population size is ℓ, for all $\ell \leq n^{1/2-\varepsilon}$.

To my family and to Joy.

Contents

Chapter 1

Overview

Massive datasets in areas such as machine learning, numerical linear algebra, and computational biology require algorithms that can deal with high dimensions, offer robustness, and tolerate noise. A common theme in the design of these algorithms is the use of randomized methods such as sketching, sampling, and hashing.

In this dissertation we apply these techniques to get improved results in two problems in theoretical computer science: low rank approximation under a non-Frobenius norm (an NP-complete problem) and population recovery under the deletion channel (a generalization of the trace reconstruction problem).

Parts I (on low rank approximations) and II (on population recovery) of this dissertation are based on [3] and [4] respectively.

1.1 Low Rank Approximations

Low rank approximation is a common way of compressing a matrix via dimensionality reduction. The goal is to replace a given $n \times d$ matrix A by a rank-k matrix A' that approximates A well, in the sense that $\|A - A'\|$ is small for some measure $\|.\|$. Since we can write the rank-k matrix A' as $U \cdot V$, where U is $n \times k$ and V is $k \times d$, it suffices to store the $k(n+d)$ entries of U and V, which is a significant reduction compared to the nd entries of A. Furthermore, computing $A'x = U(Vx)$ takes time $O(k(n+d))$, which is much less than the time $O(nd)$ for computing Ax.

Low rank approximation is extremely well studied, see the surveys [47, 60, 91] and the many references therein. In this Part, we study the following two variants of entrywise ℓ_p-low rank approximation. Given a matrix A and an integer k, one seeks to find a rank-k matrix A', minimizing $\|A - A'\|_p^p = \sum_{i,j} |A_{i,j} - A'_{i,j}|^p$ when $p > 0$

and $\|A - A'\|_0 = \sum_{i,j}[A_{i,j} \neq A'_{i,j}]$ for $p = 0$, where $[\cdot]$ is the Iverson bracket, that is, $\|A - A'\|_0$ denotes the number of entries (i, j) for which $A_{i,j} \neq A'_{i,j}$.

When $p = 2$, this coincides with the Frobenius norm error measure, which can be solved in polynomial time using the singular value decomposition (SVD); see also [91] for a survey of more efficient algorithms based on the technique of linear sketching.

Recently there has been considerable interest in obtaining algorithms for $p \neq 2$. For $0 \leq p < 2$, this error measure is often considered more robust than the SVD, since one pays less attention to noisy entries as one does not square the differences, but instead raises the difference to a smaller power. Conversely, for $p > 2$, this error measure pays more attention to outliers, and $p = \infty$ corresponds to a guarantee on each entry. This problem was shown to be NP-hard for $p \in \{0, 1\}$ [22, 34, 69].

ℓ_p-Low Rank Approximation for $p > 0$. A number of initial algorithms for ℓ_1-low rank approximation were given in [13–15, 49, 50, 53, 56, 63–65, 67, 75, 95]. There is also related work on robust PCA [16, 17, 72, 73, 92, 94] and measures which minimize the sum of Euclidean norms of rows [21, 29, 32, 33, 85], though neither directly gives an algorithm for ℓ_1-low rank approximation. Song et al. [86] gave the first approximation algorithms with provable guarantees for entrywise ℓ_p-low rank approximation for $p \in [1, 2)$. Their algorithm provides a $\text{poly}(k \log n)$ approximation and runs in polynomial time, that is, the algorithm outputs a matrix B for which $\|A-B\|_p \leq \text{poly}(k \log n) \min_{\text{rank-}k\ A'} \|A-A'\|_p$. This was generalized by Chierichetti et al. [18] to ℓ_p-low rank approximation, for every $p \geq 1$, where we also obtained a $\text{poly}(k \log n)$ approximation in polynomial time.

In Song et al. [86] it is also shown that if A has entries bounded by $\text{poly}(n)$ then an $O(1)$ approximation can be achieved, albeit in $n^{\text{poly}(k)}$ time. This algorithm depends inherently on the triangle inequality and as a result the constant factor of approximation is greater than 3. Improving this constant of approximation requires techniques that break this triangle inequality barrier. This is a real barrier, since the algorithm of [86] is based on a row subset selection algorithm, and there exist matrices for which any subset of rows contains at best a $2(1-\Theta(1/n))$-approximation (Theorem G.8 of [86]), which we discuss more below.

ℓ_0-Low Rank Approximation. When $p = 0$, one seeks a rank-k matrix A' for which $\|A - A'\|_0$ is as small as possible, where for a matrix C, $\|C\|_0$ denotes the number of non-zero entries of C. Thus, in this case, we are trying to minimize the number of disagreements between A and A'. Since A' has rank k, we can write it as $U \cdot V$ and we seek to minimize $\|A - U \cdot V\|_0$. This was studied by Bringmann et al. [12] when A, U, and V are matrices over the reals and $U \cdot V$ denotes the standard

matrix product, and the work of [12] provides a poly$(k \log n)$ bicriteria approximation algorithm. See also earlier work for $k = 1$ giving a 2-approximation [44, 84]. ℓ_0-low rank approximation is also well-studied when A, U, and V are each required to be binary matrices. In this case, there are a number of choices for the ground field (or, more generally, semiring). Specifically, for $A' = U \cdot V$ we can write the entry $A'_{i,j}$ as the inner product of the i-th row of U with the j-th column of V – and the specific inner product function $\langle ., . \rangle$ depends on the ground field.

Besides the abovementioned upper bounds, which coincide with all of these models when $k = 1$, the only other algorithm we are aware of is by Dan et al. [22], who for arbitrary k presented an $n^{O(k)}$-time $O(k)$-approximation over $\mathbb{F}_2$, and an $n^{O(k)}$-time $O(2^k)$-approximation over the Boolean semiring.

Although ℓ_p-low rank approximation is NP-hard for $p \in \{0, 1\}$, a central open question is if $(1 + \varepsilon)$-approximation is possible, namely: *Does ℓ_p-low rank approximation have a polynomial time approximation scheme (PTAS) for any constant k and ε?*

We answer this question in the affirmative in Part I and prove lower bounds as well.

1.2 Population Recovery

In recent years the unsupervised learning problem of *population recovery* has emerged as a significant focus of research attention in theoretical computer science [6, 26, 28, 31, 59, 70, 77, 90]. In the population recovery problem there is an unknown distribution $\mathbf{X}$ over an unknown set of n-bit strings from $\{0, 1\}^n$, and the learner's job is to reconstruct a high-accuracy approximation of $\mathbf{X}$ given access to noisy independent draws from $\mathbf{X}$ (so each data point which the learning algorithm receives is independently generated as follows: an n-bit string is drawn from $\mathbf{X}$ and corrupted by some noise process, and the result is provided to the learning algorithm). The two noise models which have chiefly been studied to date are the *bit-flip* noise model, in which each coordinate is independently flipped with some fixed probability, and the *erasure* noise model, in which each coordinate is independently replaced by '?' with some fixed probability.

Since the population recovery problem was first introduced in [31, 90], a number of positive results and lower bounds have been obtained for different variants of the problem. In one popular version of the problem [28, 70, 77], for a particular noise model (bit-flip or erasure) the distribution $\mathbf{X}$ may be an arbitrary distribution over $\{0, 1\}^n$, and the goal is to learn the distribution $\mathbf{X}$ with respect to ℓ_∞ distance (i.e. to output a list of strings $x^1, \ldots, x^r \in \{0, 1\}^n$ and associated weights $\tilde{\mathbf{X}}(x^i)$ such that

$|\mathbf{X}(x^i) - \tilde{\mathbf{X}}(x^i)| \leq \varepsilon$ for all $i \in [r]$ and $\mathbf{X}(x) \leq \varepsilon$ for all $x \in \{0,1\}^n \setminus \{x^1, \ldots, x^r\}$). In another well-studied version of the problem [26, 59, 90], which is closely related to the problems we shall consider, the distribution $\mathbf{X}$ is promised to be supported on at most ℓ strings in $\{0,1\}^n$ (i.e. the "population size" is promised to be at most ℓ), and the goal is to output a hypothesis distribution $\tilde{\mathbf{X}}$ over $\{0,1\}^n$ which has total variation distance at most ε from $\mathbf{X}$. Significant progress has been made on determining the sample complexity of population recovery for both of these variants under the bit-flip and erasure noise models; we refer the interested reader to [26, 28, 77] for the current state of the art.

This work: Population recovery from the deletion channel and its relation to trace reconstruction. In both the bit-flip noise model and the erasure noise model, all of the challenge in the population recovery problem stems from the fact that given a noisy draw from $\mathbf{X}$ it is *a priori* not clear which element of $\mathbf{X}$'s support was corrupted by noise to produce the noisy draw. Putting it another way, if the population size is promised to be $\ell = 1$, then under either of these two noise models it is trivially easy to learn a single unknown string from noisy examples.

In this work we study population recovery under the *deletion* noise model, which is far more challenging to handle than either bit-flip noise or erasure noise. The deletion channel is defined as follows: when a string x is passed through the deletion channel with deletion parameter δ, each coordinate x_i is independently deleted with probability δ, the surviving coordinates are concatenated, and the resulting string (of length $\boldsymbol{n}' \leq n$, where $\boldsymbol{n}'$ is distributed as $\mathrm{Bin}(n, 1-\delta)$) is the output of the noise process. Intuitively, the deletion channel is challenging because given a received word obtained by passing x through the δ-deletion channel (often referred to as a *trace* of x, and denoted by $\boldsymbol{z} \leftarrow \mathrm{Del}_\delta(x)$), it is not clear which coordinate of x gave rise to which coordinate of $\boldsymbol{z}$. Indeed, in contrast with the bit-flip and erasure noise models, even if the population size is guaranteed to be $\ell = 1$, the problem of recovering a single unknown string from independent traces is a well-known and challenging open problem, known as the *trace reconstruction problem* [7, 27, 38, 41, 42, 48, 57, 58, 66, 71, 76, 89].

There are several motivations for the study of population recovery under the deletion noise model. One motivation is the considerable recent research interest both in the trace reconstruction problem (the $\ell = 1$ case of population recovery under the deletion channel) and in population recovery problems under bit-flip and erasure models. Further motivation comes from potential relevance of the deletion channel population recovery problem both to recovery problems in computational biology and to other topics such as DNA data storage. Regarding biological recovery problems, considering population recovery (the $\ell > 1$ case) rather than trace reconstruction (the

$\ell = 1$ case) relaxes the potentially unrealistic assumption that all of the received samples (of a protein sequence, DNA sequence, etc.) are derived from a single unknown target sequence rather than from multiple unknown sequences. Heuristic algorithms for population recovery-type problems have also been applied to DNA storage (see e.g., [93] and [74]). In these settings, each string in the population comes from a DNA sequence and the noisy channel can inflict a variety of errors including bit-flips and deletions.

Thus, we feel that the time is ripe for a theoretical study of population recovery under the challenging deletion model. In this Part we initiate such a study, obtaining sample complexity upper and lower bounds when the population is of size $\ell > 1$. Before describing our results for populations of size ℓ (equivalently, target distributions supported on at most ℓ strings), we first recall known upper and lower bounds for the trace reconstruction problem ($\ell = 1$) below.

Known bounds on trace reconstruction. The trace reconstruction problem was raised more than fifteen years ago [7, 57, 58], though in fact some variants of the problem go back at least to the 1970s [45]. The first algorithm that provably succeeds with high probability in reconstructing an arbitrary $x \in \{0,1\}^n$ using subexponentially many traces is due to Mitzenmacher et al. [42], who showed that $2^{\tilde{O}(\sqrt{n})}$ many traces suffice for any constant deletion rate δ bounded away from 1. This result was improved in recent simultaneous and independent works of De et al. [27] and Nazarov and Peres [71]; these papers each showed that for any constant δ bounded away from 1, at most $2^{O(n^{1/3})}$ traces suffice to reconstruct any $x \in \{0,1\}^n$.[1]

Due to the seeming difficulty of the worst-case trace reconstruction problem (reconstructing an arbitrary $x \in \{0,1\}^n$), an average-case version of the problem (reconstructing a randomly chosen string $x \in \{0,1\}^n$), which turns out to be significantly easier in terms of sample complexity, has also received considerable attention. A number of early works [7, 48, 89] gave efficient algorithms that succeed for trace reconstruction of almost all $x \in \{0,1\}^n$ when the deletion rate δ is sufficiently low ($o_n(1)$ as a function of n). In [42] Mitzenmacher et al. gave an algorithm which uses $\mathrm{poly}(n)$ traces to perform average-case trace reconstruction when the deletion rate δ is at most some sufficiently small constant. Recently the best results on average-case trace reconstruction have been significantly strengthened in works of Peres and Zhai [76] and Holden, Pemantle and Peres [41] which build on the worst-case trace reconstruction results of [27, 71]. The latter of these papers [41] gives an algorithm which uses $\exp((\log n)^{1/3})$ traces to reconstruct a random $x \in \{0,1\}^n$ when the deletion rate is any constant bounded away from 1.

[1] Hartung, Holden and Peres [38] have recently extended this result to certain more general regimes where there can be different deletion probabilities for different coordinates and symbols.

In terms of lower bounds, it is easy to see that if the deletion rate δ is at least some positive constant, then until $\Omega(\log n)$ draws have been received there will be some bits of the target string x about which no information has been received. Improving on this simple $\Omega(\log n)$ lower bound, McGregor et al. [66] established a sample complexity lower bound of $\Omega(n)$ traces for any constant deletion rate. This was recently improved by Holden and Lyons [40] to $\tilde{\Omega}(n^{5/4})$.

Summarizing, for any constant deletion probability $0 < \delta < 1$ there is currently an exponential gap between the best lower bound of $\tilde{\Omega}(n^{5/4})$ samples and the best upper bound of $2^{O(n^{1/3})}$ samples for trace reconstruction of an arbitrary string $x \in \{0,1\}^n$.

In this work, we provide the first known upper and lower bounds for population recovery over the deletion channel.

Outline: Our work on low rank approximation variants is covered in Part I. Our work on population recovery over the deletion channel is covered in Part II.

Part I

Low Rank Approximations

Chapter 2

Introduction

2.1 Our Results

We give the first PTAS for ℓ_p-low rank approximation for $0 \leq p < 2$ in the unit cost RAM model of computation. We also give time lower bounds, assuming the Exponential Time Hypothesis (ETH) [43] and in some cases the Small Set Expansion Hypothesis [78], providing evidence that an exponential dependence on k, for $p > 0$, and a doubly-exponential dependence on k, for $p = 0$, may be necessary.

Algorithms

We first formally define the problem we consider for $0 < p < 2$. We may assume w.l.o.g. that $n \geq d$, and thus the input size is $O(n)$.

Definition 2.1.1. *(**Entrywise ℓ_p-Rank-k Approximation:**) Given an $n \times d$ matrix A with integer entries bounded in absolute value by $\mathrm{poly}(n)$, and a positive integer k, output matrices $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{k \times d}$ minimizing $\|A - UV\|_p^p := \sum_{i=1,\ldots,n,j=1,\ldots,d} |A_{i,j} - (U \cdot V)_{i,j}|^p$. An algorithm for Entrywise ℓ_p-Rank-k Approximation is an α-approximation if it outputs U and V for which $\|A - UV\|_p^p \leq \alpha \cdot \min_{U' \in \mathbb{R}^{n \times k}, V' \in \mathbb{R}^{k \times d}} \|A - U'V'\|_p^p$.*

Our main result for $0 < p < 2$ is as follows.

Theorem 2.1.1 (PTAS for $0 < p < 2$)**.** *For any $p \in (0, 2)$ and constant $\varepsilon \in (0, 1)$, there is a $(1 + \varepsilon)$-approximation algorithm to Entrywise ℓ_p-Rank-k Approximation running in time $n^{\mathrm{poly}(k/\varepsilon)}$.*

For any constants $k \in \mathbb{N}$ and $\varepsilon > 0$, Theorem 2.1.1 computes in polynomial time a $(1+\varepsilon)$-approximate solution to Entrywise ℓ_p-Rank-k Approximation. This significantly strengthens the approximation guarantees in [18, 86].

We next consider the case $p = 0$. In this setting, the base field is the finite field $\mathbb{F}_q$ (where q is a prime power and A, U, and V have entries belonging to $\mathbb{F}_q$). We obtain an algorithm running in time $n \cdot d^{\mathrm{poly}(k/\varepsilon)}$, which is an improvement for certain super-constant values of k and ε. We formally define the problem and state our result next.

Definition 2.1.2. *(**Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$:**) Given an $n \times d$ matrix A with entries that are in $\mathbb{F}_q$ for any constant q, and a positive integer k, output matrices $U \in \mathbb{F}_q^{n\times k}$ and $V \in \mathbb{F}_q^{k\times d}$ minimizing $\|A - UV\|_0$. An algorithm for Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$ is an α-approximation if it outputs matrices U and V such that $\|A - UV\|_0 \leq \alpha \cdot \min_{U' \in \mathbb{F}_q^{n\times k}, V' \in \mathbb{F}_q^{k\times d}} \|A - U'V'\|_0$.*

Our main result for Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$ is the following:

Theorem 2.1.2 ($\mathbb{F}_q$ PTAS for $p = 0$)**.** *For $\varepsilon \in (0, 1)$ there is a $(1+\varepsilon)$-approximation algorithm to Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$ running in time $n \cdot d^{\mathrm{poly}(k/\varepsilon)}$.*

Hardness

We first obtain conditional time lower bounds for Entrywise ℓ_p-Rank-k Approximation for $p \in (1, 2)$. Our results assume the Small Set Expansion Hypothesis (SSEH). Originally conjectured by Raghavendra and Steurer [78], it is still the only assumption that implies strong hardness results for various graph problems such as Uniform Sparsest Cut [80] and Bipartite Clique [61]. Assuming this hypothesis, we rule out *any* constant factor approximation α.

Theorem 2.1.3 (Hardness for Entrywise ℓ_p-Rank-k Approximation)**.** *Fix $p \in (1, 2)$ and $\alpha > 1$. Assuming the Small Set Expansion Hypothesis, there is no α - approximation algorithm for Entrywise ℓ_p-Rank-k Approximation that runs in time* $\mathrm{poly}(n)$.

Consequently, additionally assuming the Exponential Time Hypothesis, there exists

$$\delta := \delta(p, \alpha) > 0$$

such that there is no α-approximation algorithm for Entrywise ℓ_p-Rank-k Approximation that runs in time 2^{n^δ}.

This shows that assuming the SSEH and the ETH, any constant factor approximation algorithm needs at least a subexponential dependence on n (and thus k).

We also prove hardness of approximation results for $p \in (2, \infty)$ (see Theorem 5.0.2) without the SSEH. They are the first hardness results for Entrywise ℓ_p-Rank-k Approximation other than $p = 0, 1$.

Next we obtain conditional lower bounds for Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$ for any fixed q:

Theorem 2.1.4 (Hardness for Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$)**.** *Let $\mathbb{F}_q$ be a finite field and $\alpha > 1$. Assuming $\mathsf{P} \neq \mathsf{NP}$, there is no α-approximation algorithm for Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$ that runs in time poly(n).*

Consequently, assuming the Exponential Time Hypothesis, there exists $\delta := \delta(\alpha) > 0$ such that there is no α-approximation algorithm for Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$ that runs in time 2^{n^δ}.

This shows that assuming the ETH, any constant factor approximation algorithm needs at least a subexponential dependence on n (and thus k).

Additional Results

We obtain several additional results on ℓ_p-low rank approximation. We summarize our results below and defer the details to Chapter 6.

ℓ_p-low rank approximation for $p > 2$ Let g be a standard Gaussian random variable and let $\gamma_p := \mathbf{E}_g[|g|^p]^{1/p}$. We note that $\gamma_p > 1$, for any $p > 2$. Then, under ETH no $(\gamma_p^p - \varepsilon)$-approximation algorithm runs in time $O(2^{k^\delta})$. On the algorithmic side, we give a simple $(3 + \varepsilon)$-approximation algorithm running in time $n^{\mathrm{poly}(k/\varepsilon)}$.

Weighted ℓ_p-low rank approximation for $0 < p < 2$ We also generalize Theorem 2.1.1 to the following weighted setting. Given a matrix A, an integer k and a rank-r matrix W, we seek to find a rank-k matrix A' such that

$$\|W \circ (A - A')\|_p^p \leq (1 + \varepsilon) \min_{\text{rank-}k\ A_k} \|W \circ (A - A_k)\|_p^p.$$

Our algorithm runs in time $n^{r \cdot \mathrm{poly}(k/\varepsilon)}$. We defer the details to Theorem 6.2.2.

2.2 Our Techniques

We give an overview of our techniques, separating them into those for our algorithms for $0 < p < 2$, those for our algorithms for $p = 0$, and those for our hardness proofs.

Algorithms for $0 < p < 2$

We illustrate the techniques for $p = 1$; the algorithms for other $p \in (0, 2)$ follow similarly. Consider a target rank k. One of the surprising aspects of our $(1 + \varepsilon)$-approximation result is that for $p = 1$, it breaks a potential lower bound from [86]. Indeed, in Theorem G.8, they construct $(n-1) \times n$ matrices A such that the closest rank-k matrix B in the row span of A provides at best a $2(1-\Theta(1/n))$-approximation to A!

This should be contrasted with $p = 2$, for which it is well-known that for any A there exists a subset of k/ε rows of A containing a k-dimensional subspace in its span which is a $(1 + \varepsilon)$-approximation (these are called column subset selection algorithms; see [91] for a survey). In fact, for $p = 1$, all known algorithms [18, 86] find a best k-dimensional subspace in either the span of the rows or of the columns of A, and thus provably cannot give better than a 2-approximation. To bypass this, we therefore critically need to leave the row space and column space of A.

Our starting point is the "guess a sketch" technique of [81], which was used in the context of weighted low rank approximation. Let us consider the optimization problem $\min_V \|U^*V - A\|_1$, where U^* is a left factor of an optimal ℓ_1-low rank approximation for A. Suppose we could choose a *sketching matrix* S with a small number r of rows for which $\|SU^*V - SA\|_1 = (1 \pm \varepsilon)\|U^*V - A\|_1$ for all V. Then, if we somehow knew U^*, we could optimize for V in the sketched space to find a good right factor V.

Of course we do not know U^*, but if S had a small number r of rows, then we could consider instead the $\|\cdot\|_{1,2}$-norm optimization problem $\min_V \|SU^*V - SA\|_{1,2}$, where for a matrix C, $\|C\|_{1,2}$ is defined as $\sum_{i=1}^{d} \|C_{:,i}\|_2$, the sum of the $\|\cdot\|_2$-norms of its columns. The solution V to $\min_V \|SU^*V - SA\|_{1,2}$ is a $\sqrt{r}$-approximation to the original problem $\min_V \|SU^*V - SA\|_1$.

In the $\|\cdot\|_{1,2}$ norm, the solution V can be written in terms of the so-called normal equations for regression, namely, $V = (SU^*)^\dagger SA$, where $C^\dagger$ denotes the Moore-Penrose pseudoinverse of C. The key property exploited in [86] is then that although we do not know U^*, $(SU^*)^\dagger SA$ is a k-dimensional subspace in the row span of SA providing a $\sqrt{r}$-approximation, and one does know SA. This line of reasoning ultimately leads to a poly(k)-approximation.

The approach above fails to give a $(1 + \varepsilon)$-approximation for multiple reasons: (1) we may not be able to find a $(1+\varepsilon)$-approximation from the row span of A, and (2) we lose a $\sqrt{r}$ factor when we switch to the $\|\cdot\|_{1,2}$ norm.

Instead, suppose we were instead just to guess all the values of SU^*. These values might be arbitrary real numbers, but observe that we can assume there is an optimal solution U^*V^* for which V^* is a so-called ℓ_1-well conditioned basis, which loosely

speaking means that $\|yV^*\|_1 \approx \|y\|_1$ for any row vector y. Also, we can show that if $U^*V^* \neq A$, then $\|U^*V^* - A\|_1 \geq n^{-\Theta(k)}$. Furthermore, we can assume that the entries of A are bounded by poly(n). These three facts allow us to round the entries of U^* to an integer multiple of $n^{-\Theta(k)}$ of absolute value at most $n^{O(k)}$. Now suppose we could also discretize the entries of S to multiples of $n^{-\Theta(k)}$ and of absolute value at most $n^{O(k)}$. Then we would actually be able to guess the correct SU^* after $n^{\Theta(k^2 r)}$ tries, where recall r is the number of rows of S. We will show below that r can be poly(k/ε), so this will be within our desired running time.

In general, if $\mathcal{A}(Ux) = (1\pm\varepsilon)\|Ux\|$ for all x, then we say that $\mathcal{A}$ defines a subspace embedding. At this point, we can use the triangle inequality to get a constant factor approximation. If S is a subspace embedding, then

$$\begin{aligned}&\|U^*V - A\|_1\\ &\leq \|U^*(V - V^*)\|_1 + \|U^*V^* - A\|_1\\ &\leq (1 + O(\varepsilon))\|SU^*(V - V^*)\|_1 + \|U^*V^* - A\|_1\end{aligned}$$

and

$$\|SU^*(V - V^*)\|_1 \leq \|SU^*V - SA\|_1 + \|SU^*V^* - SA\|_1$$

so by taking V to be a minimizer for $\|SU^*V - SA\|_1$ we can get an approximation factor close to 3. The triangle inequality was useful here because S had a small distortion on the subspace defined by U^*. To improve this result, we would need a mapping that has small distortion on the *affine space* defined by $U^*V - A$, as V varies.

Given SU^* and SA, if in fact S has the property that $\|SU^*V - SA\|_1 = (1 \pm \varepsilon)\|U^*V - A\|_1$ for all V, then we will be in good shape. At this point we can solve for the optimal V to $\min_V \|SU^*V - SA\|_1$ by solving an ℓ_1-regression problem using linear programming. Notice that unlike [81], the approach described above does not create "unknowns" to represent the entries of SU^* and set up a polynomial system of inequalities. For Frobenius norm error, this approach is feasible because $\|SU^*V - SA\|_F^2 = \sum_{i=1}^n \|SU^*V_{:,i} - SA_{:,i}\|_F^2$ can be minimized over each column $V_{:,i}$ using the normal equations for regression. However, we do not know how to set up a polynomial system of inequalities for ℓ_1-error (which define V in terms of the SU^* variables).

Unfortunately the approach above is fatally flawed; there is no known sketching matrix S with a small number r of rows for which $\|SU^*V - SA\|_1 = (1\pm\varepsilon)\|U^*V - A\|_1$ for all V. Instead, we adapt a "median-based" embedding with a non-standard subspace embedding analysis that appeared in the context of sparse recovery [2]. In Lemma F.1 of that paper, it is shown that if L is a d-dimensional subspace of $\mathbb{R}^n$, and S is an $r \times n$ matrix of i.i.d. standard Cauchy random variables for

$r = O(d\varepsilon^{-2}\log(d/\varepsilon))$, then with constant probability, $(1-\varepsilon)\|x\|_1 \leq \text{med}(Sx) \leq (1+\varepsilon)\|x\|_1$ simultaneously for all $x \in L$. Here for a vector y, $\text{med}(y)$ denotes the median of absolute values of its entries. For a matrix M, $\text{med}(M)$ denotes the sum of the medians of its columns $\sum_i \text{med}(M_{:,i})$.

In our context, this gives us that for a fixed column $A_{:,i}$ of A and i-th column $V_{:,i}$ of V, if S is an i.i.d. Cauchy matrix with $O(k\varepsilon^{-2}\log(k/\varepsilon))$ rows, then with constant probability $\text{med}(SU^*V_{:,i} - SA_{:,i}) = (1\pm\varepsilon)\|U^*V_{:,i} - A_{:,i}\|_1$ for all vectors $V_{:,i}$. Since $V_{:,i}$ is only k-dimensional, and one can show that its entries can be taken to be integer multiples of $n^{-\text{poly}(k)}$ bounded in absolute value by $n^{\text{poly}(k)}$, we can enumerate over all $V_{:,i}$ and find the best solution. We need, however, to adapt the argument in [2] to argue that if we take a $(1/2\pm\varepsilon)$-quantile (rather than a median), we still obtain a subspace embedding. We do this in Lemma 3.0.6 and explain why this modification is crucial for the argument below.

Unfortunately, this still does not work. The issue is that S succeeds only with constant probability in achieving $\text{med}(SU^*V_{:,i} - SA_{:,i}) = (1\pm\varepsilon)\|U^*V_{:,i} - A_{:,i}\|_1$ for all vectors $V_{:,i}$. Call this property, of an index $i \in [n] := \{1, 2, \ldots, n\}$, *good*. A naïve amplification of the probability to $1 - 1/n$ would allow us to union bound over all i, but this would require S to have $\Omega(\log n)$ rows. At this point though, we would not obtain a PTAS since enumerating the entries of SU^* would take $n^{\Omega(\log n)}$ time. Nor can we use different S for different columns of A, since we may guess different SU^* for different i and not obtain a consistent solution V.

Before proceeding, we first relax the requirement that $\text{med}(SU^*V - SA) = (1\pm\varepsilon)\|U^*V - A\|_1$ for all V. We only need $\text{med}(SU^*V - SA) \geq (1-\varepsilon)\|U^*V - A\|_1$ for all V, and $\text{med}(SU^*V^* - SA) \leq (1+\varepsilon)\|U^*V^* - A\|_1$ for the fixed optimum U^*V^*. We can prove $\text{med}(SU^*V^* - SA) \leq (1+\varepsilon)\min_V \|U^*V - A\|_1$ by using tail bounds for a Cauchy random variable; we do so in Lemma 3.0.5.

Moreover, we next argue that it suffices to have the properties: i) a $(1-\text{poly}(\varepsilon/k))$-fraction of columns are good, and ii) the error introduced by bad columns is small. We can achieve (i) by increasing the number of rows of S by a $\log(k/\varepsilon)$ factor, which still allows for an enumeration in time $n^{\text{poly}(k/\varepsilon)}$. The main issue is to control the error from bad columns. In particular, it is possible to have a matrix V and a column $A_{:,i}$ such that $\|U^*V_{:,i} - A_{:,i}\|_1$ is large and yet $\text{med}(SU^*V_{:,i} - SA_{:,i})$ is small, which results in accepting a bad solution V. While for an average matrix V, the expected value of $\sum_{i \text{ is bad}} \|U^*V_{:,i} - A_{:,i}\|_1$ is small, we need to argue that this holds for every matrix V.

In order to control the error from bad columns, we first show that $\text{med}(SU^*V^* - SA) = (1\pm\varepsilon)\|U^*V^* - A\|_1$ for the fixed matrix $U^*V^* - A$, and then we demonstrate that the total contribution to $\|U^*V^* - A\|_1$ from bad columns, is small. We show the latter using Markov's bound for the fixed matrix $U^*V^* - A$. Combining this with the

former, yields that the total contribution of $\text{med}(SU^*V^*_{:,i} - SA_{:,i})$ to $\|SU^*V^* - SA\|_1$ from bad columns (in the original, unsketched space) is small.

We convert the preceding argument for bad columns of the fixed matrix $U^*V^* - A$, into an argument for bad columns of a general matrix $U^*V - A$. Inspired by ideas for $\|\cdot\|_{1,2}$ norm, established in [21], we partition the bad columns of a given matrix V into classes, using the following measurement, which differs substantially from [21]. We look at *quantiles* to handle the median operator, and we say that a bad column $A_{:,i}$ is *large* if

$$\|U^*V_{:,i} - A_{:,i}\|_1 \geq \frac{1}{\varepsilon}\left(\|U^*V^*_{:,i} - A_{:,i}\|_1 + \frac{1}{1-O(\varepsilon)}q_{1-\varepsilon/2}(S(U^*V^* - A)_{:,i})\right), \quad (2.1)$$

where $q_{1-\varepsilon/2}$ is the $(1-\varepsilon/2)$-th quantile of coordinates of column $S(U^*V^* - A)_{:,i}$ arranged in order of non-increasing absolute values. Otherwise, a bad column $A_{:,i}$ is *small*.

We show that small bad columns can be handled by applying the preceding argument for the fixed matrix $U^*V^* - A$, since intuitively, the error they introduce is dominated by the contribution of the corresponding columns of matrix $U^*V^* - A$, and we can control this contribution.

Our analysis for the large bad columns uses a different approach, which we summarize in Claim 4.1.2. The key insight is to use the additivity of a sketch matrix S, and to write

$$S(U^*V - A)_{:,i} = S(U^*V - U^*V^*)_{:,i} + S(U^*V^* - A)_{:,i}. \quad (2.2)$$

Then, by applying our "robust" version (Lemma 3.0.3) of median-based subspace embedding [2], it follows that at least a $(1/2+\varepsilon)$-fraction of the entries of column vector $S(U^*V - U^*V^*)_{:,i}$ have absolute value at least

$$\begin{aligned} & (1-O(\varepsilon)) \cdot \|U^*(V-V^*)_{:,i}\|_1 \\ \overset{(a)}{\geq} & (1-O(\varepsilon)) \cdot \left(\|(U^*V - A)_{:,i}\|_1 - \|(U^*V^* - A)_{:,i}\|_1\right) \\ \overset{(b)}{\geq} & (1-O(\varepsilon)) \cdot \|(U^*V - A)_{:,i}\|_1 + q_{1-\varepsilon/2}(S(U^*V^* - A)_{:,i}), \end{aligned}$$

where (a) follows by triangle inequality, and (b) by (2.1) since the bad column $A_{:,i}$ is *large*. Thus, at least a $(1/2+\varepsilon)$-fraction of entries of $S(U^*V - U^*V^*)_{:,i}$ have absolute value at least

$$(1-O(\varepsilon)) \cdot \|(U^*V - A)_{:,i}\|_1 + q_{1-\varepsilon/2}(S(U^*V^* - A)_{:,i}). \quad (2.3)$$

Since at most an $\varepsilon/2$ fraction of entries of $S(U^*V^* - A)_{:,i}$ have absolute value at least $q_{1-\varepsilon/2}(S(U^*V^* - A)_{:,i})$, by definition of quantile, it follows by (2.3) that in equation (2.2) at most an $\varepsilon/2$-fraction of entries of $S(U^*V - A)_{:,i}$ can have their absolute value reduced to less than $(1 - O(\varepsilon)) \cdot \|(U^*V - A)_{:,i}\|_1$. Further, by (2.3) at least $(1/2 + \varepsilon/2)$-fraction of entries of $S(U^*V - U^*V^*)$ have absolute value at least $(1 - O(\varepsilon))\|(U^*V - A)_{:,i}\|_1$. Therefore, the median of absolute value of the entries of $S(U^*V - A)_{:,i}$ is at least $(1 - O(\varepsilon))\|(U^*V - A)_{:,i}\|_1$, as desired.

Our analysis for $0 < p < 2$ uses similar arguments, but in contrast relies on p-stable random variables. In the case when $0 < p < 1$, special care is needed since the triangle inequality does not hold.

Algorithms for $p = 0$

In the case when $p = 0$ and the entries of matrix A belong to a finite field $\mathbb{F}_q$ for constant q, we use similar arguments as in the case for $p = 1$. Here, instead of p-stable random variables we apply a linear sketch for estimating the number of distinct elements, established in [46]. We show that it suffice to set the number of rows of the sketching matrix S to $\text{poly}(k/\varepsilon) \cdot \log d$. Further, since each entry of S has only q possible values, it is possible to guess matrix S by enumeration in time $q^{\text{poly}(k/\varepsilon)\cdot \log d} = d^{\text{poly}(k/\varepsilon)}$, which will lead to a total running time of $n \cdot d^{\text{poly}(k/\varepsilon)}$. This yields a PTAS for constant q. We defer the details to Chapter 4.

Hardness

Our hardness results for the ℓ_p norm for $p \in (1, 2)$ in Theorem 2.1.3 and $p \in (2, \infty)$ in Theorem 5.0.2 are established via a connection to the matrix $p \to q$ norm problem and its variants. Given a matrix $A \in \mathbb{R}^{n\times d}$, $\|A\|_{p\to q}$ is defined to be $\|A\|_{p\to q} := \max_{x\in\mathbb{R}^d, \|x\|_p=1} \|Ax\|_q$.

Approximately computing this quantity for various values of p and q has been known to have applications to the Small Set Expansion Hypothesis [5], quantum information theory [37], robust optimization [87], and the Grothendieck problem [35]. After active research [5, 8, 9, 39], it is now known that computing the $p \to q$ norm of a matrix is NP-hard to approximate within some constant $c(p, q) > 1$ except when $p = q = 2$, $p = 1$, or $q = \infty$. (Hardness of the case $p < q$ with $2 \in [p, q]$ is only known under stronger assumptions such as the Small Set Expansion Hypothesis or the Exponential Time Hypothesis.) See [9] for a survey of recent results on the approximability of these problems.

We also introduce the problem of computing the following quantity $\min_{p\to q}(A) := \min_{x\in\mathbb{R}^d, \|x\|_p=1} \|Ax\|_q$ as an intermediate problem. Recall that $p^* = p/(p - 1)$ is the

Hölder conjugate of p for which $1/p + 1/p^* = 1$. The following lemma shows that computing ℓ_p-low rank approximation when $k = d - 1$ is equivalent to computing $\min_{p^* \to p}(\cdot)$.

Lemma 2.2.1. *Let $p \in (1, \infty)$. Let $A \in \mathbb{R}^{n \times d}$ with $n \geq d$ and $k = d - 1$. Then*

$$\min_{U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{k \times d}} \|UV - A\|_p = \min_{x \in \mathbb{R}^d, \|x\|_{p^*} = 1} \|Ax\|_p = \min_{p^* \to p}(A).$$

A simple but crucial observation for the above lemma is that if we let $a_1, \ldots, a_n \in \mathbb{R}^d$ be the rows of A, computing the best $(d-1)$-rank approximation of A in the entrywise ℓ_p norm is equivalent to computing the $(d-1)$-dimensional subspace $S \subseteq \mathbb{R}^d$ (i.e., rowspace$(V) = S$) that minimizes $\|(\rho_1, \ldots, \rho_n)\|_p$, where $\rho_i := \min_{y \in S} \|y - a_i\|_p$ denotes the ℓ_p-distance between S and a_i.

If $x \in \mathbb{R}^d$ is a vector orthogonal to S, Hölder's inequality shows that

$$\rho_i = \min_{y \in S} \|y - a_i\|_p = \min_{\langle x, z + a_i \rangle = 0} \|z\|_p \geq \frac{|\langle x, z \rangle|}{\|x\|_{p^*}} = \frac{|\langle x, a_i \rangle|}{\|x\|_{p^*}}.$$

Taking z to be the *Hölder dual* of x, we can show that indeed $\rho_i = |\langle x, a_i \rangle| / \|x\|_{p^*}$. Then $\|(\rho_1, \ldots, \rho_n)\|_p = \|Ax\|_p / \|x\|_{p^*}$, finishing the lemma.

This new connection allows us to prove a number of new hardness results for low rank approximation problems. Previously, even exact hardness results were known only for $p = 0, 1$ and there was no APX-hardness result.

ℓ_p norm with $1 < p < 2$. For $p \in (1, 2)$, we reduce computing $\|\cdot\|_{2 \to p^*}$ to computing $\min_{p^* \to p}(\cdot)$.

If A is an invertible matrix, then

$$\begin{aligned} \min_{p \to p^*}(A^{-1}) &= \min_{x \neq 0} \frac{\|A^{-1}x\|_p}{\|x\|_{p^*}} = \left(\max_{x \neq 0} \frac{\|x\|_{p^*}}{\|A^{-1}x\|_p} \right)^{-1} \\ &= \left(\max_{y \neq 0} \frac{\|Ay\|_{p^*}}{\|y\|_p} \right)^{-1} = \frac{1}{\|A\|_{p \to p^*}}, \end{aligned}$$

and thus computing $\min_{p^* \to p}(\cdot)$ is equivalent to computing $\|\cdot\|_{p \to p^*}$.

By appropriately perturbing and padding 0's, we can show that computing the latter can be reduced to computing the former modulo arbitrarily small error. Standard facts from Banach spaces additionally show that $\|AA^T\|_{p \to p^*} = \|A\|^2_{2 \to p^*}$, proving the following lemma.

Lemma 2.2.2. *For any $\varepsilon > 0, p \in (1, \infty)$, there is an algorithm that runs in $poly(n, \log(1/\varepsilon))$ time and on a non-zero input matrix A, computes a matrix B satisfying*

$$(1 - \varepsilon)\|A\|^{-2}_{2 \to p^*} \leq \min_{p^* \to p}(B) \leq (1 + \varepsilon)\|A\|^{-2}_{2 \to p^*}.$$

To finish Theorem 2.1.3 for ℓ_p-low rank approximation for $p \in (1,2)$, we use the hardness of approximating the $2 \to q$ norm of a matrix proved by Barak et al. [5] assuming the Small Set Expansion Hypothesis when $q = p^* > 2$. Given a d-regular graph $G = (V, E)$ and size bound $\delta \in (0, 1/2)$, the Small Set Expansion problem asks to find a subset $U \subseteq V$ with $|U|/|V| \leq \delta$ that minimizes $\Phi(U) = \frac{|E(U,V\setminus U)|}{d|U|} = 1 - (1_U)^T A(1_U)$, where A and 1_U are the normalized adjacency matrix of G and the normalized indicator vector of U, respectively. Consequently, the problem is equivalent to finding a sparse indicator vector v with high Rayleigh quotient $v^T Av$, and one natural approach is to find a sparse vector in a subspace corresponding to large eigenvalues of A. For $q > 2$, since $\|v\|_q/\|v\|_2$ is maximized when v is supported on only one coordinate and minimized when all entries of v are equal in magnitude, $\|v\|_q/\|v\|_2$ is a natural analytic notion of sparsity, so if we let P be the orthogonal projection on to the subspace corresponding to large eigenvalues, a high $\|P\|_{2\to q}$ seems to indicate that G has a non-expanding small set. Barak et al. formalized this and proved the following theorem when $q \geq 4$ is an even integer, but the same proof essentially works for $q \in (2, \infty)$. For completeness, we present the proof in Chapter 5.

Theorem 2.2.1 ([5]). *Assuming the Small Set Expansion Hypothesis, for any $q \in (2,\infty)$ and $r > 1$, it is NP-hard to approximate the $\|\cdot\|_{2\to q}$ norm within a factor r.*

ℓ_p norm with $2 < p$. Our hardness results for $p \in (2,\infty)$ are proved directly from the above intermediate problem. The following hardness result for $\min_{p^*\to p}(\cdot)$ implies our hardness result for $p \in (2,\infty)$. It follows from a similar result by Guruswami et al. [36], which proves the same hardness for the $\min_{2\to p}(\cdot)$ norm, with some modifications that connect the 2 norm and the p^* norm. Recall that $\gamma_p := \mathbf{E}_g[|g|^p]^{1/p}$ where g is a standard Gaussian, which is strictly greater than 1 for $p > 2$.

Theorem 2.2.2. *For any $p \in (2,\infty)$ and $\varepsilon > 0$, it is NP-hard to approximate the $\min_{p^*\to p}(\cdot)$ norm within a factor $\gamma_p - \varepsilon$.*

Finite Fields. Our hardness results for finite fields rely on the following lemma.

Lemma 2.2.3. *Let $\mathbb{F}$ be a finite field and $A \in \mathbb{F}^{n\times d}$ with $n \geq d$ and $k = d - 1$. Then, we have*

$$\min_{U\in\mathbb{F}^{n\times k}, V\in\mathbb{F}^{k\times d}} \|UV - A\|_0 = \min_{x\in\mathbb{F}^d, x\neq 0} \|Ax\|_0.$$

The proof has a similar structure to Lemma 2.2.1 for the ℓ_p norm in $\mathbb{R}$. We can still identify a subspace $S \subseteq \mathbb{F}^d$ with codimension 1 with a vector x with $\langle v, x\rangle = 0$

for every $v \in S$. In finite fields, x can be possibly in S, but it does not affect the proof. Then for each row a_i of A, if $\langle a_i, x\rangle = 0$, then $a_i \in S$ and we incur no error on the ith row. If $\langle a_i, x\rangle \neq 0$, changing one entry of a_i will ensure that it will be contained in S, so the total number of errors given S is exactly $\|Ax\|_0$.

The quantity in the right-hand side, $\min_{x\in\mathbb{F}^d, x\neq 0}\|Ax\|_0$, is exactly the minimum Hamming weight of any non-zero codeword of the code that has A^T as a generator matrix, or the minimum distance of the code. Then Theorem 2.1.4 above immediately follows from the following theorem by Austrin and Khot [1].

Theorem 2.2.3 ([1])**.** *For any finite field $\mathbb{F}$ and $r > 1$, unless $\mathsf{P} = \mathsf{NP}$, there is no r-approximation algorithm for computing the minimum distance of a given linear code in polynomial time.*

Outline: In Chapter 3 we give preliminaries. In Chapter 4 we give our algorithms for ℓ_p-low rank approximation, $0 < p < 2$, and since it is technically similar, our algorithm for $p = 0$ over finite fields. In Chapter 5 we give all of our hardness results. In Chapter 6 we mention various additional results.

Chapter 3

Preliminaries

For a matrix A we write $A_{i,j}$ for its entry at position (i,j), $A_{i,:}$ for its i-th row, and $A_{:,i}$ for its i-th column.

For $0 \leq p \leq \infty$, we will let $\|A\|_p$ denote the entrywise ℓ_p-norm of A. That is, $\|A\|_0$ equals the number of non-zero entries of A, $\|A\|_\infty = \max_{i,j}\|A_{i,j}\|$, and $\|A\|_p = (\sum_{i,j} A_{i,j}^p)^{\frac{1}{p}}$.

For two matrices A, B the value $\|A - B\|_0$ is a measure of similarity that is sometimes called their Hamming distance.

We will typically give the dimensions of a matrix A as $n \times d$ when A has entries from a field such as $\mathbb{R}$ or $\mathbb{F}_q$. When the entries of A are binary, we will typically give its dimensions as $m \times n$.

We first recall some basic results about Cauchy variables. These have the property that if $x \in \mathbb{R}^n$ and Z, C_i are i.i.d standard Cauchy variables (for $i = 1, \ldots, n$) then it holds that $\sum_{i=1}^n x_i C_i \sim \|x\|_1 Z$.

Fact 3.0.1. *If C is a Cauchy variable with scale γ, then*

1. *For $\tau > 1$, $\mathbf{Pr}[|C| > \tau\gamma] \leq \frac{1}{\tau}$*
2. *For small $\varepsilon > 0$, $\mathbf{Pr}[|C| > (1+\varepsilon)\gamma] < \frac{1}{2} - \Theta(\varepsilon)$*
3. *For small $\varepsilon > 0$, $\mathbf{Pr}[|C| < (1-\varepsilon)\gamma] < \frac{1}{2} - \Theta(\varepsilon)$*

The following results are adapted from [2]. We want to analyze the quantiles of the entries of a vector after a dense Cauchy sketch is applied to it.

Definition 3.0.1. *Let $0 < \alpha < 1$. Let $v \in \mathbb{R}^m$. We let $q_\alpha(v)$ denote the $\frac{1}{\alpha}$-quantile of $|v_1|, |v_2|, \ldots, |v_m|$, or the minimum value greater than $\lceil \alpha n \rceil$ of the values $|v_1|, |v_2|, \ldots, |v_m|$. For $M \in \mathbb{R}^{m \times n}$, we let $q_\alpha(M) = \sum_{i=1}^n q_\alpha(M_{:,i})$.*

We will be particularly interested in the median of the entries of a sketched vector.

Definition 3.0.2. *For $v \in \mathbb{R}^n$, we write med(v) as shorthand for $q_{\frac{1}{2}}(v)$. For $M \in \mathbb{R}^{m\times n}$, we let med$(M) = \sum_{i=1}^{n}$ med$(M_{:,i})$.*

Lemma 3.0.2. *Let $S \in \mathbb{R}^{m\times n}$ have entries that are i.i.d. standard Cauchy variables and let $x \in R^n$. Then*

1. $\mathbf{Pr}[q_{\frac{1}{2}-\Theta(\varepsilon)}(Sx) < (1-\varepsilon)\|x\|_1] < \exp(-\Theta(\varepsilon^2)m)$
2. $\mathbf{Pr}[q_{\frac{1}{2}+O(\varepsilon)}(Sx) > (1+\varepsilon)\|x\|_1] < \exp(-\Theta(\varepsilon^2)m)$
3. *For $M > 2$,* $\mathbf{Pr}[q_{1-\frac{\varepsilon}{2}}(Sx) > \frac{M}{\varepsilon}\|x\|_1] < \exp(-\Theta(\varepsilon)Mm)$
4. *For $M > 2$,* $\mathbf{Pr}[\text{med}(Sx) > M\|x\|_1] < \exp(-\Theta(m)M)$

Proof. Note that for each $1 \le i \le m$, $(Sx)_i$ is distributed as a Cauchy variable with scale $\|x\|_1$. By Fact 3.0.1, $\mathbf{Pr}[|(Sx)_i| < (1-\varepsilon)\|x\|_1] < \frac{1}{2} - \Theta(\varepsilon)$. We want to bound the probability that more than a $\frac{1}{2} - \Theta(\varepsilon)$ fraction of the $(Sx)_i$'s are smaller than $(1-\varepsilon)\|x\|_1$. The desired upper bound follows from Chernoff's bound as $\exp(-\Theta(m)(\frac{1}{2} - \Theta(\varepsilon) - (\frac{1}{2} - \Theta(\varepsilon)))^2)$, from which (i) follows. We can prove (ii) using a similar argument.

For (iii), we know from Fact 3.0.1 that $\mathbf{Pr}[(Sx)_i > \frac{M}{\varepsilon}] < \frac{\varepsilon}{M}$. Thus a Chernoff bound gives $\mathbf{Pr}[q_{1-\frac{\varepsilon}{2}}(Sx) > \frac{M}{\varepsilon}\|x\|_1] < \exp(-\Theta(m)(\frac{\varepsilon}{2} - \frac{\varepsilon}{M})^2(\frac{\varepsilon}{M})^{-1})$ and the result follows. For (iv), a similar proof holds using $\mathbf{Pr}[(Sx)_i > M\|x\|_1] < \frac{1}{M}$.

□

Lemma 3.0.3. *Let $X \subset \mathbb{R}^n$ be a k-dimensional space and $\varepsilon, \delta > 0$. Let S have $O(\frac{1}{\varepsilon^2}k\log\frac{k}{\varepsilon\delta})$ rows, n columns, and i.i.d. Cauchy entries with scale parameter $\gamma = 1$. Then with probability at least $1 - \Theta(\delta)$, for all $x \in X$,*

$$(1-\Theta(\varepsilon))\|x\|_1 \le q_{\frac{1}{2}-\varepsilon}(Sx) \le q_{\frac{1}{2}+\varepsilon}(Sx) \le (1+O(\varepsilon))\|x\|_1$$

Proof. Let N be an $\frac{\varepsilon\delta}{k^3}$-net for the intersection of X and the unit ℓ_1 ball. Then $|N| = \exp(O(k\log\frac{k}{\varepsilon\delta}))$

By Lemma 3.0.2, $\mathbf{Pr}[q_{\frac{1}{2}-\Theta(\varepsilon)}(Sy) < (1-\varepsilon)\|y\|_1] < \exp(-\Theta(k\log\frac{k}{\varepsilon\delta}))$. Thus, for all $y \in N$, $q_{\frac{1}{2}-\Theta(\varepsilon)}(Sy) \ge 1-\varepsilon$ holds with probability $1 - \Theta(\delta)$ by a union bound.

Let X' be a matrix whose columns form an Auerbach basis ([68]) for the subspace X. That is, each column of X' has ℓ_1 norm 1 and $\|z'\|_\infty \le \|X'z'\|_1$ for all z'. By Fact 3.0.1, each entry of SX' is greater than $\Theta(\frac{k^2}{\delta})$ with probability at most $O(\frac{\delta}{k^2})$ because

each column of X' has ℓ_1 norm 1. A union bound tells us that $\|SX'\|_\infty \leq O(\frac{k^2}{\delta})$ with probability at least $1 - \frac{\delta}{2}$.

For arbitrary $z \in X$, we can write $z = X'z'$. Thus

$$\begin{aligned} \|Sz\|_\infty &= \|SX'z'\|_\infty \leq \|SX'\|_\infty \cdot \|z'\|_1 \leq O(k^2/\delta) \cdot k\|z'\|_\infty \\ &\leq O(k^3/\delta) \cdot \|X'z'\|_1 = O(k^3/\delta) \cdot \|z\|_1. \end{aligned}$$

Given any x in the intersection of the unit ℓ_1 ball and X, we can write $x = y + z$ where $y \in N$, $z \in X$, and $\|z\|_1 \leq \frac{\varepsilon\delta}{k^3}$. By the above argument, we know $\|Sz\|_\infty \leq O\left(\frac{k^3}{\delta}\right)\|z\|_1 \leq O(\varepsilon)$. Since $Sx = Sy + Sz$, then $(1 - \Theta(\varepsilon)) \leq q_{\frac{1}{2}-\Theta(\varepsilon)}(Sx)$ for any unit x. We can scale x and ε by the appropriate constants to get the desired statement.

The RHS inequality follows from a similar argument. □

We immediately have the following corollary about medians of Cauchy sketches over subspaces.

Corollary 3.0.4. *Let $X \subset \mathbb{R}^n$ be a k-dimensional space and $\varepsilon, \delta > 0$. Let S have $O(\frac{1}{\varepsilon^2} k \log \frac{k}{\varepsilon\delta})$ rows, n columns, and i.i.d. Cauchy entries with scale parameter $\gamma = 1$. With probability at least $1 - \Theta(\delta)$, for all $x \in X$,*

$$(1 - \varepsilon)\|x\|_1 \leq \mathit{med}(Sx) \leq (1 + \varepsilon)\|x\|_1$$

We can also bound the median and the $(1 - \varepsilon/2)$-quantile of a Cauchy sketch of a fixed matrix.

Lemma 3.0.5. *Let S be an $m \times n$ matrix ($m = \Theta(1/\mathit{poly}(\varepsilon))$) with i.i.d. standard Cauchy entries and let M be an $n \times d$ matrix. For $\varepsilon > 0$, with probability $1 - O(1)$,*

$$(1 - \varepsilon)\|M\|_1 \leq \mathit{med}(SM) \leq (1 + \varepsilon)\|M\|_1$$

Proof. Lemma 3.0.2 tells us that we can choose m so that $\mathbf{Pr}[\text{med}(SM_{:,i}) = (1 \pm \varepsilon)\|M_{:,i}\|_1] \geq 1 - \Theta(\varepsilon)$ for each i. Say i is good if $\text{med}(SM_{:,i}) \geq (1 - \varepsilon)\|M_{:,i}\|_1$ and bad otherwise. Then $\mathbb{E}[\sum_{\text{bad } i}\|M_{:,i}\|_1] \leq \varepsilon\|M\|_1$ so Markov's inequality tells us $\sum_{\text{bad } i}\|M_{:,i}\|_1 \leq O(\varepsilon)\|M\|_1$ with probability $1 - O(1)$ and also $\sum_{\text{good } i}\|M_{:,i}\|_1 \geq (1 - \Theta(\varepsilon))\|M\|_1$.

This implies that

$$\text{med}(SM) \geq \sum_{\text{good } i} \text{med}(SM_{:,i}) \geq (1 - \varepsilon) \sum_{\text{good } i} \|M_{:,i}\|_1 \geq (1 - \varepsilon)(1 - \Theta(\varepsilon))\|M\|_1$$

which gives our first desired inequality.

Now say that column i is small if $\text{med}(SM_{:,i}) < (1+\varepsilon)\|M_{:,i}\|_1$ and (for $k \geq 1$) k-large if

$$(k+1+\varepsilon)\|M_{:,i}\|_1 > \text{med}(SM_{:,i}) \geq (k+\varepsilon)\|M_{:,i}\|_1.$$

For $k \geq 3$, we can bound

$$\begin{aligned}\mathbb{E}\left[\sum_{k\geq 1} k \sum_{k\text{-large } i} \|M_{:,i}\|_1\right] &\leq \Theta(\varepsilon)\|M\|_1 + 2\Theta(\varepsilon)\|M\|_1 + \sum_{k\geq 3} k\varepsilon \exp(-\Theta(m)k)\|M\|_1 \\ &\leq O(\varepsilon)\|M\|_1 \sum_{k\geq 3} \frac{k}{\exp(\Theta(m)k)} \leq O(\varepsilon)\|M\|_1\end{aligned}$$

where the second inequality comes from Lemma 3.0.2 and the third inequality comes from choosing $m = \Theta(1/\text{poly}(\varepsilon))$.

For $k = 1$ or $k = 2$, note that if i is k-large, then $\text{med}(SM_{:,i}) \geq (1+\varepsilon)\|M_{:,i}\|_1$ which occurs with probability at most $\Theta(\varepsilon)$ as mentioned earlier.

This lets us bound

$$\begin{aligned}\mathbb{E}\left[\sum_{k\geq 1} k \sum_{k\text{-large } i} \|M_{:,i}\|_1\right] &\leq \Theta(\varepsilon)\|M\|_1 + 2\Theta(\varepsilon)\|M\|_1 + \sum_{k\geq 3} k\varepsilon \exp(-\Theta(m)k)\|M\|_1 \\ &\leq O(\varepsilon)\|M\|_1 \sum_{k\geq 3} \frac{k}{\exp(\Theta(m)k)} \leq O(\varepsilon)\|M\|_1\end{aligned}$$

where the last inequality occurs because the given infinite series converges by the ratio test.

Therefore

$$\begin{aligned}\text{med}(SM) &= \sum_{\text{small } i} \text{med}(SM_{:,i}) + \sum_{k\geq 1} \sum_{k\text{-large } i} \text{med}(SM_{:,i}) \\ &\leq (1+\varepsilon)\|M\|_1 + \sum_{k\geq 1} (k+1+\varepsilon) \sum_{k\text{-large } i} \|M_{:,i}\|_1 \\ &\leq (1+\varepsilon)\|M\|_1 + \sum_{k\geq 1} 3k \sum_{k\text{-large } i} \|M_{:,i}\|_1 \\ &\leq (1+O(\varepsilon)) \cdot \|M\|_1\end{aligned}$$

where the first inequality holds by the definition of k-large and the third inequality holds with probability $1 - O(1)$ by Markov's inequality. □

Lemma 3.0.6. *When S is an $m \times n$ matrix with i.i.d Cauchy entries, m equals $\Theta(1/\text{poly}(\varepsilon))$, and M is $n \times d$, then with probability $1 - O(1)$,*

$$q_{1-\varepsilon/2}(SM) \leq O\left(\frac{1}{\varepsilon}\right)\|M\|_1$$

Proof. Say that column i is small if $q_{1-\varepsilon/2}(SM_{:,i}) < \frac{3}{\varepsilon}\|M_{:,i}\|_1$ and (for $k \geq 3$) k-large if

$$\frac{k+1}{\varepsilon}\|M_{:,i}\|_1 > q_{1-\varepsilon/2}(SM_{:,i}) \geq \frac{k}{\varepsilon}\|M_{:,i}\|_1.$$

We can bound

$$\begin{aligned}\mathbf{Pr}[i \text{ is } k\text{-large}] &\leq \mathbf{Pr}[q_{1-\varepsilon/2}(SM_{:,i}) \geq \frac{k}{\varepsilon}\|M_{:,i}\|_1] \\ &< \exp(-\Theta(\varepsilon)\frac{k}{\varepsilon}m) < \exp(-\Theta(m)k),\end{aligned}$$

where the second inequality comes from Lemma 3.0.2.

This lets us bound

$$\begin{aligned}\mathbb{E}\left[\sum_{k\geq 3}\frac{k}{\varepsilon}\sum_{k\text{-large } i}\|M_{:,i}\|_1\right] &\leq \sum_{k\geq 3}\frac{k}{\varepsilon}\exp(-\Theta(m)k)\|M\|_1 \\ &\leq \frac{1}{\varepsilon}\|M\|_1\sum_{k\geq 3}\frac{k}{\exp(\Theta(m)k)} \\ &\leq O\left(\frac{1}{\varepsilon}\right)\|M\|_1\end{aligned}$$

where the last inequality occurs because the given infinite series converges by the ratio test.

Therefore

$$\begin{aligned}q_{1-\varepsilon/2}(SM) &= \sum_{\text{small } i} q_{1-\varepsilon/2}(SM_{:,i}) + \sum_{k\geq 3}\sum_{k\text{-large } i} q_{1-\varepsilon/2}(SM_{:,i}) \\ &\leq \frac{3}{\varepsilon}\|M\|_1 + \sum_{k\geq 3}\frac{k+1}{\varepsilon}\sum_{k\text{-large } i}\|M_{:,i}\|_1 \\ &\leq \frac{3}{\varepsilon}\|M\|_1 + \sum_{k\geq 3}\frac{2k}{\varepsilon}\sum_{k\text{-large } i}\|M_{:,i}\|_1 \\ &\leq O\left(\frac{1}{\varepsilon}\right)\|M\|_1\end{aligned}$$

where the first inequality holds by the definition of k-large and the third inequality holds with probability $1 - O(1)$ by Markov's inequality. □

Chebyshev's inequality. We record some basic facts. Let $Z_1, \ldots, Z_n$ be independent Bernoulli random variables, with $Z_i \sim \text{Ber}(p_i)$. Let $Z := Z_1 + \ldots + Z_n$ and $\mu := \mathbf{E}[Z]$.

Lemma 3.0.7. *For any $\Delta > 0$, we have $\mathbf{Pr}[|Z - \mu| > \Delta] \leq \mu/\Delta^2$.*

Proof. By independence, we have $\mathbf{Var}(Z) = \sum_{i=1}^{n} \mathbf{Var}(Z_i) = \sum_{i=1}^{n} p_i(1 - p_i) \leq \sum_{i=1}^{n} p_i = \mu$. By Chebyshev's inequality, for any $\Delta > 0$ we have $\mathbf{Pr}[|Z - \mu| > \Delta] \leq \mathbf{Var}(Z)/\Delta^2$. With $\mathbf{Var}(Z) \leq \mu$ we thus obtain the claim. □

Lemma 3.0.8. *For any $\Delta > 0$, we have $\mathbf{Pr}[|Z - \mu| > \Delta] \leq \sqrt{n}/\Delta$.*

Proof. As in the previous lemma's proof, we have $\mathbf{Pr}[|Z - \mu| > \Delta] \leq \mathbf{Var}(Z)/\Delta^2$, where $\mathbf{Var}(Z) \leq \mu \leq n$, and thus $\mathbf{Pr}[|Z - \mu| > \Delta] \leq n/\Delta^2$. It also follows that $\mathbf{Pr}[|Z - \mu| > \Delta] \leq \sqrt{n}/\Delta$, since if $\sqrt{n}/\Delta < 1$ we have $n/\Delta^2 \leq \sqrt{n}/\Delta$, and otherwise the inequality is trivial. □

Chapter 4

ℓ_p-Approximation Algorithms

Recall that in the Entrywise ℓ_p-Rank-k Approximation problem (for $0 < p < 2$) we are given an $n \times d$ matrix A with integer entries bounded in absolute value by $\text{poly}(n)$, a positive integer k, and we want to output matrices $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{k \times d}$ minimizing

$$\|A - UV\|_p^p := \sum_{i=1,\ldots,n, j=1,\ldots,d} |A_{i,j} - (U \cdot V)_{i,j}|^p.$$

In this chapter, we prove Theorem 2.1.1, restated here for convenience.

Theorem (PTAS for $0 < p < 2$)**.** *Let $p \in (0, 2)$ and $\varepsilon \in (0, 1)$. There is a $(1 + \varepsilon)$-approximation algorithm to Entrywise ℓ_p-Rank-k Approximation running in $n^{\text{poly}(k/\varepsilon)}$ time.*

In Section 4.1, we prove in Corollary 4.1.3 our core algorithm result which solves the Entrywise ℓ_p-Rank-k Approximation problem for $p = 1$. In Section 4.2, we give an algorithm for the case when $1 < p < 2$, and we prove its correctness in Corollary 4.2.7. In Section 4.3, we prove in Corollary 4.3.6 the correctness of our algorithm for $0 < p < 1$. We conclude the proof of Theorem 2.1.1 by combining Corollary 4.1.3, Corollary 4.2.7 and Corollary 4.3.6.

In Section 4.4, we give a $(3 + \varepsilon)$-approximation algorithm for the Entrywise ℓ_p-Rank-k Approximation problem in the case when $p > 2$.

Recall that in the Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$ problem we are given an $n \times d$ matrix A with entries in $\mathbb{F}_q$, a positive integer k, and we want to output matrices $U \in \mathbb{F}_q^{n \times k}$ and $V \in \mathbb{F}_q^{k \times d}$ minimizing $\|A - UV\|_0$. In Section 4.5, we prove Theorem 2.1.2 and for the reader's convenience we restate here our result.

Theorem ($\mathbb{F}_q$ PTAS for $p = 0$)**.** *For $\varepsilon \in (0, 1)$ there is a $(1+\varepsilon)$-approximation algorithm to Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$ running in $n \cdot d^{\text{poly}(k/\varepsilon)}$ time.*

4.1 ℓ_1-Approximation Algorithm

In this section, A_k will denote the rank k matrix closest to A in the entrywise ℓ_1-norm. We will need a claim adapted from [20].

Claim 4.1.1. *If A is n by d and has integer entries bounded by $\gamma = poly(n)$ and rank $r > k$, then we have*

$$\min_{rank\ k\ A_k} \|A - A_k\|_1 \geq \frac{1}{poly(n)^k}$$

Proof. Note that it suffices to lower bound σ_{k+1}, the $(k+1)$-th singular value of A, because $\|A - A_k\|_1 \geq \|A - A_k\|_F \geq \sigma_{k+1}$. This is clear if $\sigma_{k+1} \geq 1$ so we assume otherwise.

Since A has integer entries, then so does $A^T A$ and its characteristic polynomial has integer coefficients. Now $A^T A$ has eigenvalues σ_i^2 so its characteristic polynomial's last term is $\prod_{i=1}^{r} \sigma_i^2$ which is at least 1 because it is a positive integer. For any j, $\sigma_j^2 \leq \|A\|_F^2 \leq nd\gamma^2$.

We have

$$\sigma_{k+1}^{2(r-k)} \geq \prod_{k<i\leq r} \sigma_i^2 \geq \frac{\prod_{1\leq i\leq r} \sigma_i^2}{(nd\gamma^2)^k} \geq \frac{1}{(nd\gamma^2)^k}$$

so $\sigma_{k+1} \geq \frac{1}{(nd\gamma^2)^k}$ because $r - k \geq 1$. □

We can now describe our $(1+\varepsilon)$-approximation algorithm. For the rest of this section, let U^* and V^* be minimizers for $\|UV - A\|_1$ with $OPT = \|U^*V^* - A\|_1$. The quantities θ, ψ will be bounded above by poly(n). The quantity q will be bounded above by poly(k). The specifics of how these values are chosen will be described in the algorithm's proof of correctness. The validity of the specific sampling described in Step 1 of Algorithm 1 will be proved in Corollary 4.1.3.

Algorithm 1 $(1+\varepsilon)$-ℓ_1 low rank approximation

Input: A $n \times d$ matrix A with integer entries bounded by $\gamma = \text{poly}(n)$. An integer $k \in [d]$ and a real $\varepsilon \in (0,1)$.

Output: Matrices $U \in \mathbb{R}^{n\times k}$ and $V \in \mathbb{R}^{k\times d}$ satisfying $\|UV - A\|_1 \leq (1+\varepsilon)OPT$.

1. **If** A has rank at most k, then **return** a rank k decomposition U, V of A.
2. **Sample** an $m \times n$ matrix S satisfying the conditions of Theorem 4.1.1 (e.g. by taking $m = \text{poly}(k/\varepsilon)$ and sampling each entry of S from a standard Cauchy distribution).
3. **Round** the entries of S to the nearest multiple of $\frac{\varepsilon^2}{(\theta\psi)^k}$ where $\theta, \psi \leq \text{poly}(n)$ are chosen as described in the proof of Theorem 4.1.1.
4. **Set** U and V to be zero matrices as a default.
5. Exhaustively **guess** all possible values of SU^* with entries **rounded** to the nearest multiple of $\frac{\varepsilon}{qnk\theta^k}\frac{\varepsilon^2}{(\theta\psi)^k}$, where $q \leq \text{poly}(k)$ is chosen as described in the proof of Theorem 4.1.1.
6. For each guessed SU^*, **set** $\tilde{V} = \arg\min_V \text{med}(SU^*V - SA)$ s.t. $\|V\|_\infty \leq \frac{2nd\gamma qnk\theta^k}{\varepsilon}$.
7. For each $\tilde{V}$, **set** $\tilde{U} = \arg\min_U \|U\tilde{V} - A\|_1$.
8. **If** $\|\tilde{U}\tilde{V} - A\|_1 < \|UV - A\|_1$, then **set** $U = \tilde{U}, V = \tilde{V}$.
9. **Return** U, V.

Theorem 4.1.1. *Let A be an $n \times d$ matrix with integer entries such that $\|A\|_\infty$ is bounded by $\gamma = \text{poly}(n)$. Let $\varepsilon > 1/\text{poly}(n)$. Suppose S is an $m \times n$ random matrix such that with probability $1 - O(1)$, $\text{med}(SU^*V - SA) \geq (1-\varepsilon)\|U^*V - A\|_1$ for all V and for a fixed V^*, $\text{med}(SU^*V^* - SA) \leq (1+\varepsilon)\|U^*V^* - A\|_1$ with probability $1 - O(1)$. Suppose further that $\|S\|_\infty \leq \text{poly}(n)$. Then Algorithm 1 is a $(1+\varepsilon)$-approximation algorithm for rank k low rank approximation in the entrywise ℓ_1 norm and runs in time $\text{poly}(n)^{mk}$.*

Proof. First, if A has rank at most k, then we can just use Gaussian elimination to deduce that its optimal low rank approximation has value 0. We will assume its rank is greater than k.

We can assume V^* is an ℓ_1 well-conditioned basis since we can replace U^* and V^* with U^*R and $R^{-1}V^*$ respectively for an invertible R. Thus for all x we have $\frac{\|x\|_1}{q'} \leq \|x^TV^*\|_1 \leq q\|x\|_1$ where $q', q = \text{poly}(k)$. Using this well-conditioned basis property we see that each entry of U^* is at most $2nd\gamma q' \leq \text{poly}(n)$ because otherwise $\|U^*V^* - A\|_1 \geq \|U^*V^*\|_1 - \|A\|_1 \geq 2nd\gamma - \|A\|_1 \geq \|A\|_1$ and we could improve the ℓ_1 error by taking $U^* = 0$.

Claim 4.1.1 says that there exists $\theta \leq \text{poly}(n)$ such that $OPT \geq \frac{1}{\theta^k}$. By using the well-conditioned basis property of V^* and Claim 4.1.1, we can also assume that

each entry of U^* is rounded to nearest integral multiple of $\frac{\varepsilon}{qnk\theta^k}$ as this will incur an additive error of at most εOPT.

Thus U^* has discretized and bounded entries. Note that there are at most $\varepsilon^{-1}\text{poly}(n)^k$ possible values for each entry of U^*.

Since the entries of U^* are discretized by $\frac{\varepsilon}{qnk\theta^k}$, then the entries of V^* can be bounded above by $\frac{2nd\gamma qnk\theta^k}{\varepsilon}$ because otherwise $\|U^*V^* - A\|_1 \geq 2nd\gamma - \|A\|_1 \geq \|A\|_1$ and we might as well have set $V^* = 0$.

The well-conditioned basis property shows that $\|V^*\|_\infty \leq q$. We will be interested in matrices V where $\|V\|_\infty \leq \frac{2nd\gamma qnk\theta^k}{\varepsilon}$ (note that this includes V^* because q is less than the RHS).

We have $\|U^*V - A\|_1 \leq \varepsilon^{-1}\psi^k$ where $\psi \leq \text{poly}(n)$. We will round each entry of S to the nearest multiple of $\frac{\varepsilon^2}{(\theta\psi)^k}$, so we can write $S = \tilde{S} + \Delta$ where $\tilde{S}$ is discretized and $\|\Delta\|_\infty \leq \frac{\varepsilon^2}{(\theta\psi)^k}$. Note that $\|\Delta(U^*V - A)\|_1 \leq \frac{\varepsilon}{\theta^k} \leq \varepsilon OPT$.

Now we will prove the correctness of our algorithm. We can sample $S = \tilde{S} + \Delta$. Note that $\tilde{S}U^*$ will have entries that are multiples of $\frac{\varepsilon}{qnk\theta^k}\frac{\varepsilon^2}{(\theta\psi)^k} \geq \text{poly}(\frac{\varepsilon}{n})^k$ and bounded by $\text{poly}(\frac{n}{\varepsilon})^k$ because $\tilde{S}$ is discretized and bounded. Since $\tilde{S}U^*$ is $m \times k$, then in $\text{poly}(\frac{n}{\varepsilon})^k$ time we can exhaustively search through all possible values of $\tilde{S}U^*$ and one of them will be correct.

For each guess of $\tilde{S}U^*$ and each i we minimize $\text{med}(\tilde{S}U^*V_{:,i} - \tilde{S}A_{:,i})$ over $\|V_{:,i}\|_\infty \leq \frac{2nd\gamma qnk\theta^k}{\varepsilon}$ [1] to get $\tilde{V}_{:,i}$. We have $\text{med}(\tilde{S}U^*\tilde{V} - \tilde{S}A) \leq \text{med}(\tilde{S}U^*V^* - \tilde{S}A)$.

Now

$$\begin{aligned}
\text{med}(\tilde{S}U^*V^* - \tilde{S}A) &= \text{med}(S(U^*V^* - A) - \Delta(U^*V^* - A)) \\
&\leq \text{med}(S(U^*V^* - A)) + \varepsilon \cdot OPT \\
&\leq (1+\varepsilon)\|U^*V^* - A\|_1 + \varepsilon \cdot OPT \\
&\leq (1 + O(\varepsilon)) \cdot OPT.
\end{aligned}$$

We choose $\tilde{U}$ to minimize $\|\tilde{U}\tilde{V} - A\|_1$, so

$$\begin{aligned}
\text{med}(\tilde{S}U^*\tilde{V} - \tilde{S}A) &= \text{med}(S(U^*\tilde{V} - A) - \Delta(U^*\tilde{V} - A)) \\
&\geq \text{med}(S(U^*\tilde{V} - A)) - \varepsilon \cdot OPT \\
&\geq (1-\varepsilon)\|U^*\tilde{V} - A\|_1 - \varepsilon \cdot OPT \\
&\geq (1-\varepsilon)\|\tilde{U}\tilde{V} - A\|_1 - \varepsilon \cdot OPT
\end{aligned}$$

It follows that the best $\tilde{U}$ and $\tilde{V}$ will satisfy $\|\tilde{U}\tilde{V} - A\|_1 \leq (1 + O(\varepsilon)) \cdot OPT$. □

[1] Observe that there are at most $m!$ orderings of the entries of $\tilde{S}U^*V_{:,i} - \tilde{S}A_{:,i}$ and we are minimizing a linear function over $V_{:,i}$ subject to a linear constraint. This can be solved with linear programming, so it will be done within the $\text{poly}(n)^{mk}$ runtime.

Note that if $m = \Theta(\text{poly}(\frac{k \log d}{\varepsilon}))$, then the above algorithm is a quasipolynomial time $(1+\varepsilon)$-approximation scheme (treating k like a constant). This is because we can use Corollary 3.0.4 (with $\delta = \text{poly}(1/d)$) to see that

$$\text{med}(S \left[U^* \; A_{:,i}\right] \left[V_{:,i} \; 1\right]^T\Big]) = (1 \pm \varepsilon)\|\left[U^* \; A_{:,i}\right] \left[V_{:,i} \; 1\right]^T\|_1$$

(when $V_{:,i}$ is arbitrary) with probability at least $1 - \text{poly}(1/d)$ for each i. By a union bound, $\text{med}(S(U^*V - A)) = (1 \pm \varepsilon)\|U^*V - A\|_1$ for arbitrary V with probability $1 - \Theta(1)$. Furthermore, Fact 3.0.1 tells us that $\mathbf{Pr}[S_{i,j} \geq \text{poly}(n)] \leq \text{poly}(n)^{-1}$ so by a union bound, all entries of S are bounded by $\text{poly}(n)$ with probability $1 - \Theta(1)$.

Of course, if we could reduce m to $\Theta(\text{poly}(\frac{k}{\varepsilon}))$, then we would have a PTAS. With the target bound for m, we would still have a $(1 \pm \varepsilon)$-embedding for each column index i with probability $1 - O(1)$, but we need all d embeddings to be valid at once because a simple union bound would not suffice. We accomplish this in the next result which is a variant of Lemma 27 from [21].

Theorem 4.1.2. *Let $U \in \mathbb{R}^{n \times k}, A \in \mathbb{R}^{n \times d}$. Let V^* be chosen to minimize $\|UV^* - A\|_1$. Suppose S is an $m \times n$ matrix satisfying*

(i) $q_{\frac{1}{2}-\varepsilon}(SUx) \geq (1 - \Theta(\varepsilon))\|Ux\|_1$ *for all* x

(ii) *For each i with probability at least $1-\varepsilon^3$, $\text{med}(S[U \; A_{:,i}]x) \geq (1-\varepsilon^3)\|[U \; A_{:,i}]x\|_1$ for all x*

(iii) $\text{med}(SUV^* - SA) \leq (1 + \varepsilon^3)\|UV^* - A\|_1$

(iv) $q_{1-\varepsilon/2}(S(UV^* - A)) \leq O\left(\frac{1}{\varepsilon}\right)\|UV^* - A\|_1$

Then $\text{med}(SUV - SA) \geq (1 - O(\varepsilon))\|UV - A\|_1$ for arbitrary V.

Proof. We say a column index i is *good* if

$$\text{med}(S([U \; A_{:,i}]y)) \geq (1 - \varepsilon^3)\|[U \; A_{:,i}]y\|_1$$

for all $y \in \mathbb{R}^{k+1}$, and *bad* otherwise. We say a bad column index is *large* if

$$\varepsilon\|(UV - A)_{:,i}\|_1 \geq \frac{1}{1-\varepsilon} q_{1-\varepsilon/2}(S(UV^* - A)_{:,i}) + \|(UV^* - A)_{:,i}\|_1$$

and *small* otherwise.

By (ii), we know that $\mathbb{E}[\sum_{\text{bad } i}\|(UV^* - A)_{:,i}\|_1] \leq \varepsilon^3\|UV^* - A\|_1$. By Markov's inequality, we know that with probability $1 - O(1)$,

$$\sum_{\text{bad } i}\|(UV^* - A)_{:,i}\|_1 \leq O(\varepsilon^3)\|UV^* - A\|_1. \tag{4.1}$$

Since we were only using the probability that a column i was bad in the bound above, then by a similar Markov's inequality argument we know that with probability $1 - O(1)$,

$$\sum_{\text{bad } i} q_{1-\varepsilon/2}(S(UV^* - A)_{:,i}) \leq O(\varepsilon^3) q_{1-\varepsilon/2}(S(UV^* - A)). \tag{4.2}$$

By (iii)

$$\begin{aligned}(1+\varepsilon^3)\|UV^* - A\|_1 &\geq \text{med}(S(UV^* - A)) \\ &\geq (1-\varepsilon^3) \sum_{\text{good } i} \|(UV^* - A)_{:,i}\|_1 + \sum_{\text{bad } i} \text{med}(S(UV^* - A)_{:,i}) \\ &\geq (1-\varepsilon^3)(1-\Theta(\varepsilon^3))\|UV^* - A\|_1 + \sum_{\text{bad } i} \text{med}(S(UV^* - A)_{:,i}),\end{aligned}$$

where the second inequality comes from the definition of good, and the third inequality comes from (4.1).

Thus

$$\sum_{\text{bad } i} \text{med}(S(UV^* - A)_{:,i}) \leq O(\varepsilon^3)\|UV^* - A\|_1 \tag{4.3}$$

We also have

$$\begin{aligned}\sum_{\text{small } i} \|(UV - A)_{:,i}\|_1 &\leq \frac{1}{\varepsilon(1-\varepsilon)} \sum_{\text{small } i} q_{1-\varepsilon/2}(S(UV^* - A)_{:,i}) + \frac{1}{\varepsilon} \sum_{\text{small } i} \|(UV^* - A)_{:,i}\|_1 \\ &\leq \frac{1}{\varepsilon(1-\varepsilon)} \sum_{\text{bad } i} q_{1-\varepsilon/2}(S(UV^* - A)_{:,i}) + O(\varepsilon^2)\|UV^* - A\|_1 \\ &\leq O\left(\varepsilon^2\right) q_{1-\varepsilon/2}(S(UV^* - A)) + O(\varepsilon^2)\|UV^* - A\|_1 \\ &\leq O(\varepsilon)\|UV^* - A\|_1 + O(\varepsilon^2)\|UV^* - A\|_1 \\ &\leq O(\varepsilon)\|UV^* - A\|_1 \end{aligned} \tag{4.4}$$

where the first inequality comes from the definition of small, the second inequality comes from (4.1) and the fact that small columns are bad columns, the third inequality comes from (4.2), and the fourth inequality comes from (iv).

Claim 4.1.2.

$$\sum_{\textit{large } i} \textit{med}(S(UV - A)_{:,i}) \geq (1 - O(\varepsilon)) \sum_{\textit{large } i} \|(UV - A)_{:,i}\|_1$$

Proof. Let i be large. We can write $S(UV - A)_{:,i} = SU(V - V^*)_{:,i} + S(UV^* - A)_{:,i}$.

By (i), we know at least $\frac{1}{2}+\varepsilon$ entries of $SU(V-V^*)_{:,i}$ are larger than $(1-O(\varepsilon))\|U(V-V^*)_{:,i}\|_1$ which is at least

$$(1-O(\varepsilon))(\|(UV-A)_{:,i}\|_1-\|(UV^*-A)_{:,i}\|_1)$$

by the triangle inequality. By the definition of large, this is at least

$$(1-O(\varepsilon))((1-\varepsilon)\|(UV-A)_{:,i}\|_1+\left(\frac{1}{1-\varepsilon}\right)q_{1-\varepsilon/2}(S(UV^*-A)_{:,i}))$$

or

$$(1-O(\varepsilon))^2\|(UV-A)_{:,i}\|_1+q_{1-\varepsilon/2}(S(UV^*-A)_{:,i}).$$

By definition, less than an $\varepsilon/2$ fraction of the entries of $S(UV^*-A)_{:,i}$ are greater than $q_{1-\varepsilon/2}(S(UV^*-A)_{:,i})$ so at least half of the entries of $S(UV-A)_{:,i}$ are greater than $(1-O(\varepsilon))^2\|(UV-A)_{:,i}\|_1$. The result follows. □

Finally

$$\begin{aligned}
\operatorname{med}(S(UV-A)) &\geq \sum_{\text{good } i}\operatorname{med}(S(UV-A)_{:,i})+\sum_{\text{large } i}\operatorname{med}(S(UV-A)_{:,i})\\
&\geq (1-\varepsilon^3)\sum_{\text{good } i}\|(UV-A)_{:,i}\|_1+(1-O(\varepsilon))\sum_{\text{large } i}\|(UV-A)_{:,i}\|_1\\
&\geq (1-O(\varepsilon))\|UV-A\|_1-(1-O(\varepsilon))\sum_{\text{small } i}\|(UV-A)_{:,i}\|_1\\
&\geq (1-O(\varepsilon))\|UV-A\|_1-(1-O(\varepsilon))O(\varepsilon)\|UV^*-A\|_1\\
&\geq (1-O(\varepsilon))\|UV-A\|_1
\end{aligned}$$

where the first inequality occurs because large i are bad i, the second inequality comes from the definition of good and Claim 4.1.2, the third inequality comes from the definition of small, the fourth inequality comes from (4.4), and the last inequality holds because V^* is a minimizer. □

Corollary 4.1.3. *Let A be an $n\times d$ matrix with integer entries bounded by* $\operatorname{poly}(n)$ *and let k be a constant. There is a PTAS for finding the closest rank k matrix to A in the entrywise ℓ_1 norm.*

Proof. Let $U^*\in\mathbb{R}^{n\times k}, V^*\in\mathbb{R}^{k\times d}$ be minimizers for $\|U^*V^*-A\|_1$. It suffices to prove that an $m\times n$ ($m=\Theta(\operatorname{poly}(\frac{k}{\varepsilon}))$) matrix S with i.i.d. standard Cauchy entries satisfies the conditions of Theorem 4.1.2 with $U=U^*$, then use Theorem 4.1.1.

Indeed, S satisfies (i) through Lemma 3.0.3 and (ii) with probability $1-O(1)$ through Corollary 3.0.4. S satisfies (iii) with probability $1-O(1)$ via Lemma 3.0.5 and (iv) with probability $1-O(1)$ via Lemma 3.0.6. □

4.2 $1 < p < 2$

We can extend these ℓ_1 results to ℓ_p for $1 < p < 2$ by using p-stable variables (with scale 1) instead of Cauchy variables (or 1-stable variables). These have the property that if $x \in \mathbb{R}^n$ and Z, Z_i are i.i.d p-stable variables (for $i = 1, \ldots, n$) then $\sum_{i=1}^n x_i Z_i \sim \|x\|_p Z$.

Definition 4.2.1. *We let med_p denote the median of the absolute value of a p-stable variable.*

There is no convenient closed form expression for med_p unless $p = 1$, in which case $\text{med}_1 = 1$. However, in Appendix A.2 of [46] it is shown that a $1 \pm \varepsilon$ approximation of med_p can be computed efficiently. Since we are only interested in ε approximations, then this will suffice for our purposes. Our main sketch will be $\text{med}\left(\frac{(Sx)}{\text{med}_p}\right)$ (S has i.i.d p-stable entries with scale 1) which will concentrate around $(1 \pm \varepsilon)\|x\|_p$.

We can cite similar concentration / tail bounds for p-stable variables like the ones we used for Cauchy variables. We can also state a series of claims analagous to the ones we used in the ℓ_1 case.

Fact 4.2.1. *If Z is a p-stable variable with scale γ, then*

1. *For $\tau > 1$, $\mathbf{Pr}[|Z| > \tau\gamma\text{med}_p] \leq \Theta(\frac{1}{\tau^p})$*
2. *For small $\varepsilon > 0$, $\mathbf{Pr}[|Z| > (1+\varepsilon)\gamma\text{med}_p] < \frac{1}{2} - \Theta(\varepsilon)$*
3. *For small $\varepsilon > 0$, $\mathbf{Pr}[|Z| < (1-\varepsilon)\gamma\text{med}_p] < \frac{1}{2} - \Theta(\varepsilon)$*

Lemma 4.2.2. *Let $S \in \mathbb{R}^{m \times n}$ have entries that are i.i.d. p-stable variables with scale 1 and let $x \in R^n$. Then*

1. $\mathbf{Pr}[q_{\frac{1}{2}-\Theta(\varepsilon)}(Sx) < (1-\varepsilon)\|x\|_p\text{med}_p] < \exp(-\Theta(\varepsilon^2)m)$
2. $\mathbf{Pr}[q_{\frac{1}{2}+O(\varepsilon)}(Sx) > (1+\varepsilon)\|x\|_p\text{med}_p] < \exp(-\Theta(\varepsilon^2)m)$
3. *For $M > 3$,* $\mathbf{Pr}[q_{1-\frac{\varepsilon}{2}}(Sx) > \frac{M}{\varepsilon}\|x\|_p\text{med}_p] < \exp(-\Theta(\varepsilon)Mm)$
4. *For $M > 3$,* $\mathbf{Pr}[\text{med}(Sx) > M\|x\|_p\text{med}_p] < \exp(-\Theta(m)M)$

Proof. The proof follows the same structure as the proof for Lemma 3.0.2. We use Fact 4.2.1 in combination with Chernoff bounds. □

Since $1 < p$, then we can take advantage of Minkowski's inequality and use the triangle inequality with $\|\cdot\|_p$.

Lemma 4.2.3. *Let $X \subset \mathbb{R}^n$ be a k-dimensional space and $\varepsilon, \delta > 0$. Let S have $O(\frac{1}{\varepsilon^2} k \log \frac{k}{\varepsilon\delta})$ rows, n columns, and i.i.d. p-stable entries with scale 1. Then with probability at least $1 - \Theta(\delta)$, for all $x \in X$,*

$$(1 - \Theta(\varepsilon))\|x\|_p \leq q_{\frac{1}{2}-\varepsilon}(Sx/med_p) \leq q_{\frac{1}{2}+\varepsilon}(Sx/med_p) \leq (1 + O(\varepsilon))\|x\|_p$$

Proof. The proof follows the same structure as the proof for Lemma 3.0.3 except we use Fact 4.2.1 and p-well conditioned bases ([23]) to bound $\|(Sz)/\text{med}_p\|_\infty$ for any $z \in X$. We also use the ℓ_p ball (which is still convex) instead of the ℓ_1 ball. □

This automatically gives us the following corollary.

Corollary 4.2.4. *Let $X \subset \mathbb{R}^n$ be a k-dimensional space and $\varepsilon, \delta > 0$. Let S have $O(\frac{1}{\varepsilon^2} k \log \frac{k}{\varepsilon\delta})$ rows, n columns, and i.i.d. p-stable entries with scale 1. With probability at least $1 - \Theta(\delta)$, for all $x \in X$,*

$$(1 - \varepsilon)\|x\|_p \leq med(Sx/med_p) \leq (1 + \varepsilon)\|x\|_p$$

We also have analogous versions of our bounds on fixed matrices. The proof structures are the same as those of Lemmas 3.0.5 and 3.0.6.

Lemma 4.2.5. *Let S be an $m \times n$ matrix ($m = \Theta(1/poly(\varepsilon))$) with i.i.d. standard p-stable entries and let M be an $n \times d$ matrix. For $\varepsilon > 0$, with probability $1 - O(1)$,*

$$(1 - \varepsilon)\sum_i \|M\|_p \leq (\sum_i med(SM_{:,i})^p)^{1/p}/med_p \leq (1 + \varepsilon)\sum_i \|M\|_p$$

Lemma 4.2.6. *When S is an $m \times n$ matrix with i.i.d standard p-stable entries, $m = \Theta(1/poly(\varepsilon))$, and M is $n \times d$, then with probability $1 - O(1)$,*

$$(\sum_i q_{1-\varepsilon/2}(SM_{:,i})^p)^{1/p}/med_p \leq O\left(\frac{1}{\varepsilon}\right)\sum_i \|M\|_p$$

Finally, we have an ℓ_p form of Theorem 4.1.2 and it is proved analogously.

Theorem 4.2.1. *Let $U \in \mathbb{R}^{n\times k}, A \in \mathbb{R}^{n\times d}$. Let V^* be chosen to minimize $\|UV^* - A\|_p$. Suppose S is an $m \times n$ matrix satisfying*

1. $q_{\frac{1}{2}-\varepsilon}(SUx/med_p) \geq (1 - \Theta(\varepsilon))\|Ux\|_p$
2. *For each i with probability at least $1 - \varepsilon^3$,*

$$med(S[U\ A_{:,i}]x/med_p) \geq (1 - \varepsilon^3)\|[U\ A_{:,i}]x\|_p$$

for all x

3. $(\sum_i med(SUV^*_{:,i} - SA_{:,i})^p)^{1/p}/med_p \leq (1+\varepsilon^3)\|UV^* - A\|_p$

4. $(\sum_i q_{1-\varepsilon/2}(S(UV^* - A)_{:,i})^p)^{1/p}/med_p \leq O\left(\frac{1}{\varepsilon}\right)\|UV^* - A\|_p$

Then $(\sum_i med(SUV_{:,i} - SA_{:,i})^p)^{1/p}/med_p \geq (1-O(\varepsilon))\sum_i \|UV - A\|_p$ *for arbitrary* V.

It follows that we have a PTAS for rank k ℓ_p low rank approximation.

Corollary 4.2.7. *Let A be an $n \times d$ matrix with entries bounded by poly(n) and let k be a constant. There is a PTAS for finding the closest rank k matrix to A in entrywise ℓ_p norm for $1 < p < 2$.*

Proof. The algorithm is analogous to Algorithm 1. Correctness follows from the fact that there exist ℓ_p well-conditioned bases and that ℓ_p regression is a convex optimization problem.

Indeed, if $p > 1$ then $\|UV_{:,i} - A_{:,i}\|_p$ is convex over vectors $V_{:,i}$ and we can calculate minima in polynomial time. □

4.3 $0 < p < 1$

For $v \in \mathbb{R}^n$ we will denote v^p to mean we raise each entry of v to the pth power, i.e. $(v^p)_i = v_i^p$.

We can extend these results to ℓ_p for $0 < p < 1$ as well, but more care needs to be taken for this range of p because among other issues, $\|\cdot\|_p$ is no longer a norm. However, $\|\cdot\|_p^p$ satisfies the triangle inequality which will be enough for our purposes. We will prove that med $\left(\frac{(Sx)^p}{\text{med}_p^p}\right)$ (S has i.i.d p-stable entries) will concentrate around $(1 \pm \varepsilon)\|x\|_p^p$.

Lemma 4.3.1. *Let $S \in \mathbb{R}^{m \times n}$ have entries that are i.i.d. p-stable variables with scale 1 and let $x \in R^n$. Then*

1. $\mathbf{Pr}[q_{\frac{1}{2}-\Theta(\varepsilon)}(Sx)^p < (1-\varepsilon)\|x\|_p^p \text{med}_p^p] < \exp(-\Theta(\varepsilon^2)m)$

2. $\mathbf{Pr}[q_{\frac{1}{2}+O(\varepsilon)}(Sx)^p > (1+\varepsilon)\|x\|_p^p \text{med}_p^p] < \exp(-\Theta(\varepsilon^2)m)$

3. *For $M > 3$,* $\mathbf{Pr}[q_{1-\frac{\varepsilon}{2}}(Sx)^p > \frac{M}{\varepsilon}\|x\|_p^p \text{med}_p^p] < \exp(-\Theta(\varepsilon)Mm)$

4. *For $M > 3$,* $\mathbf{Pr}[\text{med}(Sx)^p > M\|x\|_p^p \text{med}_p^p] < \exp(-\Theta(m)M)$

Proof. These results follow from Lemma 4.2.2 and the fact that for $0 < p < 1$, we have $(1-\varepsilon)^p > 1-\varepsilon$ and $(1+\varepsilon)^p < 1+\varepsilon$. □

Using the above quantile results we can prove an embedding result similar to Lemma 4.2.3 by using the fact that $\|\cdot\|_p^p$ satisfies the triangle inequality.

Lemma 4.3.2. *Let $X \subset \mathbb{R}^n$ be a k-dimensional space and $\varepsilon, \delta > 0$. Let S have $O(\frac{1}{\varepsilon^2} k \log \frac{k}{\varepsilon\delta})$ rows, n columns, and i.i.d. p-stable entries with scale 1. Then with probability at least $1 - \Theta(\delta)$, for all $x \in X$,*

$$(1-\Theta(\varepsilon))\|x\|_p^p \leq q_{\frac{1}{2}-\varepsilon}((Sx)^p/med_p^p) \leq q_{\frac{1}{2}+\varepsilon}((Sx)^p/med_p^p) \leq (1+O(\varepsilon))\|x\|_p^p$$

This automatically gives us the following corollary.

Corollary 4.3.3. *Let $X \subset \mathbb{R}^n$ be a k-dimensional space and $\varepsilon, \delta > 0$. Let S have $O(\frac{1}{\varepsilon^2} k \log \frac{k}{\varepsilon\delta})$ rows, n columns, and i.i.d. p-stable entries with scale 1. With probability at least $1 - \Theta(\delta)$, for all $x \in X$,*

$$(1-\varepsilon)\|x\|_p^p \leq med((Sx)^p/med_p^p) \leq (1+\varepsilon)\|x\|_p^p$$

We also have analogous versions of our bounds on fixed matrices. Again, the proof structures are the same as those of Lemmas 3.0.5 and 3.0.6.

Lemma 4.3.4. *Let S be an $m \times n$ matrix ($m = \Theta(1/poly(\varepsilon))$) with i.i.d. standard p-stable entries and let M be an $n \times d$ matrix. For $\varepsilon > 0$, with probability $1 - O(1)$,*

$$(1-\varepsilon)\|M\|_p^p \leq \sum_i med((SM_{:,i})^p/med_p^p) \leq (1+\varepsilon)\|M\|_p^p$$

Lemma 4.3.5. *When S is an $m \times n$ matrix with i.i.d standard p-stable entries, $m = \Theta(1/poly(\varepsilon))$, and M is $n \times d$, then with probability $1 - O(1)$,*

$$\sum_i q_{1-\varepsilon/2}((SM_{:,i})^p/med_p^p) \leq O\left(\frac{1}{\varepsilon}\right) \sum_i \|M_{:,i}\|_p^p$$

As expected, we have an ℓ_p form of Theorem 4.1.2 and it is proved analogously.

Theorem 4.3.1. *Let $U \in \mathbb{R}^{n\times k}, A \in \mathbb{R}^{n\times d}$. Let V^* be chosen to minimize $\|UV^* - A\|_p^p$. Suppose S is an $m \times n$ matrix satisfying*

1. $q_{\frac{1}{2}-\varepsilon}((SUx)^p/med_p^p) \geq (1-\Theta(\varepsilon))\|Ux\|_p^p$
2. *For each i with probability at least $1 - \varepsilon^3$, $med((S[U\ A_{:,i}]x)^p/med_p^p) \geq (1-\varepsilon^3)\|[U\ A_{:,i}]x\|_p^p$ for all x*

3. $\sum_i med((SUV^* - SA)^p_{:,i}/med^p_p) \leq (1+\varepsilon^3)\sum_i \|(UV^* - A)_{:,i}\|^p_p$

4. $\sum_i q_{1-\varepsilon/2}(S(UV^* - A)^p_i/med^p_p) \leq O\left(\frac{1}{\varepsilon}\right)\sum_i \|(UV^* - A)_i\|^p_p)$

Then $\sum_i med((SUV - SA)^p_{:,i}/med^p_p) \geq (1 - O(\varepsilon))\|(UV - A)\|^p_p$ *for arbitrary* V.

The results above can give us the desired PTAS.

Corollary 4.3.6. *Let A be an $n \times d$ matrix with entries bounded by poly(n) and let k be a constant. There is a PTAS for finding the closest rank k matrix to A in entrywise ℓ_p norm when $0 < p < 1$.*

Proof. The algorithm is slightly different from Algorithm 1, because ℓ_p regression is no longer a convex optimization problem when $0 < p < 1$. Thus after sketching to find a minimizing V, we need a different approach to find a minimizing U. We accomplish this by sketching $UV - A$ again, but from the right and guessing the sketched V. We use the guessed V to solve for U.

Besides the above modification, we rely on the fact that $\|\cdot\|^p_p$ satisfies the triangle inequality. We also note that for $0 < p < 1$, we may not have a well-conditioned basis. However, we know that an ℓ_1 well-conditioned basis exists so there exist $q, r = \text{poly}(k)$ such that $\frac{\|x\|_1}{q} \leq \|x^T V^*\|_1 \leq r\|x\|_1$. By Holder's inequality, we know $\|x^T V^*\|^p_p \leq d^{1-p}\|x^T V^*\|^p_1 \leq d^{1-p}r^p\|x\|^p_p$ and $\|x^T V^*\|^p_p \geq \|x^T V^*\|^p_1 \geq \|x\|^p_1/q^p \geq d^{p-1}\|x\|^p_p/q^p$ so we can get a similar well-conditioned basis result saying there exist $\tilde{q}, \tilde{r} = \text{poly}(d)$ such that $\frac{\|x\|_p}{\tilde{q}} \leq \|x^T V^*\|_p \leq \tilde{r}\|x\|_p$ which will suffice for our proof. □

4.4 $p > 2$

There are no p-stable random variables when $p > 2$ so any ℓ_p-approximation algorithms in this setting will need to rely on a different technique. Our sketch will be lifted from [23]. Rather than a matrix of p-stable random variables, we use a sampling matrix that samples m rows of A with each row i having some probability p_i of being sampled. Furthermore, each sampled row is reweighted by $1/p_i$. The following claim (adapted from Theorem 5 of [23]) says we can get a subspace embedding from the right sampling matrix.

Claim 4.4.1. *Suppose U is an $n \times k$ matrix. Then there exists a $m \times n$ sampling matrix S with $m = poly(k/\varepsilon)$ such that $\|SUx\|_p = (1 \pm \varepsilon)\|Ux\|_p$ for all x.*

Theorem 4.4.1. *If A is an $n \times d$ matrix with entries bounded by poly(n), then there is a $(3+\varepsilon)$-approximation algorithm running in time $n^{poly(k/\varepsilon)}$ for finding the closest rank k matrix to A in the entrywise ℓ_p norm for $p > 2$.*

Proof. Let S be the sampling matrix of the above claim. Let $\widehat{V}$ be a minimizer for the expression $\|SU^*\widehat{V} - SA\|_p$. Again, by a similar argument as that of the proof of Theorem 4.1.1, we can guess SU^* using poly(n) tries. We can round the sampling probabilities and the entries of U^* to the nearest $1/\text{poly}(n)$ value.

We know that

$$\begin{aligned}
\|U^*\widehat{V} - A\|_p &\leq \|U^*(\widehat{V} - V^*)\|_p + \|U^*V^* - A\|_p \\
&\leq (1 + O(\varepsilon))\|SU^*(\widehat{V} - V^*)\|_p + \|U^*V^* - A\|_p \\
&\leq (1 + O(\varepsilon))\|SU^*\widehat{V} - SA\|_p + (1 + O(\varepsilon))\|SU^*V^* - SA\|_p \\
&\quad + \|U^*V^* - A\|_p \\
&\leq 2(1 + O(\varepsilon))\|SU^*V^* - SA\|_p + \|U^*V^* - A\|_p \\
&\leq (3 + \varepsilon)\|U^*V^* - A\|_p
\end{aligned}$$

where the second inequality follows from the embedding property of S and the fourth inequality comes from the definition of $\widehat{V}$ as a minimizer.

The final inequality comes from a Markov bound on S. More specifically, since S is a sampling matrix, then for an arbitrary matrix M, $\mathbb{E}[SM] = \|M\|_p$. Thus Markov's Inequality says that with probability $1 - O(1)$, we have $SM \leq O(1)\|M\|_p$. This concludes the proof.

□

4.5 Finite Fields

We can also study low rank approximation over finite fields. The ℓ_p metrics are not defined over finite fields for $p > 0$, but we can look at low rank approximation over the entrywise ℓ_0 metric (where $\|M\|_0 = |\{(i,j) : M_{i,j} \neq 0\}|$). For the rest of this section we will work over a finite field $\mathbb{F}_q$, for some prime power q.

The structure of the algorithm will be similar to that of the case $0 < p < 2$ but our sketch will be based on hashing rather than p-stable random variables. Furthermore, we will be able to sketch in the dimension d row space rather than the dimension n column space and get a running time better than that of the $0 < p < 2$ algorithms. We now describe a $(1 + \varepsilon)$-approximation sketch for the ℓ_0 metric, where ε will be sufficiently small. This sketch is inspired by the L_0 streaming algorithm in [46]. Throughout this section, we will refer to constants C and C' that are sufficiently large.

Let S_i denote a $n \times n$ matrix where column i is the standard basis column e_i with probability $p_i = \frac{1}{2^i}$ or the all zeroes column otherwise. In other words, S_i is a

sampling matrix that takes x and preserves each coordinate with probability p_i and otherwise maps the coordinate to 0. Note that $p_0 = 1$. We can generate our matrices S_i by uniformly sampling n integers between 0 to n and sampling column j in S_i if the leading 1 in the jth integer (written in binary, with indexing starting from 1) is before the ith position. Observe that under this procedure, our subsampling is nested so that if S_i does not sample entry j, then neither will $S_{i'}$ for any $i' > i$.

Note that by this nestedness property, we have $\|x\|_0 = \|S_0 x\|_0 \geq \|S_1 x\|_0 \geq \|S_2 x\|_0 \geq \cdots \geq \|S_{\log n - 1} x\|_0$. Let S denote the $n \log n \times n$ block matrix $\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ \vdots \\ S_{\log n - 1} \end{bmatrix}$.

Let h be a pairwise independent hashing function from $[n]$ to $[\frac{C'}{\varepsilon^8}]$ and let H_0 denote a $\frac{C'}{\varepsilon^8} \times n$ hashing matrix where each column equals $e_{h(i)}$. Let H denote the $\frac{C'}{\varepsilon^8} \log n \times n$ block matrix $\begin{bmatrix} H_0 S_0 \\ H_0 S_1 \\ H_0 S_2 \\ \vdots \\ H_0 S_{\log n - 1} \end{bmatrix}$ with $H^{(i)} = H_0 S_i$.

Suppose that $x = \begin{bmatrix} x^{(0)} \\ x^{(1)} \\ \vdots \\ x^{(\log n - 1)} \end{bmatrix}$ is a block vector. Then we let $\widetilde{\text{nnz}}(x)$ denote $\begin{bmatrix} \|x^{(0)}\|_0 \\ \|x^{(1)}\|_0 \\ \vdots \\ \|x^{(\log n - 1)}\|_0 \end{bmatrix}$.

We will abuse notation and let $\mathcal{C}^S(x) = \widetilde{\text{nnz}}(Sx)$ and $\mathcal{C}(x) = \widetilde{\text{nnz}}(HSx)$ with the understanding that HSx and Sx are of different dimensions but have the same number of blocks.

The main idea of the sketch is that if $\|x\|_0$ is less than a small constant and the coordinates of x are hashed into a number of buckets that is a large constant, then with high probability it will be a perfect hash. Thus the number of non-zero buckets will equal $\|x\|_0$. If x is subsampled with a low enough probability, then the subsampled vector will have an ℓ_0 value that is sufficiently small and it can be hashed as we described.

We should note that the hash is needed for dimensionality reduction, not for the

sketch to be an accurate estimator. For certain proofs we will analyze properties of the sketch without the hashing step (as in $\mathcal{C}^S(x)$).

So S will sample x with different subsampling probabilities and we will expect that one will be small enough. We can then hash that subsampled vector, count the number of non-zero entries, and rescale by the sampling probability to approximate $\|x\|_0$. It then suffices to identify a suitably subsampled vector.

To do so, we will let $\tau := \frac{C}{\varepsilon^4}$ and define estimation functions $\text{est}_j : \mathbb{R}^{\log n} \to \mathbb{R}$, where $\text{est}_j(v) = \frac{v_j}{p_j}$. If j^* denotes the maximum index such that $v_{j^*} > \gamma$ (for a value of γ to be specified later) then $\text{est}(v, \gamma) = \text{est}_{j^*}(v)$. If such an index does not exist, then $\text{est}(v, \gamma) = \text{est}_0(v)$. We let $\mathcal{E}(x, \gamma) = \text{est}(\mathcal{C}(x), \gamma)$ and $\mathcal{E}_j(x) = \text{est}_j(\mathcal{C}(x))$. We will also let $\mathcal{E}^S(x, \gamma) = \text{est}(\mathcal{C}^S(x), \gamma)$ and $\mathcal{E}^S_j(x) = \text{est}_j(\mathcal{C}^S(x))$.

Note that we can replace all instances of n in the above definitions with d and our algorithm will just sketch the row space rather than the column space. We use n in our discussion just to keep the exposition similar to the case of $0 < p < 2$ and to emphasize the similarities in technique.

For ease of notation in our proofs, we will omit the parameter γ in $\mathcal{E}(x)$, $\mathcal{E}^S(x)$, $\mathcal{E}_i(x)$, and $\mathcal{E}^S_i(x)$ if it is clear that $\gamma = \tau$.

The idea is that past j^* we can be confident that we are subsampling x with so small of a probability that we barely sample any elements. On the other hand, if all the subsampled values are too small, then we can be confident that $\|x\|_0$ itself was small.

To sketch a vector it is enough to show that at the index j^*, a p_{j^*} fraction of x is sampled up to a relative error of ε. For the purposes of our low rank approximation algorithm, we will want a slightly stronger condition that the indices around j^* will be sampled "as expected" and that the value of j^* will be approximately $\log(\|x\|_0/\gamma)$.

Throughout this section, we will let L_j denote $\|S_j x\|_0$ so

$$\mathbb{E}[L_j] = p_j \|x\|_0 \text{ and } \text{Var}[L_j] = p_j(1 - p_j)\|x\|_0 \leq \mathbb{E}[L_j].$$

Definition 4.5.1. *Given a threshold γ, let $j = \max(0, \lfloor \log_2(\|x\|_0/\gamma) \rfloor)$, so $\gamma \leq \frac{\|x\|_0}{2^j} < 2\gamma$. Let j^* be the maximum index such that $\mathcal{C}(x)_{j^*} \geq \gamma$, or 0 if none exists.*

We say that $\mathcal{E}(x, \gamma)$ is a well-behaved sampling if

1. *$j^* = j - 1, j$, or $j + 1$*
2. *If $\|x\|_0 \geq \gamma$, then $\mathcal{E}_i(x, \gamma) = (1 \pm \Theta(\varepsilon))\|x\|_0$ for $i = j - 1, j, j + 1$, and $j + 2$*
3. *If $\|x\|_0 < \gamma$, then $L_1 < 3\gamma/4$*

To prove the correctness of our sketch, it will suffice to prove that with high probability our samplings are well-behaved samplings. We will need a folklore fact about pairwise independent hashing (the proof is included for completeness).

Fact 4.5.1. *If $h : [n] \to [m]$ is a pairwise independent hash function and $m \geq \Omega(n^2/\varepsilon)$, then with probability at least $1 - \Theta(\varepsilon)$, h will perfectly hash $[n]$.*

Proof. For $i \neq j \in [n]$ let $I_{i,j}$ be an indicator variable for the event $h(i) = h(j)$. Then $I = \sum_{i \neq j} I_{i,j}$ is the total number of collisions. We have $\mathbb{E}[I] = \sum_{i \neq j} \mathbb{E}[I_{i,j}] = \sum_{i \neq j} \frac{1}{m} \leq \frac{n^2}{m} \leq O(\varepsilon)$. By Markov's Inequality, $\mathbf{Pr}\Big[I \geq 1\Big] \leq O(\varepsilon)$ and the result follows. □

To make use of this fact we will set C' to be significantly larger than C^2. These hash sizes are chosen such that they are at least $\Omega(\gamma^2)$. Thus any subsampling past level j^* will likely result in a perfect hashing.

Lemma 4.5.2. *If $O(1/\varepsilon^4) > \gamma > \Omega(1/\varepsilon^3)$, then with probability at least $1 - \Theta(\varepsilon)$ over the randomness of S and H, $\mathcal{E}(x, \gamma)$ is a well-behaved sampling.*

In particular, this holds when $\gamma = \tau$ or $\gamma = \varepsilon\tau$.

Proof. We let j^* and v be as given in the definition of well-behaved. First we consider the case when $\|x\|_0 \geq \gamma$.

Note that for $i = j - 1, j, j + 1$, or $j + 2$, we have $\mathbb{E}[L_i] \geq \|x\|_0/2^{j+2} \geq \gamma/4$. Since $\mathrm{Var}[L_i] \leq \mathbb{E}[L_i]$, then by Chebyshev's Inequality, we know

$$\mathbf{Pr}\Big[L_i \notin (1 \pm \varepsilon)\mathbb{E}[L_i]\Big] \leq \left(\frac{\sqrt{\mathrm{Var}[L_i]}}{\varepsilon\mathbb{E}[L_i]}\right)^2 \leq \frac{1}{\varepsilon^2\mathbb{E}[L_i]} \leq \frac{4}{\varepsilon^2\gamma} \leq O(\varepsilon).$$

For the given values of i, we have $\mathbb{E}[L_i] \leq \|x\|_0/2^{j-1} \leq 4\gamma$. Since H_0 hashes to a range of size $C'/\varepsilon^8 > (4\gamma)^2$, then by Fact 4.5.1, H_0 will perfectly hash the non-zero entries of $S_i x$ for the given values of i with probability at least $1 - \Theta(\varepsilon)$.

By a union bound, $\mathcal{C}(x)_i = (1 \pm \varepsilon)\mathbb{E}[L_i]$ for $i = j - 1, j, j + 1$, or $j + 2$ with probability at least $1 - \Theta(\varepsilon)$. Thus

$$\mathbf{Pr}\Big[\mathcal{E}_i(x) = (1 \pm \varepsilon)\|x\|_0\Big] = \mathbf{Pr}\Big[\mathcal{C}(x)_i = (1 \pm \varepsilon)L_i\Big] \geq 1 - \Theta(\varepsilon)$$

which satisfies (ii).

As we argued above, with probability at least $1 - \Theta(\varepsilon)$ both $\mathcal{C}(x)_{j-1} \geq (1 - \varepsilon)\mathbb{E}[L_{j-1}] \geq 3\gamma/2$ and $\mathcal{C}(x)_{j+2} \leq (1 + \varepsilon)\mathbb{E}[L_{j+2}] \leq 3\gamma/4$ hold. By the nestedness of our sampling procedure, for any $i > j + 2$ we have $\mathcal{C}(x)_i \leq 3\gamma/4$. Thus $j^* = j - 1, j$, or $j + 1$ which satisfies (i).

Now suppose $\|x\|_0 < \gamma$. This implies $j = 0$ and $j^* = 0$ by definition which satisfies (i). If $\|x\|_0 \geq \gamma/2$, then by our reasoning above, $L_1 < 3\gamma/4$ with probability at least $1 - \Theta(\varepsilon)$. If $\|x\|_0 < \gamma/2$, then $L_1 < \gamma/2$ by the nestedness property of our sampling procedure. Therefore (iii) is satisfied. □

It follows that for a given x, with probability at least $1-\Theta(\varepsilon)$, $\mathcal{E}(x) = (1\pm\varepsilon)\|x\|_0$. We can also get tail bounds for $\mathcal{E}_j(x)$ $(j = 1, \ldots, \log n)$ and $\mathcal{E}(x)$.

Lemma 4.5.3. *Let M be a large constant. Then $\mathbf{Pr}\Big[\mathcal{E}_i(x) > M\|x\|_0\Big] \leq \frac{1}{M}$ (for arbitrary i) and $\mathbf{Pr}\Big[\mathcal{E}(x) > M\|x\|_0\Big] \leq O\left(\frac{1}{M} + \varepsilon\right)$. Furthermore, $\mathbf{Pr}\Big[\mathcal{E}_i^S(x) > M\|x\|_0\Big] \leq \frac{1}{M}$ and $\mathbf{Pr}\Big[\mathcal{E}^S(x) > M\|x\|_0\Big] \leq O\left(\frac{1}{M} + \varepsilon\right)$.*

Proof. Let j^* be chosen so that $\mathcal{E}(x) = \frac{\mathcal{C}_{j^*}(x)}{p_{j^*}}$.

By Markov's Inequality, we have

$$\begin{aligned}
\mathbf{Pr}\Big[\mathcal{E}_i(x) > M\|x\|_0\Big] &= \mathbf{Pr}\Big[\mathcal{C}_i(x)/p_i > M\|x\|_0\Big] \\
&\leq \mathbf{Pr}\Big[L_i/p_i > M\|x\|_0\Big] \\
&\leq \frac{\mathbb{E}[L_i]}{p_i M\|x\|_0} \\
&= \frac{1}{M}
\end{aligned}$$

for an arbitrary index i as desired. Let W denote the event that $\mathcal{E}(x)$ is well-behaved. Then Lemma 4.5.2 tells us that

$$\begin{aligned}
\mathbf{Pr}\Big[\mathcal{E}(x) > M\|x\|_0\Big] &= \mathbf{Pr}\Big[W\Big]\mathbf{Pr}\Big[\mathcal{E}(x) > M\|x\|_0 \mid W\Big] \\
&\quad + \mathbf{Pr}\Big[\overline{W}\Big]\mathbf{Pr}\Big[\mathcal{E}(x) > M\|x\|_0 \mid \overline{W}\Big] \\
&\leq \frac{\mathbf{Pr}\Big[\mathcal{E}_i(x) > M\|x\|_0 \text{ for } i = j-1, j, j+1\Big]}{\mathbf{Pr}\Big[W\Big]} + O(\varepsilon) \\
&\leq O\left(\frac{1}{M} + \varepsilon\right)
\end{aligned}$$

and the result follows. □

Let $K = \text{poly}(k, 1/\delta, 1/\varepsilon)$ for some $\delta > 0$ and $\mathcal{E}^{(1)}, \ldots, \mathcal{E}^{(K)}$ be independent instances of the sketching procedure $\mathcal{E}$. Let $\mathcal{A}(x) = \begin{bmatrix} \mathcal{E}^{(1)}(x) \\ \vdots \\ \mathcal{E}^{(K)}(x) \end{bmatrix}$.

For a matrix M, we let $\mathcal{A}(M)$ denote the matrix whose ith column is $\mathcal{A}(M_{:,i})$.

We can also define $\mathcal{A}^S(M)$ the natural way.

We can study medians and quantiles of $\mathcal{A}(M)$ like we did the medians and quantiles of our sketches based on p-stable variables.

Lemma 4.5.4. *1.* $\mathbf{Pr}[q_{\frac{1}{2}-\Theta(\varepsilon)}(\mathcal{A}(x)) < (1-\varepsilon)\|x\|_0] < \exp(-\Theta(\varepsilon^2)K)$

2. $\mathbf{Pr}[q_{\frac{1}{2}+O(\varepsilon)}(\mathcal{A}(x)) > (1+\varepsilon)\|x\|_0] < \exp(-\Theta(\varepsilon^2)K)$

3. For $T > 2$, $\mathbf{Pr}[q_{1-\frac{\varepsilon}{2}}(\mathcal{A}(x)) > \frac{T}{\varepsilon}\|x\|_0] < \exp(-\Theta(\varepsilon)TK)$

4. For $T > 2$, $\mathbf{Pr}[\text{med}(\mathcal{A}(x)) > T\|x\|_0] < \exp(-\Theta(T)K)$

The analagous bounds for $\mathcal{A}^S(x)$ *also hold.*

Proof. We can use Chernoff bounds, Lemma 4.5.2, and Lemma 4.5.3 to prove this in a similar way to how the proof of Lemma 3.0.2 used Chernoff bounds and the tail bounds on Cauchy sketches. □

We can now deduce a finite field subspace embedding result.

Corollary 4.5.5. *Let* $X \subset \mathbb{F}_q^n$ *be a* k*-dimensional space. With probability at least* $1 - \Theta(\delta)$*, for all* $x \in X$*,*

$$(1-\varepsilon)\|x\|_0 \leq q_{\frac{1}{2}-\Theta(\varepsilon)}(\mathcal{A}(x)) \leq q_{\frac{1}{2}+O(\varepsilon)}(\mathcal{A}(x)) \leq (1+\varepsilon)\|x\|_0$$

and

$$(1-\varepsilon)\|x\|_0 \leq q_{\frac{1}{2}-\Theta(\varepsilon)}(\mathcal{A}^S(x)) \leq q_{\frac{1}{2}+O(\varepsilon)}(\mathcal{A}^S(x)) \leq (1+\varepsilon)\|x\|_0$$

Proof. We can use Lemma 4.5.4 and the fact that $|X| = q^k$ to deduce the result with a union bound. □

We can also bound the median of an ℓ_0 sketch of a fixed matrix.

Lemma 4.5.6. *Let* M *be an* $n \times d$ *matrix. For* $\varepsilon > 0$*, with probability* $1 - O(1)$*,*

$$(1-\varepsilon)\|M\|_0 \leq \textit{med}(\mathcal{A}(M)) \leq (1+\varepsilon)\|M\|_0$$

and

$$(1-\varepsilon)\|M\|_0 \leq \textit{med}(\mathcal{A}^S(M)) \leq (1+\varepsilon)\|M\|_0$$

Proof. The proof follows the same structure as the proof of Lemma 3.0.5 where we bound the expected sum of $\text{med}(\mathcal{A}(M_{:,i}))$ over values of i where $\text{med}(\mathcal{A}(M_{:,i}))$ is large and conclude with Markov's Inequality. Instead of Fact 3.0.1 and Lemma 3.0.2, we use Lemma 4.5.2 and Lemma 4.5.4. □

We can also bound the $(1-\varepsilon/2)$-quantile of an ℓ_0 sketch and we can bound the $(1-\varepsilon/2)$-quantile of a fixed index ℓ_0 sketch of a fixed matrix.

Lemma 4.5.7. *Let M be an $n \times d$ matrix. Let J be a set of indices with $|J| = d$. With probability $1 - O(1)$,*

$$q_{1-\varepsilon/2}(\mathcal{A}(M)) \leq O\left(\frac{1}{\varepsilon}\right) \|M\|_0$$

and

$$q_{1-\varepsilon/2}(\mathcal{A}^S(M)) \leq O\left(\frac{1}{\varepsilon}\right) \|M\|_0$$

Proof. These inequalities can be proved using the same argument that was used to prove Lemma 3.0.6, but using Lemma 4.5.4 instead of Lemma 3.0.2. □

Theorem 4.5.1. *Let $U \in \mathbb{F}_q^{n\times k}, A \in \mathbb{F}_q^{n\times d}$. With probability $1 - O(1)$,*

$$med(\mathcal{A}(UV - A)) \geq (1 - O(\varepsilon))\|UV - A\|_0$$

for arbitrary V.

Proof. Let V^* be chosen to minimize $\|UV^* - A\|_0$.

For column indices i, let $J_i = \max(0, \log(\|(UV^* - A)_{:,i}\|_0/\gamma))$

By Lemmas 4.5.2, 4.5.6, and 4.5.7, the following statements hold with probability $1 - O(1)$:

(i) $\mathcal{E}$ is well-behaved on Ux for all x

(ii) For each i with probability at least $1 - \varepsilon^3$,

$$\text{med}(\mathcal{A}([U\ A_{:,i}]x)) \geq (1 - \varepsilon^3)\|[U\ A_{:,i}]x\|_0$$

for all x

(iii) $\text{med}(\mathcal{A}(UV^* - A)) \leq (1 + \varepsilon^3)\|UV^* - A\|_0$

(iv) $q_{1-\varepsilon/2}(\mathcal{A}^S(UV^* - A, \varepsilon\tau)) \leq O\left(\frac{1}{\varepsilon}\right) \|UV^* - A\|_0$

We say a column index i is *good* if

$$\text{med}(\mathcal{A}([U\ A_{:,i}]y)) \geq (1 - \varepsilon^3)\|[U\ A_{:,i}]y\|_0$$

for all $y \in \mathbb{R}^{k+1}$, and *bad* otherwise. Let $Q_i = q_{1-\varepsilon/2}(\mathcal{A}^S(UV^* - A, \varepsilon\tau)_{:,i}))$. We say a bad column index is *large* if

$$\varepsilon\|(UV - A)_{:,i}\|_0 \geq \frac{2}{1-\varepsilon}Q_i + \|(UV^* - A)_{:,i}\|_0.$$

By (ii), we know that $\mathbb{E}[\sum_{\text{bad } i}\|(UV^* - A)_{:,i}\|_0] \leq \varepsilon^3\|UV^* - A\|_0$. By Markov's inequality, we know that with probability $1 - O(1)$,

$$\sum_{\text{bad } i}\|(UV^* - A)_{:,i}\|_0 \leq O(\varepsilon^3)\|UV^* - A\|_0 \tag{4.5}$$

By (iii)

$$\begin{aligned}(1+\varepsilon^3)\|UV^* - A\|_0 &\geq \text{med}(\mathcal{A}(UV^* - A)) \\ &\geq (1-\varepsilon^3)\sum_{\text{good } i}\|(UV^* - A)_{:,i}\|_0 + \sum_{\text{bad } i}\text{med}(\mathcal{A}(UV^* - A)_{:,i}) \\ &\geq (1-\varepsilon^3)(1-\Theta(\varepsilon^3))\|UV^* - A\|_0 + \sum_{\text{bad } i}\text{med}(\mathcal{A}(UV^* - A)_{:,i}),\end{aligned}$$

where the second inequality comes from the definition of good, and the third inequality comes from (4.5).

Thus

$$\sum_{\text{bad } i}\text{med}(\mathcal{A}(UV^* - A)_{:,i}) \leq O(\varepsilon^3)\|UV^* - A\|_0 \tag{4.6}$$

We also have

$$\begin{aligned}\sum_{\text{small } i}\|(UV - A)_{:,i}\|_0 &\leq \frac{2}{\varepsilon(1-\varepsilon)}\sum_{\text{small } i}Q_i + \frac{1}{\varepsilon}\sum_{\text{small } i}\|(UV^* - A)_{:,i}\|_0 \\ &\leq O\left(\frac{1}{\varepsilon^2(1-\varepsilon)}\right)\left(\sum_{\text{small } i}\|(UV^* - A)_{:,i}\|_0\right) + O(\varepsilon^2)\|UV^* - A\|_0 \\ &\leq O(\varepsilon)\|UV^* - A\|_0 \end{aligned} \tag{4.7}$$

where the first inequality comes from the definition of small, the second inequality comes from (iv) and (4.5) and the third inequality comes from (4.6).

Claim 4.5.8.

$$\sum_{\textit{large } i}\textit{med}(\mathcal{A}(UV - A)_{:,i}) \geq (1 - O(\varepsilon))\sum_{\textit{large } i}\|(UV - A)_{:,i}\|_0$$

Proof. Let column i be large. We have $H(UV-A)_{:,i} = HU(V-V^*)_{:,i}+H(UV^*-A)_{:,i}$.

By the triangle inequality, we have

$$\begin{aligned}
&(1-\Theta(\varepsilon))\|U(V-V^*)_{:,i}\|_0 \\
\geq&(1-\Theta(\varepsilon))(\|(UV-A)_{:,i}\|_0 - \|(UV^*-A)_{:,i}\|_0) \\
\geq&(1-\Theta(\varepsilon))((1-\varepsilon)\|(UV-A)_{:,i}\|_0 + \frac{2}{1-\varepsilon}Q_i) \\
\geq&(1-\Theta(\varepsilon))\|(UV-A)_{:,i}\|_0 + Q_i \\
\geq&(1-\Theta(\varepsilon))\|(UV-A)_{:,i}\|_0
\end{aligned}$$

where the second inequality follows from the definition of large.

Since $\|U(V-V^*)_{:,i}\|_0 \geq (1-\Theta(\varepsilon))\|(UV-A)_{:,i}\|_0$ and $\varepsilon\|(UV-A)_{:,i}\|_0 \geq Q_i$, then $Q_i/\varepsilon \leq \|U(V-V^*)_{:,i}\|_0$.

If we run K independent instances of H, then by (i), we know that at least $\frac{1}{2}+\varepsilon$ of those instances will have estimations $\mathcal{E}(U(V-V^*)_{:,i})$ that are well-behaved and satisfy $\mathcal{E}(U(V-V^*)_{:,i}) \geq (1-\Theta(\varepsilon))\|U(V-V^*)_{:,i}\|_0$.

At least $1-\varepsilon/2$ of those instances will satisfy $Q_i > \text{est}^S((UV^*-A)_{:,i}, \varepsilon\tau)$. In each of these instances, there is some index t which is the maximum index where $\mathcal{C}^S((UV^*-A)_{:,i}) > \varepsilon\tau$. This index t satisfies $\text{est}((UV^*-A)_{:,i}, \varepsilon\tau) \geq 2^t\varepsilon\tau$ which implies that

$$t \leq \log_2\left(\frac{\text{est}^S((UV^*-A)_{:,i}, \varepsilon\tau)}{\varepsilon\tau}\right) < \log_2\left(\frac{Q_i}{\varepsilon\tau}\right) \leq \log_2\left(\frac{\|U(V-V^*)_{:,i}\|_0}{\tau}\right) \leq J_i$$

and by the nestedness property of S, for every index $l \geq J_i - 1$ we have $\mathcal{C}^S((UV^*-A)_{:,i})_l < \varepsilon\tau$. Furthermore, $\mathcal{C}((UV^*-A)_{:,i})_l < \varepsilon\tau$ because $\|H_0y\|_0 \leq \|y\|_0$ for all y.

Thus, for at least $\frac{1}{2}+\varepsilon/2$ instances of H, it is true that $\mathcal{E}(U(V-V^*)_{:,i})$ is well-behaved and for every index $l \geq J_i - 1$ we have $\mathcal{C}((UV^*-A)_{:,i})_l < \varepsilon\tau$. We first consider the case that $\|U(V-V^*)_{:,i}\|_0 > \tau$.

We know that for $l = J_i-1$, J_i, or J_i+1, one of those values will be the maximum value such that the lth block of $HU(V-V^*)_{:,i}$ has at least τ non-zero entries, and all the later blocks will have at most $3\tau/4$ non-zero entries. Each block of $H(UV^*-A)_{:,i}$ after the J_i-1th one will have fewer than $\varepsilon\tau$ non-zero entries. By well-behavedness, it follows that $\mathcal{E}(UV-A)_{:,i} = \mathcal{E}(U(V-V^*)_{:,i}+(UV^*-A)_{:,i}) \geq (1-\Theta(\varepsilon))\|U(V-V^*)_{:,i}\|_0$ because the salient blocks of $H(UV-A)_{:,i}$ will have a number of non-zero entries differing from those blocks of $HU(V-V^*)_{:,i}$ by an additive $\Theta(\varepsilon)$ error.

If $\|U(V-V^*)_{:,i}\|_0 \leq \tau$, then by well-behavedness we know that block 1 of $HU(V-V^*)_{:,i}$ will have fewer than $3\tau/4$ non-zero entries. In this case all blocks of $H(UV^*-A)_{:,i}$ will have fewer than $\varepsilon\tau$ non-zero entries so all blocks of $H(UV-A)_{:,i}$ besides

the zeroth block will have fewer than τ non-zero entries. Thus, $\mathcal{E}(UV-A)_{:,i} \geq (1-\Theta(\varepsilon)) \geq \|U(V-V^*)_{:,i}\|_0$.

Therefore in a majority of the instances of H, we have $\mathcal{E}(UV-A)_{:,i} \geq (1-\Theta(\varepsilon))\|U(V-V^*)_{:,i}\|_0 \geq (1-\Theta(\varepsilon))\|(UV-A)_{:,i}\|_0$ and the result follows. □

Finally,

$$\begin{aligned}
\operatorname{med}(\mathcal{A}(UV-A)) &\geq \sum_{\text{good } i} \operatorname{med}(\mathcal{A}(UV-A)_{:,i}) + \sum_{\text{large } i} \operatorname{med}(\mathcal{A}(UV-A)_{:,i}) \\
&\geq (1-\varepsilon^3) \sum_{\text{good } i} \|(UV-A)_{:,i}\|_0 + (1-O(\varepsilon)) \sum_{\text{large } i} \|(UV-A)_{:,i}\|_0 \\
&\geq (1-O(\varepsilon))\|UV-A\|_0 - (1-O(\varepsilon)) \sum_{\text{small } i} \|(UV-A)_{:,i}\|_0 \\
&\geq (1-O(\varepsilon))\|UV-A\|_0 - (1-O(\varepsilon))O(\varepsilon)\|UV^*-A\|_0 \\
&\geq (1-O(\varepsilon))\|UV-A\|_0
\end{aligned}$$

where the first inequality occurs because large i are bad i, the second inequality comes from the definition of good and Claim 4.5.8, the third inequality comes from the definition of small, the fourth inequality comes from (4.7), and the last inequality holds because V^* is a minimizer. □

Theorem ($\mathbb{F}_q$ PTAS for $p=0$)**.** *For $\varepsilon \in (0,1)$ there is a $(1+\varepsilon)$-approximation algorithm to Entrywise ℓ_0-Rank-k Approximation over $\mathbb{F}_q$ running in $n \cdot d^{\operatorname{poly}(k/\varepsilon)}$ time.*

Proof. Suppose U^* and V^* ($n\times k$ and $k\times d$ respectively) are minimizers for $\|UV-A\|_0$. By Theorem 4.5.1, $\operatorname{med}(\mathcal{A}(U^*V-A)) = (1\pm\varepsilon)\|U^*V-A\|_0$. Since H is a $\frac{C'}{\varepsilon^4}\log n \times n$ block matrix, then HU^* has $\frac{C'}{\varepsilon^4}k \cdot \log n$ entries and we need K instances of HU^* for a total of $(\log n)\cdot \operatorname{poly}(k/\varepsilon)$ entries each having q possible values. Thus we can exhaustively guess all possible values of HU^* in $n^{\operatorname{poly}(k/\varepsilon)}$ time.

For each guess of HU^* and each column i, we can try all q^k possible vectors V_i and choose the minimizer. Once a V has been identified, we can solve for its optimal U and throughout this whole process keep the best U and V that minimize $\|UV-A\|_0$. Since there are d rows, the algorithm will have a total runtime of $d \cdot n^{\operatorname{poly}(k/\varepsilon)}$.

As we stated in the opening exposition of this section, we could have sketched over the dimension d row space instead. In this case we would be guessing values for $H(V^*)^T$, a $\frac{C'}{\varepsilon^4}\log d \times k$ matrix, which would take $d^{\operatorname{poly}(k/\varepsilon)}$ time. We would then minimize over each of the n rows of U for a total runtime of $n \cdot d^{\operatorname{poly}(k/\varepsilon)}$. □

Chapter 5

Hardness

In this chapter, we prove hardness of approximately computing the best rank k-approximation of a given $n \times d$ matrix A, where $n \geq d$. Indeed all hardness results in this chapter hold when $k = d-1$, indicating that reducing the rank by 1 is indeed hard to even approximate. This complements our efficient approximation schemes when $k = O(1)$.

Our results for $p \in (1, 2)$ assume the Small Set Expansion Hypothesis. Originally conjectured by Raghavendra and Stuerer [78], it is still the only assumption that implies strong hardness results for various graph problems such as Uniform Sparsest Cut [80] and Bipartite Clique [61]. Assuming this hypothesis, we prove even stronger results than above that rules out *any* constant factor approximation in $\text{poly}(n, k)$. The following theorem immediately implies Theorem 2.1.3 in the introduction.

Theorem 5.0.1. *Fix $p \in (1, 2)$ and $r > 1$. Assuming the Small Set Expansion Hypothesis, there is no r-approximation algorithm for rank k approximation of a matrix $A \in \mathbb{R}^{n \times d}$ with $n \geq d$ and $k = d - 1$ in the entrywise ℓ_p norm that runs in time poly(n).*

Consequently, additionally assuming the Exponential Time Hypothesis, there exists $\delta := \delta(p, r) > 0$ such that there is no r-approximation algorithm for rank k approximation of a matrix $A \in \mathbb{R}^{n \times d}$ with $n \geq d$ and $k = d - 1$ in the entrywise ℓ_p norm that runs in time 2^{n^δ}.

For $p \in (2, \infty)$, we do not rely on the Small Set Expansion Hypothesis though the hardness factor is bounded by a constant. Recall that $\gamma_p := \mathbf{E}_g[|g|^p]^{1/p}$ where g is a standard Gaussian, which is strictly greater than 1 for $p > 2$.

Theorem 5.0.2. *Fix $p \in (2, \infty)$ and $\varepsilon > 0$. Assuming $\mathsf{P} \neq \mathsf{NP}$, there is no $(\gamma_p^p - \varepsilon)$ - approximation algorithm for rank k approximation of a matrix $A \in \mathbb{R}^{n \times d}$ with $n \geq d$ and $k = d - 1$ in the entrywise ℓ_p norm that runs in time poly(n).*

Consequently, assuming the Exponential Time Hypothesis, there exists

$$\delta := \delta(p, \varepsilon) > 0$$

such that there is no $(\gamma_p^p - \varepsilon)$ - approximation algorithm for rank k approximation of a matrix $A \in \mathbb{R}^{n \times d}$ with $n \geq d$ and $k = d - 1$ in the entrywise ℓ_p norm that runs in time 2^{n^δ}.

We also prove similar hardness results for ℓ_0-low rank approximation in finite fields. The following theorem immediately implies Theorem 2.1.4 in the introduction.

Theorem 5.0.3. *Fix a finite field $\mathbb{F}$ and $r > 1$. Assuming $\mathsf{P} \neq \mathsf{NP}$, there is no r-approximation algorithm for rank k approximation of a matrix $A \in \mathbb{F}^{n \times d}$ with $n \geq d$ and $k = d - 1$ in the entrywise ℓ_0 metric that runs in time poly(n).*

Consequently, assuming the Exponential Time Hypothesis, there exists $\delta > 0$ (dependent on r) such that there is no r-approximation algorithm for rank k approximation of a matrix $A \in \mathbb{F}^{n \times d}$ with $n \geq d$ and $k = d - 1$ in the entrywise ℓ_0 metric that runs in time 2^{n^δ}.

Section 5.1 proves Lemma 2.2.1, showing that computing $\min_{p^* \to p}(A)$ is equivalent to finding the best rank k approximation of $A \in \mathbb{R}^{n \times d}$ when $n \geq d$ and $k = d - 1$. Section 5.2 proves Lemma 2.2.2, reducing $\|\cdot\|_{2 \to p^*}$ to $\min_{p^* \to p}(\cdot)$. Section 5.3 presents the Barak et al. [5]'s proof of hardness of $\|\cdot\|_{2 \to p^*}$ with modifications for all $q > 2$, finishing the proof of Theorem 5.0.1 for $p \in (1, 2)$. Section 5.4 proves the hardness of $\min_{p^* \to p}(\cdot)$ for $p > 2$, using the result of [36], and finishes the proof of Theorem 5.0.2. Finally, Theorem 5.0.3 is proved in Section 5.5.

Numerical issues. In the proofs of Theorem 5.0.1 and Theorem 5.0.2, we consider our matrices as having real entries for simplicity, but our results will hold even when all entries are rescaled to polynomially bounded integers. The instance in Theorem 5.0.2 is explicitly constructed and it can be easily checked that all entries are polynomially bounded integers. For Theorem 5.0.1, our hard instance B for $\|\cdot\|_{p \to p^*}$ is simply a projection matrix and the final instance A is obtained by $(\varepsilon I + B)^{-1}$, so by ensuring that $\varepsilon \geq 1/\text{poly}(n)$, we can ensure that eigenvalues of A are within $[1, \text{poly}(n)]$.

5.1 ℓ_p-Low Rank Approximation and $\min_{p^* \to p}(A)$

In this section, we prove the following lemma showing that computing $\min_{p^* \to p}(A)$ is equivalent to finding the best rank k approximation of $A \in \mathbb{R}^{n \times d}$ when $n \geq d$ and $k = d - 1$.

Lemma (Restatement of Lemma 2.2.1)**.** *Let $p \in (1, \infty)$. Let $A \in \mathbb{R}^{n \times d}$ with $n \geq d$ and $k = d - 1$. Then*

$$\min_{U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{k \times d}} \|UV - A\|_p = \min_{x \in \mathbb{R}^d, \|x\|_{p^*} = 1} \|Ax\|_p.$$

Proof. Assume that the rank of A is d; otherwise the lemma becomes trivial. We first prove $(\geq)$. Given $V^* \in \mathbb{R}^{k \times d}$ that achieves the best rank k approximation, assume without loss of generality that the rank of V^* is $k = d - 1$. Let $x \in \mathbb{R}^d$ be the unique vector (up to sign) that is orthogonal to the rowspace of V^* and $\|x\|_{p^*} = 1$. Let $a_1, \ldots, a_n$ be the rows of A. For fixed V^*, for $i \in [n]$, the ith row $u_i^* \in \mathbb{R}^k$ of U^* must be obtained by computing

$$\min_{u_i^* \in \mathbb{R}^k} \|u_i^* V^* - a_i\|_p = \min_{y \in \text{rowspace}(V^*)} \|y - a_i\|_p = \min_{z \in \mathbb{R}^d : \langle x, z \rangle = -\langle x, a_i \rangle} \|z\|_p.$$

Note that by Hölder's inequality, the last quantity is at least

$$|\langle x, z \rangle| / \|x\|_{p^*} = |\langle x, a_i \rangle| / \|x\|_{p^*} = |\langle x, a_i \rangle|.$$

Indeed, taking $z \in \mathbb{R}^d$ with $z_j := (-\langle x, a_i \rangle) \cdot (\text{sgn}(x_j)|x_j|^{p^*/p})$ for each $j \in [d]$ implies

$$\langle x, z \rangle = -\langle x, a_i \rangle \cdot \sum_{j \in [d]} \text{sgn}(x_j)|x_j|^{p^*/p} \cdot x_j = -\langle x, a_i \rangle \cdot \|x\|_{p^*}^{p^*} = -\langle x, a_i \rangle,$$

and

$$\|z\|_p = |\langle x, a_i \rangle| \cdot \Big(\sum_{j \in [d]} |x_j|^{p^*} \Big)^{1/p} = |\langle x, a_i \rangle| \cdot \|x\|_{p^*}^{p^*/p} = |\langle x, a_i \rangle|,$$

so we can conclude $\|u_i^* V^* - a_i\|_p = \min_{z \in \mathbb{R}^d : \langle x, z \rangle = -\langle x, a_i \rangle} \|z\|_p = |\langle x, a_i \rangle|$. Summing over $i \in [n]$,

$$\|U^* V^* - A\|_p = \Big(\sum_{i \in [n]} \|u_i^* V^* - a_i\|_p^p \Big)^{1/p} = \Big(\sum_{i \in [n]} |\langle x, a_i \rangle|^p \Big)^{1/p} = \|Ax\|_p.$$

This proves that $\min_{U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{k \times d}} \|UV - A\|_p \geq \min_{x \in \mathbb{R}^d, \|x\|_{p^*} = 1} \|Ax\|_p$. For the other direction, given $x \in \mathbb{R}^d$ with $\|x\|_{p^*} = 1$, let $V^* \in \mathbb{R}^{k \times d}$ be a matrix whose rowspace is a k-dimensional subspace orthogonal to x, and compute U^* as above. The above analysis shows that $\|U^* V^* - A\|_p = \|Ax\|_p$, which completes the proof. □

5.2 Reducing $\|\cdot\|_{2\to p^*}$ to $\min_{p^*\to p}(\cdot)$

In this section, we show that computing $\min_{p^*\to p}(\cdot)$ is as hard as computing $\|\cdot\|_{2\to p^*}$, proving the following lemma.

Lemma (Restatement of Lemma 2.2.2)**.** *For any $\varepsilon > 0, p \in (1,\infty)$, there is an algorithm that runs in $\mathrm{poly}(n, \log(1/\varepsilon))$ and on a non-zero input matrix A, computes a matrix B satisfying*

$$(1-\varepsilon)\|A\|_{2\to p^*}^{-2} \le \min_{p^*\to p}(B) \le (1+\varepsilon)\|A\|_{2\to p^*}^{-2}.$$

The lemma is proved in the following two steps.

Reducing $\|\cdot\|_{2\to p^*}$ to $\|\cdot\|_{p\to p^*}$. We first prove the following claim. This follows from standard tools from Banach space theory that *factor* an operator from ℓ_p to ℓ_p^* via ℓ_2.

Claim 5.2.1. $\|AA^T\|_{p\to p^*} = \|A\|_{2\to p^*}^2$.

Proof. By the definitions of $p \to q$ norms,

$$\begin{aligned}\|AA^T\|_{p\to p^*} &= \sup_x \tfrac{\|AA^Tx\|_{p^*}}{\|x\|_p}\\ &\le \sup_x \tfrac{\|A\|_{2\to p^*}\|A^Tx\|_2}{\|x\|_p}\\ &\le \|A\|_{2\to p^*}\|A^T\|_{p\to 2}\\ &= \|A\|_{2\to p^*}^2,\end{aligned}$$

where the last line follows from the fact that

$$\|A\|_{2\to p^*} = \sup_{\|y\|_p=1}\sup_{\|x\|_2=1}\langle y, Ax\rangle = \sup_{\|x\|_2=1}\sup_{\|y\|_p=1}\langle A^Ty, x\rangle = \|A^T\|_{p\to 2}.$$

For the other direction,

$$\begin{aligned}\|AA^T\|_{p\to p^*} &= \sup_{\|x\|_p=1}\sup_{\|y\|_p=1}\langle y, AA^Tx\rangle = \sup_{\|x\|_p=1}\sup_{\|y\|_p=1}\langle A^Ty, A^Tx\rangle\\ &\ge \sup_{\|x\|_p=1}\|A^Tx\|_2^2 = \|A^T\|_{p\to 2}^2 = \|A\|_{2\to p^*}^2,\end{aligned}$$

which completes the proof. □

Reducing $\|\cdot\|_{p\to p^*}$ to $\min_{p^*\to p}(\cdot)$. We now relate two quantities $\|A\|_{p\to p^*}$ and $\min_{p^*\to p}(B)$ for two related matrices A and B. If A is invertible, this can be seen easily.

Fact 5.2.2. *If A is an invertible matrix, then $\min_{p\to q}(A^{-1}) = (\|A\|_{q\to p})^{-1}$*

Proof. First observe that the condition $A^{-1}x \neq 0$ is equivalent to the condition $x \neq 0$ since A is invertible. Then we have,

$$\inf_{x\neq 0}\frac{\|A^{-1}x\|_q}{\|x\|_p} = \inf_{Ax\neq 0}\frac{\|A^{-1}x\|_q}{\|x\|_p} = \left(\sup_{A^{-1}x\neq 0}\frac{\|x\|_p}{\|A^{-1}x\|_q}\right)^{-1} = \left(\sup_{y\neq 0}\frac{\|A^{-1}y\|_p}{\|y\|_q}\right)^{-1}.$$

The leftmost quantity is $\min_{p\to q}(A^{-1})$ and the rightmost quantity is $(\|A\|_{q\to p})^{-1}$. □

Even if A is not invertible, there is an invertible matrix B whose $p\to q$ norm is close to that of A for any p and q.

Claim 5.2.3. *Let A be a non-zero $n\times d$ matrix. For any $p, q \in (1,\infty)$ and any $\varepsilon > 0$, there is an invertible and polynomial time computable $\max(n,d)\times\max(n,d)$ matrix B such that $(1-\varepsilon)\|A\|_{p\to q} \leq \|B\|_{p\to q} \leq (1+\varepsilon)\|A\|_{p\to q}$.*

Proof. Let $\oplus$ denote vector concatenation. We start by exhibiting a square matrix with the same norm. If $d \geq n$, we pad 0's to the bottom of A to obtain an $d\times d$ matrix A'. Now for any $x\in\mathbb{R}^d$, $\|A'x\|_q = \|Ax\oplus 0^{d-n}\|_q = \|Ax\|_q$. So $\|A\|_{p\to q} = \|A'\|_{p\to q}$.

If $d \leq n$, we pad 0's to the right of A to obtain an $n\times n$ matrix A'. Consider any $y\in\mathbb{R}^n$ and let $x\in\mathbb{R}^d$, $z\in\mathbb{R}^{n-d}$ be such that $y = x\oplus z$. Then we have $\|A'y\|_q = \|Ax\|_q$. Now since $\|y\|_p \geq \|x\|_p$, we have $\|A\|_{p\to q} \geq \|A'\|_{p\to q}$. On the other hand, $\|A\|_{p\to q} \leq \|A'\|_{p\to q}$ since $\|A'(x\oplus 0^{n-d})\|_q = \|Ax\|_q$ and $\|x\oplus 0^{n-d}\|_p = \|x\|_p$.

Next to obtain an invertible matrix, we set $B := A' + \varepsilon'\cdot \mathrm{I}$ where $\varepsilon' := \varepsilon\cdot M/\|\mathrm{I}\|_{p\to q}$ and M is the max magnitude of an entry of A which must be non-zero since A is non-zero. First we observe that $\|A\|_{p\to q} \geq M$ since one can substitute $x = e_i$ where i is the index of the column containing the max magnitude entry. Lastly, applying triangle inequality (since $\|\cdot\|_{p\to q}$ is a norm) implies the claim. □

5.3 Hardness of $2\to q$ norm for all $q\in(2,\infty)$

In this section, we prove Theorem 2.2.1 for hardness of $\|\cdot\|_{2\to q}$ for $q\in(2,\infty)$. Barak et al. [5] proved that under the Small Set Expansion Hypothesis, for any $r > 1$ and an even integer $q \geq 4$, it is NP-hard to approximate the $2\to q$ norm problem within a factor r. The same proof essentially works for all $q\in(2,\infty)$ with slight modifications. For completeness, we present their proof here, with additional remarks when we generalize an even integer $q\geq 4$ to all $q\in(2,\infty)$.

Preliminaries for Small Set Expansion. For a vector $x \in \mathbb{R}^d$, every p-norm in this section denotes the expectation norm defined as $\|x\|_{L_p} := (\mathbf{E}_{i\in[d]}[|x_i|^p])^{1/p}$. For a regular graph $G = (V, E)$ and a subset $S \subseteq V$, we define the measure of S to be $\mu(S) = |S|/|V|$ and we define $G(S)$ to be the distribution obtained by picking a random $x \in S$ and then outputting a random neighbor y of x. We define the expansion of S to be

$$\Phi_G(S) = \mathbf{Pr}_{y\in G(S)}[y \notin S].$$

For $\delta \in (0, 1)$, we define $\Phi_G(\delta) = \min_{S\subseteq V:\mu(S)\leq\delta} \Phi_G(S)$. We identify G with its normalized adjacency matrix. For every $\lambda \in [-1, 1]$, we denote by $V_{\geq\lambda}(G)$ the subspace spanned by the eigenvectors of G with eigenvalue at least λ. The projector into this subspace is denoted $P_{\geq\lambda}(G)$. For a distribution D, we let $\mathrm{cp}(D)$ denote the collision probability of D (the probability that two independent samples from D are identical). The Small Set Expansion Hypothesis, posed by Raghavendra and Steurer [78] states the following.

Conjecture 5.3.1. *For any $\varepsilon > 0$, there exists $\delta > 0$ such that it is NP-hard to decide whether $\Phi_G(\delta) \leq \varepsilon$ or $\Phi_G(\delta) \geq 1 - \varepsilon$.*

This implies strong hardness results for various graph problems such as Uniform Sparsest Cut [80] and Bipartite Clique [61]. The main theorem of this section is the following, which corresponds to Theorem 2.4 of [5].

Theorem 5.3.1. *For every regular graph $G, \lambda \in (0, 1)$, and $q \in (2, \infty)$,*

1. *For all $\delta > 0, \varepsilon > 0$, $\|P_{\geq\lambda}(G)\|_{L_2\to L_q} \leq \varepsilon/\delta^{(q-2)/2q}$ implies that $\Phi_G(\delta) \geq 1 - \lambda - \varepsilon^2$.*

2. *There is a constant $a = a(q)$ such that for all $\delta > 0$, $\Phi_G(\delta) > 1 - a\lambda^{2q}$ implies $\|P_{\geq\lambda}(G)\|_{L_2\to L_q} \leq 2/\sqrt{\delta}$.*

Given this theorem, the hardness of $2 \to q$ norm can be proved as follows. This corresponds to Corollary 8.1 of [5].

Proof of Theorem 2.2.1. Using [79], the Small Set Expansion Hypothesis implies that for any sufficiently small numbers $0 < \delta \leq \delta'$, there is no polynomial time algorithm that can distinguish between the following cases for a given graph G:

- Yes case: $\Phi_G(\delta) < 0.1$.
- No case: $\Phi_G(\delta') > 1 - 2^{-a' \log(1/\delta')}$. ($a'$ is a fixed universal constant.)

In particular, for all $\eta > 0$, if we let $\delta' = \delta^{(q-2)/8q}$ and make δ small enough, then in the No case $\Phi_G(\delta^{(q-2)/8q}) > 1 - \eta$. (Since $q > 2$, $\delta' \to 0$ as $\delta \to 0$.)

Using Theorem 5.3.1, in the Yes case we know $\|P_{\geq 1/2}\|_{L_2 \to L_q} \geq 1/(10\delta^{(q-2)/2q})$, while in the No case, if we choose δ sufficiently small so that η is smaller than $a(1/2)^{2q}$, then we know that $\|P_{\geq 1/2}\|_{L_2 \to L_q} \leq 2/\sqrt{\delta'} = 2/\delta^{(q-2)/4q}$. The gap between the Yes case and the No case is at least $\delta^{-(q-2)/4q}/20$, which goes to ∞ as δ decreases. □

We now prove Theorem 5.3.1. The first part that proves small set expansion of G given a $2 \to q$ norm bound indeed follows from older work (e.g., [52]).

Lemma 5.3.1 (Lemma B.1 of [5]). *For all $\delta > 0, \varepsilon > 0$, $\|P_{\geq\lambda}(G)\|_{L_2 \to L_q} \leq \varepsilon/\delta^{(q-2)/2q}$ implies that $\Phi_G(\delta) \geq 1 - \lambda - \varepsilon^2$*

Proof. Let $q^* = q/(q-1)$ be the Hölder conjugate of q such that $1/q + 1/q^* = 1$. Since $P_{\geq\lambda}$ is a projector,

$$\|P_{\geq\lambda}(G)\|_{L_{q^*} \to L_2} = \|P_{\geq\lambda}(G)^T\|_{L_{q^*} \to L_2} = \|P_{\geq\lambda}(G)\|_{L_2 \to L_q}.$$

Given $S \subseteq V$ with $\mu(S) = \mu \leq \delta$, let $f = 1_S/\sqrt{\mu}$ be the normalized indicator vector of S so that $\|f\|_{L_2} = 1$. Let $f = f' + f''$ where f' is its projection to the eigenvalues at least λ (i.e., $f' = P_{\geq\lambda} f$) and f'' is its projection to the eigenvalues strictly less than λ. Since $\|1_S\|_{L_{q^*}} = \mu^{1/q^*} = \mu^{(q-1)/q}$, we have $\|f\|_{L_{q^*}} = \mu^{((q-1)/q)-1/2} \leq \delta^{((q-1)/q)-1/2}$ (since $q > 2$ and $\delta \geq \mu$), and

$$\|f'\|_{L_2} \leq \|f\|_{L_{q^*}} \cdot \|P_{\geq\lambda}(G)\|_{L_{q^* \to 2}} \leq \delta^{((q-1)/q)-1/2} \cdot (\varepsilon/\delta^{(q-2)/2q}) = \varepsilon.$$

Then

$$\langle f, Gf \rangle = \langle f', Gf' \rangle + \langle f'', Gf'' \rangle \leq \|f'\|_{L_2}^2 + \lambda \|f''\|_{L_2}^2 \leq \varepsilon^2 + \lambda.$$

Since $\Phi_G(S) = 1 - \langle f, Gf \rangle$, the lemma follows. □

The second part of Theorem 5.3.1 requires more technical proofs.

Lemma 5.3.2 (Lemma 8.2 of [5]). *There is a constant $a = a(q)$ such that for all $\delta > 0$, $\Phi_G(\delta) > 1 - a\lambda^{2q}$ implies $\|P_{\geq\lambda}(G)\|_{L_2 \to L_q} \geq 2/\sqrt{\delta}$.*

Proof. Let f be a function in $V_{\geq\lambda}$ with $\|f\|_{L_2} = 1$ that maximizes $\|f\|_{L_q}$. We write $f = \sum_{i=1}^m \alpha_i \chi_i$ where $\chi_1, \ldots, \chi_m$ denote the eigenfunctions of G with values $\lambda_1, \ldots, \lambda_m$ that are at least λ. Assume towards contradiction that $\|f\|_{L_q} < 2/\sqrt{\delta}$. We will prove that $g = \sum_{i=1}^m (\alpha_i/\lambda_i)\chi_i$ satisfies $\|g\|_{L_q} \geq 5\|f\|_{L_q}/\lambda$. Note that g is defined such that $f = Gg$. This is a contradiction since (using $\lambda_i \in [\lambda, 1]$) $\|g\|_{L_2} \leq \|f\|_{L_2}/\lambda$, and we assumed f is a function in $V_{\geq\lambda}$ with a maximal ratio $\|f\|_{L_q}/\|f\|_{L_2}$.

Let $U \subseteq V$ be the set of vertices such that $|f(x)| \geq 1/\sqrt{\delta}$ for all $x \in U$. Using the Markov inequality and the fact that $\mathbf{E}_{x \in V}[f(x)^2] = 1$, we know that $\mu(U) = |U|/|V| \leq \delta$. On the other hand, because $\|f\|_{L_q}^q \geq 2^q/\delta^{q/2}$, we know that U contributes at least half of the term $\|f\|_{L_q}^q = \mathbf{E}_{x \in V}[|f(x)|^q]$. That is, if we define α to be $\mu(U)\,\mathbf{E}_{x \in U}[|f(x)|^q]$ then $\alpha \geq \|f\|_{L_q}^q/2$. We will prove the lemma by showing that $\|g\|_{L_q}^q \geq (10\lambda^{-1})^q \alpha$.

Let $c = c(q)$ and $d = d(c, q)$ be sufficiently large constants that will be determined later, and $e = d \cdot \lambda^{-q}$. By the variant local Cheeger bound obtained in Theorem 2.1 of [88], there exists $a = a(d, q)$ such that $\Phi_G(\delta) > 1 - a\lambda^{2q}$ implies that $\mathrm{cp}(G(S)) \leq 1/(e|S|)$ for all S with $\mu(S) \leq \delta$.

We define U_i to be the set $\{x \in U : f(x) \in [c^i/\sqrt{\delta}, c^{i+1}/\sqrt{\delta}]\}$, and let I be the maximal i such that U_i is non-empty. Thus, the sets $U_0, \ldots, U_I$ form a partition of U (where some of these sets may be empty). We let α_i be the contribution of U_i to α. That is, $\alpha_i = \mu_i\,\mathbf{E}_{x \in U_i}[|f(x)|^q]$, where $\mu = \mu(U_i)$. Note that $\alpha = \alpha_0 + \cdots + \alpha_I$. We will show that there are some indices $i_1, \ldots, i_J$ such that

1. $\alpha_{i_1} + \cdots + \alpha_{i_J} \geq \alpha/(2c^q)$.
2. For all $j \in [J]$, there is a non-negative function $g_j : V \to \mathbb{R}$ such that $\mathbf{E}_{x \in V}[|g_j(x)|^q] \geq e\alpha_{i_j}/(10c^2)^{q/2}$.
3. For every $x \in V$, $g_1(x) + \cdots + g_J(x) \leq |g(x)|$.

Showing these will complete the proof, since it is easy to see that for non-negative functions g', g'' and $q \in [1, \infty)$

$$\mathbf{E}[(g'(x) + g''(x))^q] \geq \mathbf{E}[g'(x)^q] + \mathbf{E}[g''(x)^q],$$

and hence 2. and 3. imply that

$$\|g\|_{L_q}^q = \mathbf{E}[|g(x)|^q] \geq (e/(10c^2)^{q/2}) \sum_j \alpha_{i_j}. \tag{5.1}$$

Using 1., we conclude that for $e \geq 2c^q \cdot (10c^2)^{q/2} \cdot (10/\lambda)^q$, the right-hand side of (5.1) will be larger than $(10/\lambda)^q \alpha$. In particular, we set $d = d(c, q) = 2c^q \cdot (10c^2)^{q/2} \cdot 10^q$.

We find the indices $i_1, \ldots, i_J$ iteratively. We let $\mathcal{I}$ be initially the set $\{0, ..., I\}$ of all indices. For $j = 1, 2, \ldots$, we do the following as long as $\mathcal{I}$ is not empty:

- Let i_j be the largest index in $\mathcal{I}$.
- Remove from $\mathcal{I}$ every index i such that $\alpha_i \leq c^q \alpha_{i_j}/2^{i - i_j}$.

We let J denote the step we stop. Note that our indices $i_1, \ldots, i_J$ are sorted in descending order. For every step j, the total of the α_i's for all indices we removed is less than $c^q \alpha_{i_j}$ and hence we satisfy 1. We use the following claim, whose proof is omitted here since it does not involve q at all. This follows from the fact that $\mathrm{cp}(G(S)) \leq 1/(e|S|)$ for all S with $\mu(S) \leq \delta$.

Claim 5.3.3 (Claim 8.3 of [5]). *Let $S \subseteq V$ and $\beta > 0$ such that $\mu(S) \leq \delta$ and $|f(x)| \geq \beta$ for all $x \in S$. Then there is a set of size at least $e|S|$ such that $\mathbf{E}_{x \in T}[g(x)^2] \geq \beta^2/4$.*

We will construct the functions $g_1, \ldots, g_J$ by applying iteratively Claim 5.3.3. We do the following for $j = 1, \ldots, J$:

1. let T_j be the set of size $e|U_{i_j}|$ that is obtained by applying Claim 5.3.3 to the function f and the set U_{i_j}. Note that $\mathbf{E}_{x \in T_j}[g(X)^2] \geq \beta_{i_j}^2/4$, where we let $\beta_i = c^i/\sqrt{\delta}$ (and hence for every $x \in U_i$, $\beta_i \leq |f(x)| \leq c\beta_i$).

2. Let g'_j be the function on input x that outputs $\gamma \cdot |g(x)|$ if $x \in T_j$ and 0 otherwise, where $\gamma \leq 1$ is a scaling factor that ensures that $\mathbf{E}_{x \in T_j}[g'(x)^2]$ equals exactly $\beta_{i_j}^2/4$.

3. We define $g_j(x) = \max(0, g'_j(x) - \sum_{k<j} g_k(x))$.

Note that the second step ensures $g'_j(x) \leq |g(x)|$, while the third step ensures that $g_1(x) + \cdots + g_j(x) \leq g'_j(x)$ for all j, and in particular $g_1(x) + \cdots + g_J(x) \leq |g(x)|$. Hence the only thing left to prove is the following.

Claim 5.3.4 (Claim 8.5 of [5]). $\mathbf{E}_{x \in V}[|g_j(x)|^q] \geq e\alpha_{i_j}/(10c^2)^{q/2}$.

Proof. Recall that for every i, $\alpha_i = \mu_i \, \mathbf{E}_{x \in U_i}[|f(x)|^q]$, and hence (using $f(x) \in [\beta_i, c\beta_i)$ for $x \in U_i$):

$$\mu_i \beta_i^q \leq \alpha_i \leq \mu_i c^q \beta_i^q. \tag{5.2}$$

Now fix $T = T_j$. Since $\mathbf{E}_{x \in V}[|g_j(x)|^q] = \mu(T) \cdot \mathbf{E}_{x \in T}[|g_j(x)|^q]$ and $\mu(T) = e\mu(U_{i_j})$, we can use (5.2) and $\mathbf{E}_{x \in T}[|g_j(x)|^q] \geq (\mathbf{E}_{x \in T}[g_j(x)^2])^{q/2}$ (since $q > 2$), to reduce proving the claim to showing the following:

$$\mathbf{E}_{x \in T}[g_j(x)^2] \geq (c\beta_{i_j})^2/(10c^2) = \beta_{i_j}^2/10. \tag{5.3}$$

We know that $\mathbf{E}_{x \in T}[g'_j(x)^2] = \beta_{i_j}^2/4$. We claim that (5.3) will follow by showing that for every $k < j$,

$$\mathbf{E}_{x \in T}[g'_k(x)^2] \leq 100^{-i'} \cdot \beta_{i_j}^2/4, \tag{5.4}$$

where $i' = i_k - i_j$. (Note that $i' > 0$ since in our construction the indices $i_1, \ldots, i_J$ are sorted in descending order.)

Indeed, (5.4) means that if we let momentarily $\|g_j\|_{L_2}$ denote $\sqrt{\mathbf{E}_{x\in T}[g_j(x)^t]}$ then

$$\|g_j\|_{L_2} \geq \|g'_j\|_{L_2} - \|\sum_{k<j} g_k\|_{L_2} \geq \|g'_j\|_{L_2} - \sum_{k<j}\|g_k\|_{L_2} \geq \|g'_j\|_{L_2}(1-\sum_{i'=1}^{\infty} 10^{-i'}) \geq 0.8\|g'_j\|_{L_2}. \tag{5.5}$$

The first inequality holds we can write g_j as $g'_j - h_j$, where $h_j = \min(g'_j, \sum_{k<j} g_k)$. Then, on the other hand, $\|g_j\|_{L_2} \geq \|g'_j\|_{L_2} - \|h_j\|_{L_2}$, and on the other hand, $\|h_j\|_{L_2} \leq \|\sum_{k<j} g_k\|_{L_2}$ since $g'_j \geq 0$. The second inequality holds because $\|g_k\|_{L_2} \leq \|g'_k\|_{L_2}$. By squaring (5.5) and plugging in the value of $\|g'_j\|^2_{L_2}$ we get (5.3).

Proof of (5.4). By our construction, it must hold that

$$c^q \alpha_{i_k}/2^{i'} \leq \alpha_{i_j}, \tag{5.6}$$

since otherwise the index i_j would have been removed from the $\mathcal{I}$ at the kth step. Since $\beta_{i_k} = \beta_{i_j} c^{i'}$, we can plug (5.2) in (5.6) to get

$$\mu_{i_k} c^{q+qi'}/2^{i'} \leq c^q \mu_{i_j}$$

or

$$\mu_{i_k} \leq \mu_{i_j} \cdot 2^{i'} \cdot c^{-qi'}.$$

Since $|T_i| = e|U_i|$ for all i, it follows that $|T_k|/|T| \leq 2^{i'} \cdot c^{-qi'}$. On the other hand, we know that $\mathbf{E}_{x\in T_k}[g'_k(x)^2] = \beta^2_{i_k}/4 = c^{2i'}\beta^2_{i_j}/4$. Thus,

$$\mathbf{E}_{x\in T}[g'_k(x)^2] \leq 2^{i'} c^{2i'-qi'} \beta^2_{i_j}/4 = (2/c^{q-2})^{i'} \beta^2_{i_j}/4,$$

and we now just choose c sufficiently large so that $2/c^{q-2} > 100$. □

□

□

5.4 Hardness of $\min_{p^*\to p}(\cdot)$

In this section, we prove Theorem 2.2.2 that for any $\varepsilon > 0$ and $p \in (2, \infty)$, it is NP-hard to approximate $\min_{p^*\to p}(\cdot)$ within a factor $(\gamma_p - \varepsilon)$, where

$$\gamma_p = (\mathbf{E}_{g\sim\mathcal{N}(0,1)}[|g|^p])^{1/p} > 1$$

is the absolute pth moment of the standard Gaussian.

Our result is obtained by using the result of Guruswami et al. [36] that proved the same hardness of $\min_{2\to p}(\cdot)$. When $\|\cdot\|_{L_p}$ denotes expectation p-norm defined by $\|x\|_{L_p} := \mathbf{E}_i[|x_i|^p]^{1/p}$, since $p^* < 2$, any x satisfies $\|x\|_{L_{p^*}} \le \|x\|_{L_2}$. This implies that for any matrix A, the optimal value of $\min_{p^*\to p}(A)$ is at least the optimal value of $\min_{2\to p}(A)$. We modify the reduction of [36] slightly such that in the Yes case, x that minimizes $\min_{2\to p}(A)$ has either $+1$ or -1 in each coordinate. This implies $\|x\|_{L_2} = \|x\|_{L_{p^*}}$, and certifies that $\min_{p^*\to p}(A) = \min_{2\to p}(A)$. In the No case, $\min_{p^*\to p}(A)$ is always at least $\min_{2\to p}(A)$, so the gap between the Yes case and the No case for $\min_{p^*\to p}(\cdot)$ is at least as large as the gap for $\min_{2\to p}(\cdot)$.

Our presentation closely follows the recent work by Bhattiprolu et al. [9].

Fourier Analysis. To present the reduction, we first introduce some basic facts about Fourier analysis.

Let $R \in \mathbb{N}$ be a positive integer, and consider a function $f : \{\pm 1\}^R \to \mathbb{R}$. For any subset $S \subseteq [R]$ let $\chi_S := \prod_{i\in S} x_i$. Then we can represent f as

$$f(x_1, \ldots, x_R) = \sum_{S\subseteq[R]} \widehat{f}(S) \cdot \chi_S(x_1, \ldots x_R), \tag{5.7}$$

where

$$\widehat{f}(S) = \mathbf{E}_{x\in\{\pm1\}^R}[f(x) \cdot \chi_S(x)] \text{ for all } S \subseteq [R]. \tag{5.8}$$

The *Fourier transform* refers to a linear operator F that maps f to $\widehat{f}$ as defined as (5.8). We interpret $\widehat{f}$ as a 2^R-dimensional vector whose coordinates are indexed by $S \subseteq [R]$. In this section, we let $\|\cdot\|_{\ell_p}$ to denote the counting p-norm and $\|\cdot\|_{L_p}$ to denote the expectation p-norm. Endow the expectation norm and the expectation norm to f and $\widehat{f}$ respectively; i.e.,

$$\|f\|_{L_p} := \left(\mathbf{E}_{x\in\{\pm1\}^R} |f(x)|^p\right)^{1/p} \quad \text{and} \quad \|\widehat{f}\|_{\ell_p} := \left(\sum_{S\subseteq[R]} |\widehat{f}(S)|^p\right)^{1/p}.$$

as well as the corresponding inner products $\langle f, g\rangle$ and $\langle \widehat{f}, \widehat{g}\rangle$ consistent with their 2-norms. We also define the *inverse Fourier transform* F^T to be a linear operator that maps a given $\widehat{f} : 2^R \to \mathbb{R}$ to $f : \{\pm1\}^R \to \mathbb{R}$ defined as in (5.7). We state the following well-known facts from Fourier analysis.

Observation 5.4.1 (Parseval's Theorem)**.** For any $f : \{\pm1\}^R \to \mathbb{R}$, $\|f\|_{L_2} = \|Ff\|_{\ell_2}$.

Observation 5.4.2. F and F^T form an adjoint pair; i.e., for any $f : \{\pm 1\}^R \to \mathbb{R}$ and $\widehat{g} : 2^R \to \mathbb{R}$,

$$\langle \widehat{g}, Ff \rangle = \langle F^T \widehat{g}, f \rangle.$$

Observation 5.4.3. $F^T F$ is the identity operator.

Smooth Label Cover. An instance of Label Cover is given by a quadruple $\mathcal{L} = (G, [R], [L], \Sigma)$ that consists of a regular connected graph $G = (V, E)$, a label set $[R]$ for some positive integer n, and a collection $\Sigma = ((\pi_{e,v}, \pi_{e,w}) : e = (v, w) \in E)$ of pairs of maps both from $[R]$ to $[L]$ associated with the endpoints of the edges in E. Given a *labeling* $\ell : V \to [R]$, we say that an edge $e = (v, w) \in E$ is *satisfied* if $\pi_{e,v}(\ell(v)) = \pi_{e,w}(\ell(w))$. Let $\mathrm{OPT}(\mathcal{L})$ be the maximum fraction of satisfied edges by any labeling.

The following hardness result for Label Cover, given in [36], is a slight variant of the original construction due to [51]. The theorem also describes the various structural properties, including smoothness, that are identified by the hard instances.

Theorem 5.4.4. *For any $\xi > 0$ and $J \in \mathbb{N}$, there exist positive integers $R = R(\xi, J), L = L(\xi, J)$ and $D = D(\xi)$, and a Label Cover instance $(G, [R], [L], \Sigma)$ as above such that*

- *(Hardness): It is NP-hard to distinguish between the following two cases:*
  - *Yes case:* $\mathrm{OPT}(\mathcal{L}) = 1$.
  - *No case:* $\mathrm{OPT}(\mathcal{L}) \leq \xi$.
- *(Structural Properties):*
  - *(J-Smoothness): For every vertex $v \in V$ and distinct $i, j \in [R]$, we have*

$$\mathbf{Pr}_{e \ni v}\Big[\pi_{e,v}(i) = \pi_{e,v}(j)\Big] \leq 1/J.$$

  - *(D-to-1): For every vertex $v \in V$, edge $e \in E$ incident on v, and $i \in [L]$, we have $|\pi_{e,v}^{-1}(i)| \leq D$; that is at most D elements in $[R]$ are mapped to the same element in $[L]$.*
  - *(Weak Expansion): For any $\delta > 0$ and vertex set $V' \subseteq V$ such that $|V'| = \delta \cdot |V|$, the number of edges among the vertices in $|V'|$ is at least $(\delta^2/2)|E|$.*

Reduction. Let $\mathcal{L} = (G, [R], [L], \Sigma)$ be an instance of Label Cover with $G = (V, E)$. Our reduction will construct a linear operator $\mathbf{A} : \mathbb{R}^N \to \mathbb{R}^M$ with $N = |V| \cdot 2^R$ and $M = 2|V| \cdot 2^R - |V| + |E| \cdot |L|$. The space $\mathbb{R}^N$ will be endowed the expectation norm (and call its elements functions) and $\mathbb{R}^M$ will be endowed the counting norm (and call its elements vectors). We define $\mathbf{A}$ by giving a linear transformation from a function $\mathbf{f} : V \times \{\pm 1\}^R \to \mathbb{R}$ to a vector $\mathbf{a} \in \mathbb{R}^M$. Let $C := M^3$. Given $\mathbf{f}$, a vertex $v \in V$ induces $f_v \in \mathbb{R}^{2^R}$ defined by $f_v(x) := \mathbf{f}(v, x)$ for $x \in \{\pm 1\}^R$. Let $\widehat{\mathbf{g}} \in V \times [R]$ be the vectors of linear coefficients; $\widehat{\mathbf{g}}(v, i) = \widehat{f_v}(i)$ for $v \in V, i \in [R]$. Given $\mathbf{f}$ (that determines $\{\widehat{f_v}\}_v \in V$ and $\widehat{\mathbf{g}}$), $\mathbf{a} = \mathbf{A}\mathbf{f}$ is defined as follows.

- For $v \in V$ and $x \in \{\pm 1\}^R$, $\mathbf{a}(v, x) = \sum_{i=1}^R \widehat{\mathbf{g}}(v, i) x_i$.
- For $v \in V$ and $S \subseteq [R]$ with $|S| \neq 1$, $\mathbf{a}(v, S) = C \cdot \widehat{f_v}(S)$.
- For $e = (u, v) \in E$ and $i \in [L]$, $\mathbf{a}(e, i) = C \cdot \left(\sum_{j \in \pi_{e,u}^{-1}(i)} \widehat{f_u}(i) - \sum_{j \in \pi_{e,v}^{-1}(i)} \widehat{f_v}(i) \right)$.

Since $\widehat{\mathbf{g}}$ and $\mathbf{a}$ are all linear in $\mathbf{f}$, the matrix $\mathbf{A}$ that satisfies $\mathbf{a} = \mathbf{A}\mathbf{f}$ is well-defined, which is our instance of $\min_{p^* \to p}(\cdot)$. Intuitively, C will be chosen large enough so that every $\widehat{f_v}$ has almost all Fourier mass on its linear coefficients, and their linear coefficients correctly indicate the labels that satisfy all constraints of the Label Cover instance.

Completeness. We prove the following lemma for the Yes case.

Lemma 5.4.1 (Completeness)**.** *Let $\ell : V \to [R]$ be a labeling that satisfies every edge of $\mathcal{L}$. There exists a function $\mathbf{f} \in \mathbb{R}^{V \times 2^R}$ such that $\mathbf{f}(v, x)$ is either $+1$ or -1 for all $v \in V, x \in \{\pm 1\}^R$ and $\|\mathbf{A}\mathbf{f}\|_{\ell_p} = (|V| \cdot 2^R)^{1/p}$. In particular, $\|\mathbf{A}\mathbf{f}\|_{\ell_p} / \|\mathbf{f}\|_{L_{p^*}} = (|V| \cdot 2^R)^{1/p}$.*

Proof. Let $\mathbf{f}(v, x) := x_{\ell(v)}$ for every $v \in V, x \in \{\pm 1\}^R$. Consider $\mathbf{a} = \mathbf{A}\mathbf{f}$. Since every $\widehat{f_v}$ is linear, for each $v \in V$ and $S \subseteq [R]$ with $|S| \neq 1$, $\mathbf{a}(v, S) = 0$. For each $v \in V$ and $i \in [R]$, $\widehat{\mathbf{g}}(v, i) = 1$ if and only if $i = \ell(v)$ and 0 otherwise. Since ℓ satisfies every edge of $\mathcal{L}$, $\mathbf{a}(e, i) = 0$ for every $e \in E$ and $i \in [L]$. This implies that for every $v \in V, x \in \{\pm 1\}^R$, $\mathbf{a}(v, x) = x_{\ell(v)} = \mathbf{f}(v, x)$. Therefore, $\|\mathbf{A}\mathbf{f}\|_{\ell_p} = (|V| \cdot 2^R)^{1/p}$. □

Soundness. We prove the following lemma for the soundness. Combined with Theorem 5.4.4 for hardness of Label Cover and observing that $\|\mathbf{f}\|_{L_{p^*}} \leq \|\mathbf{f}\|_{L_2}$, it finishes the proof of Theorem 2.2.2.

Lemma 5.4.2. *For any $\eta > 0$, there exists $\xi > 0$ (that determines $D = D(\xi)$ as in Theorem 5.4.4) and $J \in \mathbb{N}$ such that if $\mathrm{OPT}(\mathcal{L}) \le \xi$, $\mathcal{L}$ is D-to-1 and $\mathcal{L}$ is J-smooth, for every $\mathbf{f}$ with $\|\mathbf{f}\|_{L_2} = 1$, $\|\mathbf{Af}\|_{\ell_p} \ge (\gamma_p - \eta)(|V| \cdot 2^R)^{1/p}$.*

Proof. We will prove contrapositive; if $\|\mathbf{Af}\|_{\ell_p} \le (\gamma_p - \eta)(|V| \cdot 2^R)^{1/p}$ for some $\mathbf{f}$ is small then $\mathrm{OPT}(\mathcal{L}) \ge \xi$ with the choice of the parameters that will determined later. Fix such an $\mathbf{f}$ with $\|\mathbf{f}\|_{L_2} = 1$ that determines f_v and $\widehat{f_v}$ for each $v \in V$. Let $\mathbf{a} = \mathbf{Af}$. Suppose that there is $v \in V$ and $S \subseteq [R]$ with $|S| \neq 1$ such that $|\widehat{f_v}(S)| > 1/M^2$. It means that $|\mathbf{a}(v, S)| > C/M^2$. Since $C = M^3$, it already implies $\|\mathbf{a}\|_{\ell_p} \ge M \gg (\gamma_p - \eta)(|V| \cdot 2^R)^{1/p}$, so suppose that there is no such v and S.

Let $\widehat{\mathbf{g}} \in V \times [R]$ be defined as above; $\widehat{\mathbf{g}}(v, i) = \widehat{f_v}(i)$ for $v \in V, i \in [R]$. By Parseval,

$$\sum_{v \in V} \|\widehat{f_v}\|_{\ell_2}^2 = \sum_{v \in V} \|f_v\|_{L_2}^2 = |V| \cdot \mathbf{E}_{v \in V} \|f_v\|_{L_2}^2 = |V| \cdot \|\mathbf{f}\|_{L_2}^2 = |V|.$$

and the fact that $|\widehat{f_v}(S)| < 1/M^2$ for every $v \in V$, $S \subseteq [R]$ with $|S| \neq 1$, we have $\|\widehat{\mathbf{g}}\|_{\ell_2} \in [\sqrt{|V|} - 1/M, \sqrt{|V|}]$.

Furthermore, suppose that there is $e = (u, v) \in E$ and $i \in [L]$ such that

$$\left| \sum_{j \in \pi_{e,u}^{-1}(i)} \widehat{\mathbf{g}}(u, i) - \sum_{j \in \pi_{e,v}^{-1}(i)} \widehat{\mathbf{g}}(v, j) \right| \ge 1/M^2.$$

This implies that $|\mathbf{a}(e, i)| \ge C/M^2$. Since $C = M^3$, it already implies $\|\mathbf{a}\|_{\ell_p} \ge M \gg (\gamma_p - \eta)(|V| \cdot 2^R)^{1/p}$, so we can assume that there is no such e and i.

To bound $\|\mathbf{a}\|_{\ell_p}$, it only remains to analyze

$$\sum_{v \in V} \sum_{x \in \{\pm 1\}^R} \left| \mathbf{a}(v, x) \right|^p = \sum_{v \in V} \sum_{x \in \{\pm 1\}^R} \left| \sum_{i \in [R]} \widehat{\mathbf{g}}(v, i) x_i \right|^p. \tag{5.9}$$

The rest of the proof closely follows [36], and we explain high-level intuitions and why their proofs work in our settings. First, let us assume that $\|\widehat{\mathbf{g}}\|_{\ell_2} = \sqrt{|V|}$. It involves a multiplicative error of $(1 - 1/M)$, which is negligible in our proof. To simplify notations, let $\widehat{g}_v \in \mathbb{R}^R$ be such that $\widehat{g}_v(i) := \widehat{f_v}(\{i\}) = \widehat{\mathbf{g}}(v, i)$ for each $v \in V$ and $i \in [R]$. Call a vertex $v \in V$ *τ-irregular* if there exists $i \in [R]$ such that $|\widehat{\mathbf{g}}(v, i)| > \tau \|\widehat{g}_v\|_{\ell_2}^2$. If not, v is *τ-regular*. Also, call a vertex $v \in V$ *small* if $\|\widehat{g}_v\|_{\ell_2} < 1/M$. Otherwise, call it *big*.

For each $v \in V$, we consider $\sum_{x \in \{\pm 1\}^R} \left| \sum_{i \in [R]} \widehat{g}(v, i) x_i \right|^p$. By Khintchine inequality, it is at most $2^R \cdot \gamma_p^p \cdot \|\widehat{g}\|_{\ell_2}^p$. The following lemma, based on standard applications of the Berry-Esseen theorem, shows that the converse is almost true when v is τ-regular, implying the contribution from irregular vertices to (5.9) is large.

Lemma 5.4.3 ([54]). *For sufficiently small τ (depending only on p), if $v \in V$ is τ-regular, then*

$$\sum_{x\in\{\pm 1\}^R} \left| \sum_{i\in[R]} \widehat{g}(v,i)x_i \right|^p \geq 2^R \cdot \gamma_p^p \cdot \|\widehat{g}\|_{\ell_2}^p (1-\sqrt{\tau}).$$

Let S be the set of big τ-irregular vertices. Based on the above, the following lemma shows that S must be a large set. Originally, [36] only argued for τ-irregular vertices. (The notion of big and small vertices does not appear there.) However, since the contribution of small vertices to (5.9) is negligible, the same proof essentially works.

Lemma 5.4.4 (Lemma 4.4 of [36]). *There are τ and θ, depending only on p and η, such that S, the set of big τ-irregular vertices, satisfies $|S| \geq \theta|V|$.*

By the weak expansion property of $\mathcal{L}$ guaranteed in Theorem 5.4.4, S induces at least $\theta^2|E|$ edges of $\mathcal{L}$. To finish the proof, [36] showed that we can satisfy a significant fraction of the edges from $\mathcal{L}$. The only difference in their setting and our setting is that

- [36]: S is the set of all τ-irregular veritces. For each $e = (u, v)$ and $i \in [L]$,

$$\sum_{j\in\pi_{e,u}^{-1}(i)} \widehat{g}_u(j) = \sum_{j\in\pi_{e,v}^{-1}(i)} \widehat{g}_v(j). \tag{5.10}$$

- Here: S is the set of all big τ-irregular veritces. For each $e = (u, v)$ and $i \in [L]$,

$$\left| \sum_{j\in\pi_{e,u}^{-1}(i)} \widehat{g}_u(j) - \sum_{j\in\pi_{e,v}^{-1}(i)} \widehat{g}_v(j) \right| < 1/M^2. \tag{5.11}$$

These differences do not affect their proof since in the only place (5.10) was used for $e = (u, v)$ and $i \in [L]$, they indeed used the fact the left-hand side of (5.11) is at most $0.3\tau \cdot \max(\|\widehat{g}_u\|_{\ell_2}, \|\widehat{g}_v\|_{\ell_2})$. Since we additionally assumed that S is big, $\|\widehat{g}_u\|_{\ell_2} \geq 1/M$ for each $u \in S$, so it is always satisfied from (5.11).

Lemma 5.4.5 ([36]). *Let $\beta := 10000D^4/\tau^4 J$. Then $\mathrm{OPT}(\mathcal{L}) \geq (\tau^4/16)(\theta^2 - 2/\beta)$.*

Since θ and τ only depend on η and p, fixing small enough ξ (that determines D) and large enough J will ensure $\mathrm{OPT}(\mathcal{L}) \geq (\tau^4/16)(\theta^2 - 2/\beta) \geq \xi$, finishing the proof of the lemma. □

5.5 Hardness for Finite Fields

In this section, we prove Lemma 2.2.3, which in turn finishes the proof of Theorem 5.0.3 for hardness of ℓ_0-row lank approximation for matrices whose entries are from a finite field $\mathbb{F}$.

Lemma (Restatement of Lemma 2.2.3)**.** *Let $\mathbb{F}$ be a finite field and $A \in \mathbb{F}^{n\times d}$ with $n \geq d$ and $k = d - 1$. Then*

$$\min_{U\in\mathbb{F}^{n\times k},V\in\mathbb{F}^{k\times d}} \|UV - A\|_0 = \min_{x\in\mathbb{F}^d,x\neq 0} \|Ax\|_0.$$

Proof. Assume that the rank of A is d; otherwise the lemma becomes trivial. We first prove ($\geq$). Given $V^* \in \mathbb{F}^{k\times d}$ that achieves the best rank k approximation, assume without loss of generality that the rank of V^* is $k = d - 1$. Let $x \in \mathbb{F}^d$ be a nonzero vector that is orthogonal to the rowspace of V; i.e., $\langle v, x\rangle = 0$ if and only if $v \in$ rowspace(V). Note that unlike in $\mathbb{R}$, x can be in rowspace(V), but it does not affect the proof. Let $a_1, \ldots, a_n$ be the rows of A. For fixed V^* and $i \in [n]$, if $a_i \in$ rowspace(V), then we can compute the ith row of U^* (denoted by u_i^*) such that $u_i^*V^* = a_i$. Otherwise, $\langle a_i, x\rangle = b$ for some $b \neq 0$, since x is nonzero, there is u_i^* such that $\|u_i^*V^* = a_i\|_0 = 1$. Therefore, $\|U^*V^* - A\|_0 = \|Ax\|_0$, which implies that $\min_{U\in\mathbb{F}^{n\times k},V\in\mathbb{F}^{k\times d}}\|UV - A\|_0 \geq \min_{x\in\mathbb{F}^d,x\neq 0}\|Ax\|_0$.

For the other direction, given $x \in \mathbb{F}^d \setminus \{0\}$, the set of vectors u with $\langle u, x\rangle = 0$ forms a k-dimensional subspace. (Again, this space may contain x unlike in $\mathbb{R}$, but it does not matter.) Let $V^* \in \mathbb{R}^{k\times d}$ be a matrix whose rows span that space, and compute U^* as above. The above analysis shows that $\|U^*V^* - A\|_0 = \|Ax\|_0$, which completes the proof. □

Chapter 6

Additional Results

Here we list some additional results on variants of the ℓ_p low rank approximation problem.

6.1 Bicriteria Algorithm

In this section we show that we can develop low rank approximations that apply to matrices whose entries are not bounded by poly(n) so long as we accept bicriteria algorithms. That is, instead of a target rank k approximation, the algorithm will output an approximating matrix of rank $3k$.

Theorem 6.1.1. *If A is an $n \times d$ matrix, our target rank k is a constant, and $1 \leq p < 2$, then there exists a polynomial time algorithm that outputs a matrix M of rank at most $3k$ such that $\|M - A\|_p \leq (1+\varepsilon)OPT$ where OPT is the best rank k ℓ_p-low rank approximation value for A with probability $1 - O(1)$.*

Proof. Let C_l denote the best rank l approximation to a matrix C in the ℓ_p norm (i.e. the matrix that minimizes $\|C_l - C\|_p$).

Let B be the best rank k approximation to A in the Frobenius norm. Then

$$\|A - B\|_p \leq \text{poly}(n)\|A - B\|_F \leq \text{poly}(n)\|A - A_k\|_F \leq \text{poly}(n)OPT.$$

We can find a rank $2k$ $(1+\varepsilon)$-approximation to $A - B$ using the same techniques as in Theorem 4.1.1, where we sample a matrix S of p-stable variables, guess values for SU^*, and then minimize $\|SU^*V^* - S(A-B)\|_p$. Now the entries of $A-B$ are not necessarily bounded by poly(n) so we need to justify that it suffices to guess poly(n) values for SU^*.

Indeed, by a well-conditioned basis argument, no entry of U^* has absolute value greater than $\text{poly}(n)\|A-B\|_p$. Furthermore, we can round each entry of U^* (similar to the proof of Theorem 4.1.1) to the nearest multiple of $\frac{\varepsilon\|A-B\|_p}{\text{poly}(n)}$ and incur an additive error of at most εOPT because $\|A-B\|_p \leq \text{poly}(n)OPT$. This error is small enough for the purposes of our approximation.

Let $(A-B)^*_{2k} = U^*V^*$ and let $M = (A-B)^*_{2k} + B$. We have

$$\begin{aligned}\|M-A\|_p &= \|(A-B)^*_{2k} + B - A\|_p \\ &= \|(A-B)^*_{2k} - (A-B)\|_p \\ &\leq (1+\varepsilon)\|(A-B)_{2k} - (A-B)\|_p + \varepsilon OPT \\ &\leq (1+\varepsilon)\|A_k - B - (A-B)\|_p + \varepsilon OPT \\ &\leq (1+\varepsilon)OPT\end{aligned}$$

where the first inequality follows from our argument above and the second inequality follows because $A_k - B$ has rank at most $k+k = 2k$.

Since M has rank at most $2k+k = 3k$, then the result follows. □

6.2 Weighted Low Rank Approximation

For $0 < p < 2$, we can also design a PTAS for the weighted ℓ_p low rank approximation problem. In this setting we have a matrix A, a weight matrix W of rank r, and we want to output a rank k matrix A' such that, for $\varepsilon > 0$,

$$\|W \circ (A - A')\|_p^p \leq (1+\varepsilon) \min_{\text{rank k } A_k} \|W \circ (A - A_k)\|_p^p.$$

Our main tool will be a multiple regression concentration result based on that of [81].

Theorem 6.2.1. *Let S be a poly$(k/\varepsilon) \times n$ matrix whose entries are i.i.d p-stable random variables with scale 1. Let $M^{(1)}, M^{(2)}, \ldots, M^{(m)}$ be $n \times d$ matrices and let $b^{(1)}, b^{(2)}, \ldots, b^{(m)} \in \mathbb{R}^n$. Let*

$$x^{(i)} = \arg\min_x \|M^{(i)}x - b^{(i)}\|_p^p$$

and

$$y^{(i)} = \arg\min_y med(SM^{(i)}y - Sb^{(i)})/med_p.$$

Then w.h.p we have

$$\sum_i \|M^{(i)}y^{(i)} - b^{(i)}\|_p^p \leq (1+O(\varepsilon)) \sum_i \|M^{(i)}x^{(i)} - b^{(i)}\|_p^p$$

Proof. By Lemmas 4.2.5 and 4.3.4, w.h.p.

$$\sum_i \frac{\text{med}(S(M^{(i)}x^{(i)} - b^{(i)}))^p}{\text{med}_p^p} \leq (1 + O(\varepsilon)) \sum_i \|M^{(i)}x^{(i)} - b^{(i)}\|_p^p.$$

Let T be the set of all i such that

$$\frac{\text{med}(S[M^{(i)}\ b^{(i)}]y)^p}{\text{med}_p^p} \geq (1 - \Theta(\varepsilon))\|[M^{(i)}\ b^{(i)}]y\|_p^p$$

for all y. By Corollary 4.2.4, we know that for each i, the probability that $i \in T$ is at least $1 - \Theta(\varepsilon)$.

Thus

$$\mathbb{E}[\sum_{i \notin T} \|[M^{(i)}\ b^{(i)}]y\|_p^p] \leq \Theta(\varepsilon) \sum_i \|[M^{(i)}\ b^{(i)}]y\|_p^p$$

so by Markov's inequality, w.h.p we have

$$\sum_{i \notin T} \|[M^{(i)}\ b^{(i)}]y\|_p^p \leq \Theta(\varepsilon) \sum_i \|[M^{(i)}\ b^{(i)}]y\|_p^p.$$

Let y be arbitrary. Since

$$\sum_i \frac{\text{med}(S[M^{(i)}\ b^{(i)}]y)^p}{\text{med}_p^p} \geq \sum_{i \notin T} \frac{\text{med}(S[M^{(i)}\ b^{(i)}]y)^p}{\text{med}_p^p} \geq (1 - \Theta(\varepsilon)) \sum_{i \notin T} \|[M^{(i)}\ b^{(i)}]y\|_p^p,$$

it follows that for all y we have

$$\sum_i \frac{\text{med}(S[M^{(i)}\ b^{(i)}]y)^p}{\text{med}_p^p} \geq (1 - \Theta(\varepsilon)) \sum_i \|[M^{(i)}\ b^{(i)}]y\|_p^p.$$

Therefore w.h.p we have

$$(1 - \Theta(\varepsilon)) \sum_i \|M^{(i)}y^{(i)} - b^{(i)}\|_p^p \leq \sum_i \frac{S(M^{(i)}y^{(i)} - b^{(i)})}{\text{med}_p^p}$$

$$\leq \sum_i \frac{S(M^{(i)}x^{(i)} - b^{(i)})}{\text{med}_p^p} \leq (1 + O(\varepsilon)) \sum_i \|M^{(i)}x^{(i)} - b^{(i)}\|_p^p$$

because $0 < p < 2$. The result follows.

□

Theorem 6.2.2. *Suppose A and W are $n \times d$ matrices with entries bounded by poly(n), and $r = \text{rank}(W)$. There is an algorithm that for any integer k, $p \in (0, 2)$ and $\varepsilon \in (0, 1)$, outputs in time $n^{r \cdot \text{poly}(k/\varepsilon)}$ a $n \times k$ matrix U^* and a $k \times d$ matrix V^* such that*

$$\|W \circ (A - U^*V^*)\|_p^p \leq (1 + O(\varepsilon)) \min_{\text{rank-}k\ A_k} \|W \circ (A - A_k)\|_p^p.$$

Proof. To achieve a relative-error low rank approximation $W \circ (UV - A)$, for each column i we can guess sketches for $W_{:,i} \circ UV_i$ using a similar argument as in Theorem 4.1.1. Indeed, we can apply Theorem 6.2.1 with $M^{(i)} = W_{:,i} \circ U^*V_i^*$ and $b^{(i)} = W_{:,i} \circ A_{:,i}$. To do so, we need to be able to guess $SW_{:,i} \circ U^*$, a poly$(\frac{k}{\varepsilon}) \times k$ matrix, in poly(n) tries. We will follow the same reasoning as in the proof of Theorem 4.1.1. Since the entries of W and A are bounded by poly(n), then we can bound the entries of U^* by poly(n) using a well-conditioned basis. Furthermore, we can round each entry of U^* to the nearest multiple of poly(n^{-1}) while incurring an error factor of only $(1 + O(\varepsilon))$. Thus, we need only $n^{\text{poly}(k/\varepsilon)}$ guesses.

Of course, there d columns so this is not enough to achieve a PTAS. However, we only need to guess sketches for r values of j because W has rank r so we can express any column of W as a linear combination of those r columns. That is, we choose a subset S of the columns such that $|S| = r$ and guess the sketches of $W_{:,i} \circ UV_i$ for each $i \in S$ as described in the previous paragraph. Therefore, we require $n^{r \cdot \text{poly}(k/\varepsilon)}$ time in total for a $(1 + O(\varepsilon))$ approximation algorithm. Since k and r are constants this results in a PTAS. □

Part II

Population Recovery

Chapter 7

Introduction

Our results

Positive result. As our main positive result, we obtain an algorithm which learns any unknown distribution $\mathbf{X}$ supported on at most ℓ strings under the deletion channel. For any constant ℓ (and in fact even for ℓ as large as $o(\log n/\log\log n)$, its sample complexity is exponential in $n^{1/2+o(1)}$. In more detail, our main positive result is the following:

Theorem 7.0.1 (Learning an arbitrary mixture of ℓ strings under the deletion channel)**.** *There is an algorithm with the following performance guarantee: Let $\mathbf{X}$ be an arbitrary distribution over at most ℓ strings in $\{0,1\}^n$. For any deletion rate $0 < \delta < 1$ and any accuracy parameter ε, if the algorithm is given access to independent draws from $\mathbf{X}$ that are independently corrupted with deletion noise at rate δ, then the algorithm uses*

$$\frac{1}{\varepsilon^2}\cdot\left(\frac{2}{1-\delta}\right)^{\sqrt{n}\cdot(\log n)^{O(\ell)}}$$

many samples and with probability at least 0.99 outputs a hypothesis $\tilde{\mathbf{X}}$ which is supported over at most ℓ strings and has total variation distance at most ε from the unknown target distribution $\mathbf{X}$.

It is easy to see that if the target distribution is promised to be uniform over (a multi-set of) at most ℓ strings, then the algorithm of Theorem 7.0.1 can be used to exactly reconstruct the unknown multi-set.

As we explain in Section 7.1, while Theorem 7.0.1 extends prior results on trace reconstruction (the $\ell = 1$ case), it is proved using very different techniques from recent works [27, 38, 41, 42, 71, 76] on trace reconstruction.

We note that for deletion rates δ that are bounded away from 1 by a constant, the $2^{O(n^{1/3})}$ sample complexity bounds of [27, 71] for trace reconstruction are better than the $\ell = 1$ case of our result. However, our bounds apply even if the deletion rate δ is very close to 1; in particular, [27, 71] give no results for very high deletion rates $\delta = 1 - o(1/\sqrt{n})$, while Theorem 7.0.1 gives a $2^{\tilde{O}(\sqrt{n})}$ bound for $\delta = 1 - 1/2^{\mathrm{polylog}(n)}$ and a $2^{o(n)}$ bound even for δ as large as $1 - 1/2^{\sqrt{n}/\mathrm{polylog}(n)}$. Of course, the main feature of Theorem 7.0.1 is that it applies when $\ell > 1$ (unlike [27, 71]).

Negative result. Complementing the sample complexity upper bound, we obtain a lower bound on the sample complexity of population recovery. Our lower bound shows that for a wide range of values of ℓ, at least $n^{\Omega(\ell)}$ samples are required when the population is of size at most ℓ. An informal version of our lower bound is as follows (see Theorem 10.0.1 in Chapter 10 for a detailed statement):

Theorem 7.0.2 (Sample complexity lower bound, informal statement)**.** *Let $0 < \delta < 1$ be any constant deletion probability and suppose that A is an algorithm which, when run on samples drawn from the δ-deletion channel over an arbitrary distribution $\mathbf{X}$ supported over at most $\ell \leq n^{0.499}$ many strings, with probability at least 0.51 outputs a hypothesis distribution $\tilde{\mathbf{X}}$ that has total variation distance at most 0.49 from the unknown target distribution $\mathbf{X}$. Then A must use $n^{\Omega(\ell)}$ many samples.*

7.1 Our techniques

As noted earlier, our positive result (Theorem 7.0.1) gives a sample complexity upper bound for the original ($\ell = 1$) trace reconstruction problem as a special case. We remark that both of the recent $2^{O(n^{1/3})}$ sample complexity upper bounds for the trace reconstruction problem [27, 71], as well as the earlier work of [42], employed essentially the same algorithmic approach, which is referred to in [27] as a "mean-based algorithm." At a high level, mean-based algorithms use their samples (traces) only to compute empirical estimates of the n expectations[1]

$$\mathbf{E}_{\boldsymbol{z} \leftarrow \mathrm{Del}_\delta(x)}[\boldsymbol{z}_0],\ \cdots, \mathbf{E}_{\boldsymbol{z} \leftarrow \mathrm{Del}_\delta(x)}[\boldsymbol{z}_{n-1}] \tag{7.1}$$

corresponding to the coordinate means of the received traces; they then only use those n estimates to reconstruct the unknown target string x. Both of the algorithms in [27, 71], as well as the algorithm from [42] for trace reconstruction from an arbitrary string

[1] In this context, the original unknown target string x is viewed as belonging to $\{-1, 1\}^n$, and a trace z obtained from $\mathrm{Del}_\delta(x)$ is viewed as a string in $\{-1, 1\}^{n'}$ for some $n' \leq n$ with $n - n'$ zeros appended to the end. We use $[0 : n-1] = \{0, \ldots, n-1\}$ to index entries of a string of length n.

x, are mean-based algorithms. (Both [27] and [71] show that their sample complexity upper bounds are essentially best possible for any mean-based trace reconstruction algorithm.)

While mean-based algorithms have led to state-of-the-art results for trace reconstruction of a single string, this approach breaks down even for the simplest non-trivial cases of population recovery under the deletion channel. Indeed, even when $\ell = 2$ and the unknown distribution $\mathbf{X}$ is promised to be uniform over two strings, it is easy to see that the coordinate means do not provide enough information to recover $\mathbf{X}$. For example, if (x^1, x^2) and (y^1, y^2) are two pairs of strings whose sums (as vectors in $\mathbb{R}^n$) $x^1 + x^2$ and $y^1 + y^2$ are equal (such as $x^1 = 0^n$, $x^2 = 1^n$, $y^1 = 0^{n/2}1^{n/2}$, $y^2 = 1^{n/2}0^{n/2}$), it is easy to see that the coordinate means of received traces will match perfectly:

$$\mathop{\mathbf{E}}_{j\in\{1,2\}} \mathop{\mathbf{E}}_{z\leftarrow \mathrm{Del}_\delta(x^j)}[z_i] = \mathop{\mathbf{E}}_{j\in\{1,2\}} \mathop{\mathbf{E}}_{z\leftarrow \mathrm{Del}_\delta(y^j)}[z_i], \quad \text{for every } i \in \{0,\dots,n-1\}.$$

Thus the mean-based approach of [27, 42, 71] does not suffice for even the simplest version of the population recovery problem when $\ell = 2$. Indeed, our sample complexity upper bounds are obtained using a completely different approach, which we explain below.

Warm-up: A different approach to trace reconstruction (the $\ell = 1$ case)

As a warm-up to our main results, we first give a high-level explanation of how our approach can be used to obtain a simple $2^{\tilde{O}(\sqrt{n})}$-sample algorithm for the trace reconstruction problem. While this is a higher sample complexity than the state-of-the-art mean-based approach of [27, 71] (though our approach does better for very high deletion rates, as noted earlier), our approach has the crucial advantage that it can be adapted to go beyond the $\ell = 1$ case, whereas the mean-based approach cannot handle $\ell > 1$ as described above.

In a nutshell, the essence of our approach is to work with *subsequence frequencies* in the *original string* x (in contrast, note that the mean-based approach uses *single-coordinate frequencies* in the *received traces*). To explain further we introduce some useful terminology: the *k-deck* of a string $x \in \{0,1\}^n$, denoted $\mathsf{D}_k(x)$, is the multi-set of all $\binom{n}{k}$ subsequences of x with length exactly k. Thus, the k-deck encapsulates all frequency information about length-k subsequences of x.

A question that arises naturally in the combinatorics of words is the following: what is the smallest value of k (as a function of n) so that for every string $x \in \{0,1\}^n$, the k-deck of x uniquely identifies x? Despite significant investigation dating back

to the 1970s [45], this basic quantity is still poorly understood. Improving on earlier $k \leq n/2$ bounds of Kalashnik [45] and Manvel et al. [62] and a simultaneous $k = O(\sqrt{n \log n})$ bound of Scott [83], Krasikov and Roddity [55] showed that $k = O(\sqrt{n})$ suffices. On the lower bounds side, the best lower bound known is $k = 2^{\Omega(\sqrt{\log n})}$, due to Dudík and Schulman [30] (improving on earlier $k = \Omega(\log n)$ lower bounds of [62] and [19]).

The relevance of upper bounds on k to the trace reconstruction problem is intuitively clear, and indeed, McGregor et al. [66] observed that if the deletion rate δ is at most $1 - c\sqrt{(\log n)/n}$, then it is trivially easy to extract a random length-$O(\sqrt{n \log n})$ subsequence of x from a typical trace of x. Combining this with the $k = O(\sqrt{n \log n})$ upper bound of Scott [83] and a straightforward sampling-based procedure (which estimates the frequency of each string in $\{0,1\}^k$ to high enough accuracy to determine its exact multiplicity in the k-deck), they obtained an information-theoretic sample complexity upper bound on trace reconstruction: for $\delta \leq 1 - c\sqrt{(\log n)/n}$, at most $n^{O(\sqrt{n \log n})}$ traces suffice to reconstruct any $x \in \{0,1\}^n$ with high probability.

As an initial observation, we slightly strengthen the [66] result by showing that for *any* value of $\delta < 1$, an algorithm which combines sampling and dynamic programming can exactly infer the k-deck of an unknown string $x \in \{0,1\}^n$ with high probability using $(n/(1-\delta))^{O(k)}$ traces from $\mathrm{Del}_\delta(x)$. (See Theorem 9.0.2 for a detailed statement and proof of a more general version of this result.) Combining this with the [55] upper bound $k = O(\sqrt{n})$, we get that any string x can be reconstructed from δ-deletion noise using $(n/(1-\delta))^{O(\sqrt{n})}$ samples.

The above-outlined approach to trace reconstruction (the $\ell = 1$ case of population recovery) is the starting point for our main positive result, Theorem 7.0.1. In the next subsection we give a high-level description of some of the challenges that arise in dealing with multiple strings and how this work overcomes them.

Ingredients in the proof of Theorem 7.0.1

Recall that in the setting of Theorem 7.0.1 the unknown $\mathbf{X}$ is an arbitrary distribution supported on at most ℓ strings $x^1, \ldots, x^\ell$ in $\{0,1\}^n$. Viewing $\mathbf{X}$ as a mixture of individual strings, there is a natural notion of the k-deck of $\mathbf{X}$, which we denote by $\mathsf{D}_k(\mathbf{X})$ and which is the weighted multi-set corresponding to the $\mathbf{X}$-mixture of the decks $\mathsf{D}_k(x^1), \ldots, \mathsf{D}_k(x^\ell)$.[2]

[2]By a weighted-multiset we mean a multiset in which each element has a weight. Alternatively, one can interpret (after normalization) $\mathsf{D}_k(x)$ as a probability distribution over the 2^k strings in $\{0,1\}^k$ and in this case, $\mathsf{D}_k(\mathbf{X})$ can be viewed as a probability distribution that is the $\mathbf{X}$-mixture of $\mathsf{D}_k(x^1), \ldots, \mathsf{D}_k(x^\ell)$.

As a result, Theorem 7.0.1 will follow if we can show the following: if two distributions $\mathbf{X}, \mathbf{Y}$ over $\{0,1\}^n$ (each supported on at most ℓ strings) have $d_{\mathrm{TV}}(\mathbf{X}, \mathbf{Y}) > \varepsilon$, then for a not-too-large value of k, the k-decks $\mathsf{D}_k(\mathbf{X})$ and $\mathsf{D}_k(\mathbf{Y})$ (note that these are two weighted multi-sets of strings in $\{0,1\}^k$) must be "noticeably different." This is established in Lemma 9.0.6, which is the technical heart of our upper bound.

To explain our proof of Lemma 9.0.6 it is useful to revisit the $\ell = 1$ setting; the analogous (and much easier to prove) statement in this context is that given any two strings $x \neq y \in \{0,1\}^n$, the k-decks $\mathsf{D}_k(x)$ and $\mathsf{D}_k(y)$ are not identical when $k \geq C\sqrt{n}$ for some large enough constant C. This is the main result of [55] (and a similar statement, with a slightly weaker quantitative bound on k, is also proved in [83]). Since the k-deck in and of itself is somewhat difficult to work with (being a multi-set over $\{0,1\}^k$), both [55] and [83] work instead with the *summed k-deck*, which we denote by $\mathsf{SD}_k(x)$ and which is simply the vector in $\mathbb{N}^k$ obtained by summing all $\binom{n}{k}$ elements of the k-deck $\mathsf{D}_k(x)$ (recall that each element of $\mathsf{D}_k(x)$ is a vector in $\{0,1\}^k$). Both [55] and [83] actually show that for a suitable value of k, the *summed* k-deck $\mathsf{SD}_k(x)$ uniquely identifies x among all strings in $\{0,1\}^n$. (Both papers also observe that by a simple counting argument, the smallest such k is at least $\tilde{\Omega}(\sqrt{n})$.) The [55] proof reduces the analysis of the summed k-deck to an extremal problem about univariate polynomials. The key ingredient of their proof is the following result about univariate polynomials, which was established in [10] in their work on the Prouhet-Tarry-Escott problem:

> Given any nonzero vector $\delta \in \{-1,0,1\}^n$, there is a univariate polynomial p of degree $O(\sqrt{n})$ such that
>
> $$\sum_{0 \leq i < n} \delta_i \cdot p(i) \neq 0. \qquad (\dagger)$$

Setting $\delta = x - y \neq 0$, to finish the proof of $\mathsf{SD}_k(x) \neq \mathsf{SD}_k(y)$ when $x \neq y$ and $k \geq C\sqrt{n}$, [55] shows that choosing k to be $\deg(p) + 1$, the inequality $(\dagger)$ implies that $\mathsf{SD}_k(x) \neq \mathsf{SD}_k(y)$.

Returning to our ℓ-string setting, we remark that several challenges arise which are not present in the one-string setting. To highlight one of these, due to the difficulty of analyzing the entire k-deck of $\mathbf{X}$ it is natural to try to work with the summed k-deck $\mathsf{SD}_k(\mathbf{X})$ (a nonnegative vector in $\mathbb{R}^k$), which is obtained by summing all elements of the weighted multi-set $\mathsf{D}_k(\mathbf{X})$. Indeed it can be shown via a rather straightforward extension of the [55] analysis that, when $\mathbf{X}$ is uniform over $x^1, \ldots, x^\ell$, the summed k-deck with $k = O(\sqrt{n} \log \ell)$ suffices to exactly reconstruct the *sum* $x^1 + \cdots + x^\ell$ (a vector in $\mathbb{N}^n$).

But even for uniform distributions, a difficulty which arises is that the summed k-deck (even with $k = n$) cannot distinguish between two uniform distributions over $x^1, \ldots, x^\ell$ versus $y^1, \ldots, y^\ell$ that have the same coordinate-wise sums, i.e. that satisfy $x^1 + \cdots + x^\ell = y^1 + \cdots + y^\ell$.[3] Indeed, considering the same example as earlier, in which $\ell = 2$ and $x^1 = 0^n$, $x^2 = 1^n, y^1 = 0^{n/2}1^{n/2}$ and $y^2 = 1^{n/2}0^{n/2}$, the summed k-deck is $(\binom{n}{k}, \ldots, \binom{n}{k})/2 \in \mathbb{R}^k$ in both cases.

At a high level our Lemma 9.0.6 can be viewed as a *robust* generalization of the [55] result. A key technical ingredient in its proof is a robust generalization of the [10] result to *multivariate* polynomials. (The summed k-deck corresponds to univariate polynomials, so at a high level our analysis involving multivariate polynomials can be viewed as how we get around the obstacle noted in the previous paragraph.) The proof of Lemma 9.0.6 consists of three steps which we outline below.

The first conceptual step of our argument is to show that if two support-ℓ distributions $\mathbf{X}$ and $\mathbf{Y}$ over $\{0,1\}^n$ satisfy $d_{\mathrm{TV}}(\mathbf{X}, \mathbf{Y}) \geq \varepsilon$, then there exists a subset $T \subset [0 : n-1]$ of size $d = \lfloor \log(2\ell) \rfloor$ such that $\mathbf{X}$ and $\mathbf{Y}$ "differ significantly" just on the coordinates in T. In particular, there is some $|T|$-bit string c such that $\mathbf{Pr}_{\boldsymbol{x}\sim\mathbf{X}}[\boldsymbol{x}_T = c]$ is significantly different from $\mathbf{Pr}_{\boldsymbol{y}\sim\mathbf{Y}}[\boldsymbol{y}_T = c]$, where we use x_T to denote the restriction of a string $x \in \{0,1\}^n$ on coordinates in T. (This is made precise in Lemma 9.0.1.) Let $\Delta : \binom{[0:n-1]}{d} \to \mathbb{R}$ be the following function over size-d subsets of $[0 : n-1]$:

$$\Delta(S) = \mathbf{Pr}_{\boldsymbol{x}\sim\mathbf{X}}\Big[\boldsymbol{x}_S = c\Big] - \mathbf{Pr}_{\boldsymbol{y}\sim\mathbf{Y}}\Big[\boldsymbol{y}_S = c\Big]. \tag{7.2}$$

Then Lemma 9.0.1 implies that $\|\Delta\|_\infty$ is not too small.

The second (and central) conceptual step of our argument can be viewed as a robust generalization of the [10] result to d-variate polynomials, as alluded to earlier. The key result giving this step, Lemma 9.0.7, roughly speaking states the following:

> Given the Δ as defined in (7.2), there is a d-variate polynomial ϕ of not-too-high degree (roughly $\sqrt{n}$) such that[4]
>
> $$\left| \sum_{0 \leq t_1 < \cdots < t_d < n} \phi(t_1, \ldots, t_d) \cdot \Delta\Big(\{t_1, \ldots, t_d\}\Big) \right| \tag{††}$$
>
> can be lower bounded in terms of $\|\Delta\|_\infty$, which is not too small by Lemma 9.0.1.

[3] This is conceptually similar to the inability of mean-based algorithms to handle multiple strings noted earlier.

[4] The reader who has peeked ahead to the statement of Lemma 9.0.7 may have noticed that the lemma statement also bounds the magnitudes of coefficients of the polynomial ϕ. This is done for technical reasons, and we skip these technical details in the high-level description here.

The technical details of this step are deferred to [4].

The third conceptual step relates (††) to the distance between the k-decks $\mathsf{D}_k(\mathbf{X})$ and $\mathsf{D}_k(\mathbf{Y})$, by showing that if (††) is not too small then $\mathsf{D}_k(\mathbf{X})$ and $\mathsf{D}_k(\mathbf{Y})$ must be "noticeably different" when k is chosen to be $\deg(\phi) + d$. We refer the reader to Lemma 9.0.8.

At a high level this is analogous to, but technically more involved than, the [55] proof that the inequality (†) for $\delta = x - y$ implies that $\mathsf{SD}_k(x) \neq \mathsf{SD}_k(y)$ with $k = \deg(p) + 1$.

Lemma 9.0.6 then follows by combining all three steps, i.e. $d_{\mathrm{TV}}(\mathbf{X}, \mathbf{Y})$ being large implies that $\mathsf{D}_k(\mathbf{X})$ is "noticeably different" from $\mathsf{D}_k(\mathbf{Y})$ for k that is roughly $\sqrt{n}$.

Our lower bounds

We begin by recalling the $\Omega(n)$ lower bound of McGregor et al. [66]. This lower bound is obtained via a simple analysis of the two distributions of traces resulting from the two strings $x^1 = 0^{n/2}10^{n/2-1}$ and $x^2 = 0^{n/2-1}10^{n/2}$. The starting point of the [66] analysis is the observation that under the δ-deletion channel, conditioned on the sole "1" coordinate being retained, the distribution of a trace of x^1 corresponds to $(\boldsymbol{a}, \boldsymbol{b})$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ are independent draws from $\mathrm{Bin}(n/2, 1-\delta)$ and $\mathrm{Bin}(n/2-1, 1-\delta)$ respectively, whereas the distribution of a trace of x^2 corresponds to $(\boldsymbol{b}, \boldsymbol{a})$. [66] used this to show that the squared Hellinger distance between these two distributions of traces is $O(1/n)$, and in turn use this squared Hellinger distance bound to infer an $\Omega(n)$ sample complexity lower bound for determining whether a collection of received traces came from x^1 or from x^2.

Our lower bound approach may be viewed as an extension of the [66] lower bound to *mixtures* of distributions similar to the ones they consider. The high-level idea of our lower bound proof is as follows: we show that there exist two distributions $\mathbf{X}, \mathbf{Y}$ over $\{0,1\}^n$ (in fact, over n-bit strings with precisely one 1) which have disjoint supports, each of size at most 2ℓ, but are such that the total variation distance $d_{\mathrm{TV}}(\mathrm{Del}_\delta(\mathbf{X}), \mathrm{Del}_\delta(\mathbf{Y}))$, between traces of strings drawn from $\mathbf{X}$ versus traces of strings drawn from $\mathbf{Y}$, is very small. This is easily seen to imply Theorem 7.0.2.

For simplicity in introducing the main ideas of our analysis, in this expository overview we will first consider an "$n = +\infty$" version of our population recovery scenario. We begin by considering the distribution $\mathrm{Del}_\delta(\tilde{e}_{m+i})$ where m is some fixed value and $\tilde{e}_{m+i}$ is an infinite string with a single 1 in position $m+i$ and all other coordinates 0. A δ fraction of the outcomes of $\mathrm{Del}_\delta(\tilde{e}_{m+i})$ are the infinite all-0 string, which conveys no information. The other $1-\delta$ fraction of the outcomes each have precisely one 1, occurring in position $1 + \boldsymbol{a}$ where $\boldsymbol{a}$ is distributed according to the binomial distribution $\mathrm{Bin}(m+i, 1-\delta)$. In this infinite-$n$ setting, two distributions

$\mathbf{X}, \mathbf{Y}$ over strings of the form $\tilde{e}_{m+i}$ with disjoint supports correspond to two mixtures of distinct binomial distributions (all with second parameter $1-\delta$, but with a set of first parameters in the first mixture that is disjoint from the set of first parameters in the second mixture). The animating idea behind our construction and analysis is that it is possible for two distinct mixtures of binomials like this to be very close to each other in total variation distance.[5]

In order to show that two distinct mixtures of binomial distributions as described above can be very close to each other in total variation distance, our lower bounds employ technical machinery due to Roos [82] and Daskalakis and Papadimitriou [25]. Roos [82] developed a "Krawtchouk expansion" which provides an *exact* expression for the probability that a Poisson binomial distribution (a sum of n independent Bernoulli random variables with expectations $p_1, \ldots, p_n$) puts on any given outcome in $\{0, 1, \cdots, n\}$. Daskalakis and Papadimitriou [25] used Roos's Krawtchouk expansion to show that under mild technical conditions, low-order moments of any Poisson binomial distribution essentially determine the entire distribution. In more detail, their main result is that if $\mathbf{X}, \mathbf{Y}$ are two Poisson binomial distributions (satisfying mild technical conditions) whose t-th moments match, i.e. $\mathbf{E}[\mathbf{X}^t] = \mathbf{E}[\mathbf{Y}^t]$ for $t = 1, \ldots, O(\log(1/\varepsilon))$, then the total variation distance between $\mathbf{X}$ and $\mathbf{Y}$ is at most ε.

Our analysis proceeds in two main steps. In the first step, we show that there exist two mixtures of pairs of binomial distributions, which we denote by $\mathbf{D}_S$ and $\mathbf{D}_T$, with certain desirable properties. S and T are both subsets of $\{0, \ldots, 2\ell\}$, and $\mathbf{D}_S$ is a certain mixture of pairs of binomial distributions $(\mathrm{Bin}(n/2+i, 1-\delta), \mathrm{Bin}(n/2-i, 1-\delta))$ for $i \in S$ while $\mathbf{D}_T$ is a certain mixture of pairs of binomial distributions $(\mathrm{Bin}(n/2+j, 1-\delta), \mathrm{Bin}(n/2-j, 1-\delta))$ for $j \in T$. We establish the existence of *disjoint* sets S, T such that the resulting mixtures $\mathbf{D}_S$ and $\mathbf{D}_T$ have matching t-th moments for all $t = 1, \ldots, \ell$. This is proved using known algebraic expressions for the moments of binomial distributions and simple linear algebraic arguments. In the second main step, we extend the analysis of Daskalakis and Papadimitriou [25] and apply this extension to our setting, in which we are dealing with mixtures of (pairs of) binomial distributions (as opposed to their and Roos's setting of Poisson binomial distributions). We show that the matching first ℓ moments of $\mathbf{D}_S$ and $\mathbf{D}_T$ imply

[5] We remark that our actual scenario is more complicated than this idealized version because n is a finite value rather than $+\infty$. For $n = 2m+1$, this means that a received trace $0^{\boldsymbol{a}}10^{\boldsymbol{b}}$ which contains a 1 and came from $\mathrm{Del}_\delta(e_{m+i})$ provides a pair of values $(\boldsymbol{a}, \boldsymbol{b})$ where $\boldsymbol{a}$ is distributed according to $\mathrm{Bin}(m+i, \rho)$ and $\boldsymbol{b}$ is independently distributed according to $\mathrm{Bin}(m-i, \rho)$ where $\rho = 1-\delta$ is the retention probability. This second value $\boldsymbol{b}$ provides additional information which is not present in the $n = +\infty$ version of the problem, and this makes it more challenging and more technically involved to prove a lower bound. We deal with these issues in Section 10.1.

that the distributions $\mathrm{Del}_\delta(\mathbf{X})$ and $\mathrm{Del}_\delta(\mathbf{Y})$ are very close, where $\mathbf{X}$ corresponds to the mixture of Hamming-weight-one strings in $\{0,1\}^n$ corresponding to $\mathbf{D}_S$ and $\mathbf{Y}$ likewise corresponds to the mixture of Hamming-weight-one strings corresponding to $\mathbf{D}_T$. (In fact, in our setting having ℓ matching moments leads to $n^{-\Omega(\ell)}$-closeness in total variation distance, whereas in [25] the resulting closeness from ℓ matching moments was $2^{-\Omega(\ell)}$.)

We close this subsection by observing that while the results of [25, 82] were used in a crucial way in subsequent work of Daskalakis et al. [24] to obtain a sample complexity *upper bound* on learning Poisson binomial distributions, in our context we use these results to obtain a sample complexity *lower bound* for population recovery. Intuitively, the difference is that in the [24] scenario of learning an unknown Poisson binomial distribution, there is no noise process affecting the samples: the learning algorithm is assumed to directly receive draws from the underlying Poisson binomial distribution being learned. In such a noise-free setting, the existence of a small ε-cover for the space of all Poisson binomial distributions (which is established in [25] as a consequence of their moment-matching result) means, at least on a conceptual level, that a learning algorithm "need only search a small space of candidates" to find a high-accuracy hypothesis. In contrast, in our context of deletion-channel noise, our arguments show that it is possible for two underlying true distributions $\mathbf{X}, \mathbf{Y}$ over $\{0,1\}^n$ to be very different (indeed, to have disjoint supports) but to be such that their deletion-noise-corrupted versions have low-order moments which match each other exactly. In this scenario, the [25, 82] results can be used to show that the variation distance between the two distributions of noisy samples received by the learner is very small, and this gives a sample complexity lower bound for distinguishing $\mathbf{X}$ and $\mathbf{Y}$ on the basis of such noisy samples.

Outline: In Chapter 8, we give preliminaries. In Chapter 9, we present our population recovery algorithm. In Chapter 10, we prove our lower bounds.

Chapter 8

Preliminaries

Notation. Given a nonnegative integer n, we write $[n]$ to denote $\{1, \ldots, n\}$. Given integers $a \leq b$ we write $[a : b]$ to denote $\{a, \ldots, b\}$. It will be convenient for us to index a binary string $x \in \{0,1\}^n$ using $[0 : n-1]$ as $x = (x_0, \ldots, x_{n-1})$. Given a vector $v = (v_1, \ldots, v_d) \in \mathbb{R}^d$, we write $\|v\|_\infty$ to denote $\max_{i \in [d]} |v_i|$. Given a function $\Delta : A \to \mathbb{R}$ over a finite domain A, we write $\|\Delta\|_\infty = \max_{a \in A} |\Delta(a)|$. Given a polynomial p (which may be univariate or multivariate), we write $\|p\|_1$ to denote the sum of magnitudes of p's coefficients. All logarithms and exponents are binary (base 2) unless otherwise specified.

Distributions. We use bold font letters to denote probability distributions and random variables, which should be clear from the context. We write "$\boldsymbol{x} \sim \mathbf{X}$" to indicate that random variable $\boldsymbol{x}$ is distributed according to distribution $\mathbf{X}$. The total variation distance between two distributions $\mathbf{X}$ and $\tilde{\mathbf{X}}$ over a finite set $\mathcal{X}$ is defined as

$$d_{\mathrm{TV}}(\mathbf{X}, \tilde{\mathbf{X}}) = \frac{1}{2} \sum_{x \in \mathcal{X}} \left| \mathbf{X}(x) - \tilde{\mathbf{X}}(x) \right|,$$

where $\mathbf{X}(x)$ denotes the amount of probability mass that the distribution $\mathbf{X}$ puts on outcome x.

Population recovery from the deletion channel. Throughout this Part the parameter $0 < \delta < 1$ denotes the *deletion probability*. Given a string $x \in \{0,1\}^n$, we write $\mathrm{Del}_\delta(x)$ to denote the distribution of a random trace of x after it has been passed through the δ-deletion channel (so the distribution $\mathrm{Del}_\delta(x)$ is supported on $\{0,1\}^{\leq n}$). Recall that a random trace $\boldsymbol{y} \sim \mathrm{Del}_\delta(x)$ is obtained by independently deleting each bit of x with probability δ and concatenating the surviving bits.[1]

[1]For simplicity in this work we assume that the deletion probability δ is known to the learning

We now define the problem of population recovery from the deletion channel that we will study in this Part. In this problem the goal is to learn an unknown *target distribution* $\mathbf{X}$ supported on at most ℓ strings from $\{0,1\}^n$.

The learning algorithm has access to independent samples, each of which is generated independently by first drawing a string $\boldsymbol{x} \sim \mathbf{X}$ and then outputting a trace from $\mathrm{Del}_\delta(\boldsymbol{x})$. For conciseness we write $\mathrm{Del}_\delta(\mathbf{X})$ to denote this distribution.

The goal for the learning algorithm is to output with high probability (say at least 0.99) a *hypothesis distribution* $\tilde{\mathbf{X}}$ for $\mathbf{X}$ which is ε-accurate in total variation distance: $d_{\mathrm{TV}}(\mathbf{X}, \tilde{\mathbf{X}}) \le \varepsilon$. We are interested in the number of samples needed for this learning task in terms of n, ℓ, ε and δ.

Decks. Given a subset $T = \{t_1, \ldots, t_k\} \subseteq [0 : n-1]$ of size k with $t_1 < \cdots < t_k$, and two strings $v \in \{0,1\}^k$, $x \in \{0,1\}^n$, we say that v *matches* x *at* T if $x_T = v$, where $x_T = (x_{t_1}, \ldots, x_{t_k}) \in \{0,1\}^k$ denotes the string x restricted to positions in T. We say that the *number of occurrences of* v *in* x is the number of size-k subsets $T \subseteq [0 : n-1]$ such that v matches x at T, and we write $\#(v, x)$ to denote this quantity. Given a distribution $\mathbf{X}$ over $\{0,1\}^n$, we write $\#(v, \mathbf{X})$ to denote the expected number of occurrences of v in $\boldsymbol{x} \sim \mathbf{X}$, i.e.

$$\#(v, \mathbf{X}) = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathbf{X}} \big[\#(v, \boldsymbol{x})\big].$$

Given a string $x \in \{0,1\}^n$, we write $\mathsf{D}_k(x)$ to denote the (*normalized*[2]) *k-deck* of x.

This is a 2^k-dimensional vector indexed by strings $v \in \{0,1\}^k$ such that

$$\big(\mathsf{D}_k(x)\big)_v = \frac{\#(v, x)}{\binom{n}{k}}.$$

So $\mathsf{D}_k(x)$ is a nonnegative vector that sums to 1.

Similarly, for a distribution $\mathbf{X}$ over strings from $\{0,1\}^n$, we write $\mathsf{D}_k(\mathbf{X})$ to denote the (*normalized*[3]) *k-deck* of $\mathbf{X}$, given by

$$\big(\mathsf{D}_k(\mathbf{X})\big)_v = \frac{\#(v, \mathbf{X})}{\binom{n}{k}},$$

for each $v \in \{0,1\}^k$. So $\mathsf{D}_k(\mathbf{X})$ is also a 2^k-dimensional nonnegative vector that sums to 1.

---

algorithm. We note that it is possible to obtain a high-accuracy estimate of δ simply by measuring the average length of traces received from the deletion channel.

[2]It will be more convenient for us to use the notion of (normalized) k-decks defined here; note that we can recover from it the multi-set of all subsequences of x with length k, and vice versa.

[3]Similarly, the (normalized) k-deck here is equivalent to the weighted multi-set version used in the introduction up to a simple rescaling.

Chapter 9

Upper bounds

Our goal is to prove Theorem 9.0.1, which is restated below:

Theorem 9.0.1. *There is an algorithm A which has the following performance guarantee: For any distribution $\mathbf{X}$ supported over at most ℓ strings in $\{0,1\}^n$, if A is given*

$$\frac{1}{\varepsilon^2} \cdot \left(\frac{2}{1-\delta}\right)^{\sqrt{n}\cdot(\log n)^{O(\ell)}} \tag{9.1}$$

many samples from $\mathrm{Del}_\delta(\mathbf{X})$, then with probability at least 0.99 the algorithm outputs a probability distribution $\tilde{\mathbf{X}}$ supported over at most ℓ strings such that $d_{\mathrm{TV}}(\mathbf{X}, \tilde{\mathbf{X}}) \le \varepsilon$.

In Section 9 we introduce the notion of a *restriction*, which is a "local view" of a distribution $\mathbf{X}$ confined to a specific subset of coordinates and a specific outcome for those coordinates. We then provide some terminology and prove three useful lemmas about restrictions in Section 9.

Next in Section 9 we describe the algorithm A, state our main technical lemma, Lemma 9.0.6, and use it to prove the correctness of algorithm A.

We prove Lemma 9.0.6 in Sections 9 and 9.

Notational convention. Our argument below involves many integer-valued index variables which take values in a range of different intervals. To help the reader keep track, we will use the following convention (the values L and m will be defined later):

- $s, t, s_1, t_1, \ldots$ will denote an index ranging over $[0 : n-1]$;
- $j, j_1, \ldots$ will denote an index ranging over $[0 : k-1]$;
- $a, a', a_1, \ldots$ will denote an index ranging over $[L]$;
- $b, b', b_1, \ldots$ will denote an index ranging over $[0 : m]$;

- $i, i_1, \ldots,$ $\alpha, \alpha_1, \ldots$ and $\beta, \beta_1, \ldots$ will denote an index in all other places.

Restrictions

Let $\mathbf{X}$ be a distribution over strings from $\{0,1\}^n$ and let $d \in [n]$ be a parameter (which should be thought of as quite small; we will set $d = O(\log \ell)$ below). Given a size-d subset $T = \{t_1, \ldots, t_d\}$ of $[0 : n-1]$ with $0 \le t_1 < \cdots < t_d < n$ and a string $c \in \{0,1\}^d$, we define

$$\mathrm{restrict}(\mathbf{X}, T, c) := \mathbf{Pr}_{\boldsymbol{x} \sim \mathbf{X}} \Big[(\boldsymbol{x}_{t_1}, \ldots, \boldsymbol{x}_{t_d}) = c \Big],$$

the probability that a draw of $\boldsymbol{x} \sim \mathbf{X}$ matches c in the coordinates of T.

Let $\mathbf{X}$ and $\mathbf{Y}$ be two distributions, each supported over at most ℓ strings from $\{0,1\}^n$. Our first lemma shows that if $d_{\mathrm{TV}}(\mathbf{X}, \mathbf{Y})$ is large, then there are a size-d subset T and a string $c \in \{0,1\}^d$ with $d = \lfloor \log(2\ell) \rfloor$ such that there is a reasonably big gap between $\mathrm{restrict}(\mathbf{X}, T, c)$ and $\mathrm{restrict}(\mathbf{Y}, T, c)$.

Lemma 9.0.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two distributions, each supported over at most ℓ strings from $\{0,1\}^n$. Then there exist a size-d subset T of $[0 : n-1]$ and a string $c \in \{0,1\}^d$ with $d = \lfloor \log(2\ell) \rfloor$ such that*

$$\Big| \mathrm{restrict}(\mathbf{X}, T, c) - \mathrm{restrict}(\mathbf{Y}, T, c) \Big| \ge \frac{d_{\mathrm{TV}}(\mathbf{X}, \mathbf{Y})}{\ell^{O(\ell)}}.$$

Proof. Let $\mathrm{supp}(\mathbf{X}) \cup \mathrm{supp}(\mathbf{Y}) = \{z^1, \ldots, z^{\ell'}\}$ for some $\ell' \le 2\ell$. For each $i \in [\ell']$, let $p_i \ge 0$ be the magnitude of the difference between the probabilities of z^i in $\mathbf{X}$ and in $\mathbf{Y}$. Let $\varepsilon = d_{\mathrm{TV}}(\mathbf{X}, \mathbf{Y})$. Then by definition we have $\sum_i p_i = 2\varepsilon$. Without loss of generality we assume that $p_1 \ge \cdots \ge p_{\ell'} \ge 0$ and prove the following claim (where we set $p_{\ell'+1} = 0$ by default for convenience):

Claim 9.0.2. *There exists an $i^* \in [\ell']$ such that $p_{i^*} \ge \varepsilon/(4\ell)^{\ell'}$ and $p_{i^*+1} \le p_{i^*}/(4\ell)$.*

Proof. First we notice that $p_1 \ge \varepsilon/\ell$ given that $\sum_i p_i = 2\varepsilon$ and $\ell' \le 2\ell$. Now given that the p_i's are nonnegative, there exists an $i \in [\ell']$ (e.g., by taking $i = \ell'$) such that $p_{i+1} \le p_i/(4\ell)$. Take i^* to be the smallest such index i. Then we have

$$\frac{p_{i^*}}{p_1} = \frac{p_{i^*}}{p_{i^*-1}} \cdots \frac{p_2}{p_1} > \frac{1}{(4\ell)^{i^*-1}}$$

by the choice of i^* as the smallest such index. As a result, we have

$$p_{i^*} \ge \frac{\varepsilon}{(4\ell)^{i^*}} \ge \frac{\varepsilon}{(4\ell)^{\ell'}}.$$

This finishes the proof of the claim. □

Let $i^* \in [\ell']$ be the integer given by the claim above, and we consider the first i^* strings $z^1, \ldots, z^{i^*}$. Given that $i^* \leq \ell' \leq 2\ell$, there exist a d-subset T of $[0 : n-1]$ with $d = \lfloor \log(2\ell) \rfloor$, a string $c \in \{0,1\}^d$ and an $i' \leq i^*$ such that the restriction of $z^{i'}$ matches c but the restriction of z^i does not match c for any other $i \leq i^*$. (This can be achieved by repeatedly selecting a coordinate that splits the remaining strings into two nonempty subsets and setting c to reduce the size by at least half each time.) Using properties of i^* given in the claim above, we have

$$\left|\text{restrict}(\mathbf{X}, T, c) - \text{restrict}(\mathbf{Y}, T, c)\right| \geq p_{i^*} - \sum_{i>i^*} p_i \geq p_{i^*} - 2\ell \cdot \frac{p_{i^*}}{4\ell} = \frac{p_{i^*}}{2} \geq \frac{\varepsilon}{\ell^{O(\ell)}}.$$

This finishes the proof of the lemma. □

Given two size-d subsets $S = \{s_1, \ldots, s_d\}$ and $T = \{t_1, \ldots, t_d\}$ of $[0 : n-1]$ with $s_1 < \cdots < s_d$ and $t_1 < \cdots < t_d$, we say that S is *dominated* by T if $s_i \leq t_i$ for every $i \in [d]$. Let $\Delta : \binom{[0:n-1]}{d} \to \mathbb{R}$ be a function over size-d subsets of $[0 : n-1]$. We use $\text{supp}(\Delta)$ to denote the set of subsets T with $\Delta(T) \neq 0$. We need the following definitions of a *cover* and a *group cover* of such a function Δ.

Definition 9.0.1 (Covers and group covers)**.** *We say that a function* $\Delta : \binom{[0:n-1]}{d} \to \mathbb{R}$ *has an* L-cover $\{(T_a, \mathcal{S}_a) : a \in [L]\}$ *for some* $L \geq 0$ *if*

1. $\mathcal{S}_1, \ldots, \mathcal{S}_L$ *form an* L*-way partition of* $\text{supp}(\Delta)$*;*
2. $T_a \in \mathcal{S}_a$ *for each* $a \in [L]$*;*
3. $\Delta(T) = \Delta(T_a)$ *for every* $T \in \mathcal{S}_a$*; and*
4. T_a *is dominated by every* $T \in \mathcal{S}_a$*.*

We refer to the set T_a *as the* anchor set *of the collection* $\mathcal{S}_a$*.*

Furthermore we say that Δ *has an* (L, q, λ)-group cover *if* Δ *has an* L*-cover* $\{(T_a, \mathcal{S}_a) : a \in [L]\}$ *and a* q*-way partition of* $[L]$ *into* $A_1, \ldots, A_q$ *such that for each* $i \in [q]$*, for all* $a, a' \in A_i$ *we have*

$$\frac{|\Delta(T_a)|}{|\Delta(T_{a'})|} \leq \lambda.$$

Given distributions $\mathbf{X}$ and $\mathbf{Y}$ over strings from $\{0,1\}^n$ and a string $c \in \{0,1\}^d$, we write $\Delta_{\mathbf{X},\mathbf{Y},c}$ to denote the function over size-d subsets of $[0 : n-1]$ that maps a size-d subset T to

$$\Delta_{\mathbf{X},\mathbf{Y},c}(T) := \text{restrict}(\mathbf{X}, T, c) - \text{restrict}(\mathbf{Y}, T, c).$$

The second lemma shows that when d and the supports of $\mathbf{X}, \mathbf{Y}$ are small, the function $\Delta_{\mathbf{X},\mathbf{Y},c}$ has a small cover for any string $c \in \{0,1\}^d$. Taking as an example when $\ell = d = 2$ and $\text{supp}(\mathbf{X}) = \{x^1, x^2\}$, we have that $\text{restrict}(\mathbf{X}, S, c) = \text{restrict}(\mathbf{X}, T, c)$ if $x^1_S = x^1_T$ and $x^2_S = x^2_T$ (note that this is a sufficient but not necessary condition in general). Letting $S = \{s_1, s_2\}$ for some $s_1 < s_2$ and $T = \{t_1, t_2\}$ for some $t_1 < t_2$, this condition can be written equivalently as

$$(x^1_{s_1}, x^2_{s_1}) = (x^1_{t_1}, x^2_{t_1}) \quad \text{and} \quad (x^1_{s_2}, x^2_{s_2}) = (x^1_{t_2}, x^2_{t_2}).$$

This implies that $\text{restrict}(\mathbf{X}, \cdot, c)$, as a function over size-2 subsets, has the following combinatorial "rectangular" structure: one can partition indices $t \in [0 : n-1]$ into four types 00,01,10,11 according to values of x^1_t and x^2_t; this induces a partition of all size-2 subsets into 16 "rectangles," [1] where $S = \{s_1 < s_2\}$ and $T = \{t_1 < t_2\}$ belong to the same "rectangle" iff the type of s_1 is the same as that of t_1 and the type of s_2 is the same as that of t_2. It follows that all T in the same "rectangle" share the same value $\text{restrict}(\mathbf{X}, T, c)$. We use this observation to obtain a small cover for $\Delta_{\mathbf{X},\mathbf{Y},c}$.

Lemma 9.0.3. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two distributions, each supported over at most ℓ strings from $\{0,1\}^n$. For any $d \in [n]$ and any string $c \in \{0,1\}^d$, $\Delta_{\mathbf{X},\mathbf{Y},c}$ has an L-cover for some $L \le 2^{2d\ell}$.*

Proof. Suppose that $\mathbf{X}$ is supported on $x^1, \ldots, x^{\ell'}$ and $\mathbf{Y}$ is supported on $y^1, \ldots, y^{\ell''}$ with $\ell', \ell'' \le \ell$. We say an index $t \in [0 : n-1]$ is of *type-(u, v)*, where $u \in \{0,1\}^{\ell'}$ and $v \in \{0,1\}^{\ell''}$, if

$$(x^1_i, \ldots, x^{\ell'}_i) = u \quad \text{and} \quad (y^1_i, \ldots, y^{\ell''}_i) = v.$$

This allows us to classify size-d subsets of $[0 : n-1]$ into at most $(2^{\ell'+\ell''})^d \le 2^{2d\ell}$ many equivalence classes: $S \sim T$ if $S = \{s_1, \ldots, s_d\}$ with $s_1 < \cdots < s_d$ and $T = \{t_1, \ldots, t_d\}$ with $t_1 < \cdots < t_d$ are such that s_i and t_i are of the same type for all $i \in [d]$.

Let $\mathcal{S}_a$ be a nonempty equivalence class of $\sim$ such that $S = \{s_1, \ldots, s_d\} \in \mathcal{S}_a$ if $s_1 < \cdots < s_d$ and s_i has type-$(u^{(i)}, v^{(i)})$ for each $i \in [d]$. It follows from the definition of $\sim$ that all $S \in \mathcal{S}_a$ have the same $\text{restrict}(\mathbf{X}, S, c)$ and $\text{restrict}(\mathbf{Y}, S, c)$, and hence the same value of $\Delta_{\mathbf{X},\mathbf{Y},c}(S)$. Moreover, we let $T_a = \{t_1, \ldots, t_d\}$ be the following set: t_1 is the smallest index of type-$(u^{(1)}, v^{(1)})$ and for each i from 2 to d, t_i is the smallest index that is larger than t_{i-1} and has type-$(u^{(i)}, v^{(i)})$. Because $\mathcal{S}_a$ is nonempty, T_a is well defined and it is easy to verify that T_a is dominated by every $S \in \mathcal{S}_a$. As a result, $\Delta_{\mathbf{X},\mathbf{Y},c}$ has the following L-cover:

$$\Big\{(T_a, \mathcal{S}_a) : \mathcal{S}_a \text{ is nonempty and } \Delta_{\mathbf{X},\mathbf{Y},c}(T_a) \neq 0\Big\},$$

[1] Strictly speaking, these are not rectangles since we always need to order indices of a subset in ascending order.

for some $L \leq 2^{2d\ell}$. This finishes the proof of the lemma. □

The last lemma shows that the function $\Delta_{\mathbf{X},\mathbf{Y},c}$ actually has an (L, q, λ)-group cover, for some parameters $L \leq 2^{2d\ell}$, $q \leq \ell$ and $\lambda \leq \ell^{O(\ell)}$.

Lemma 9.0.4. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two distributions, each supported over at most* ℓ *strings from* $\{0,1\}^n$. *For any* $d \in [n]$ *and* $c \in \{0,1\}^d$, $\Delta_{\mathbf{X},\mathbf{Y},c}$ *has an* $(L, q, \ell^{O(\ell)})$*-group cover for some* $L \leq 2^{2d\ell}$ *and* $q \leq \ell$.

Proof. First we apply Lemma 9.0.3 to obtain an L-cover $\{(T_a, \mathcal{S}_a) : a \in [L]\}$ of $\Delta := \Delta_{\mathbf{X},\mathbf{Y},c}$ for some $L \leq 2^{2d\ell}$. It suffices to show that the L positive numbers $|\Delta(T_a)|$, $a \in [L]$, can be divided into at most ℓ groups such that any two in the same group have the ratio bounded from above by $\ell^{O(\ell)}$.

Let $p_1, \ldots, p_{\ell'} > 0$ be probabilities of strings in $\mathbf{X}$ for some $\ell' \leq \ell$ and $q_1, \ldots, q_{\ell''} > 0$ be probabilities of strings in $\mathbf{Y}$ for some $\ell'' \leq \ell$. The observation is that every number $|\Delta(T_a)|$ is a linear form over the p_i's and q_i's with coefficients $-1, 0$ or 1. This motivates the following claim:

Claim 9.0.5. *Let* $u_1, \ldots, u_g > 0$ *be* g (not necessarily distinct) *positive numbers. Let*

$$V = \Big\{ v > 0 : v = c_1 u_1 + \cdots + c_g u_g \text{ for some } c_1, \ldots, c_g \in \{-1, 0, 1\} \Big\}.$$

Then there cannot exist $g+1$ *numbers* $v_1, \ldots, v_{g+1}$ *in* V *satisfying* $v_{g+1} > \cdots > v_1$ *and*

$$\frac{v_{i+1}}{v_i} \geq (g+2)!, \quad \text{for all } i \in [g].$$

Proof. Assume for a contradiction that such $g+1$ numbers $v_1, \ldots, v_{g+1}$ exist in V and let

$$v_i = c_{i,1} u_1 + \cdots + c_{i,g} u_g$$

where $c_{i,j} \in \{-1, 0, 1\}$ for each $i \in [g+1]$. Given that these are $g+1$ many g-dimensional vectors $c_i = (c_{i,1}, \ldots, c_{i,g})$, let $i^* \leq g+1$ be the smallest integer such that c_{i^*} can be written as a linear combination of $c_1, \ldots, c_{i^*-1}$: $c_{i^*} = \alpha_1 c_1 + \cdots + \alpha_{i^*-1} c_{i^*-1}$, which implies that

$$v_{i^*} = \alpha_1 v_1 + \cdots + \alpha_{i^*-1} v_{i^*-1} \leq |\alpha_1| \cdot v_1 + \cdots + |\alpha_{i^*-1}| \cdot v_{i^*-1}. \tag{9.2}$$

We show below that the magnitude of coefficients $\alpha_1, \ldots, \alpha_{i^*-1}$ is relatively small, which leads to a contradiction because we assumed that v_{i^*} is much bigger than $v_{i^*-1}, \ldots, v_1$.

To see this, note that $(\alpha_1, \ldots, \alpha_{i^*-1})$ is the solution to a $(i^*-1) \times (i^*-1)$ linear system $Ax = b$ where A is a $\{-1, 0, 1\}$-valued $(i^*-1) \times (i^*-1)$ full-rank matrix

and b is a $\{-1, 0, 1\}$-valued vector. (In more detail, one can take A to be a full-rank $(i^* - 1) \times (i^* - 1)$ submatrix of the matrix that consists of $c_1, \ldots, c_{i^*-1}$ as columns and take the vector b to be the corresponding entries of c_{i^*}.) It follows from Cramer's rule that each entry of A^{-1} has magnitude at most $(i^* - 1)!$ and thus, each entry of $A^{-1}b$ has absolute value at most $(i^* - 1) \cdot (i^* - 1)! < i^*! \leq (g + 1)!$ This contradicts with (9.2) and the assumption that $v_1 < \ldots < v_{i^*-1} \leq v_{i^*}/(g + 2)!$. □

Claim 9.0.5 gives us the following procedure to partition $[L]$ into $A_1, \ldots, A_q$ for some $q \leq \ell$:

1. Set $i = 1$ and $\mathcal{L} = [L]$.
2. While $\mathcal{L}$ is nonempty do
3. Let v be the smallest $|\Delta(T_a)|$, $a \in \mathcal{L}$.
4. Remove from $\mathcal{L}$ and add to A_i every $a \in \mathcal{L}$ with $|\Delta(T_a)| \leq (2\ell + 2)! \cdot v$, and increment i.

It follows from Claim 9.0.5 that when $\mathcal{L}$ becomes empty at the end, the number of A_i's we created can be no more than ℓ. Furthermore, every a and a' that belong to the same A_i have the ratio of $|\Delta(T_a)|$ and $|\Delta(T_{a'})|$ bounded by $(2\ell+2)! = \ell^{O(\ell)}$. This finishes the proof of the lemma. □

Main Algorithm

We start with an algorithm, based on dynamic programming, for estimating the k-deck of a distribution $\mathbf{X}$ over $\{0, 1\}^n$.

Theorem 9.0.2. *Let $k \in [n]$. There is an algorithm with the following performance guarantee: for any distribution $\mathbf{X}$ over strings in $\{0, 1\}^n$, if the algorithm is given*

$$M = O\left(\frac{k}{\xi^2(1-\delta)^{2k}}\right)$$

many samples from $\mathrm{Del}_\delta(\mathbf{X})$ then with probability at least 0.99 the algorithm outputs a nonnegative 2^k-dimensional vector Q with $\|Q - \mathsf{D}_k(\mathbf{X})\|_\infty \leq \xi$. Its running time is $2^k M \cdot \mathrm{poly}(n)$.

Proof. Let $x^1, \ldots, x^p$ be the support of $\mathbf{X}$. Then for each string $v \in \{0, 1\}^k$, we have

$$\begin{aligned}\mathop{\mathbf{E}}_{\boldsymbol{z}\sim\mathrm{Del}_\delta(\mathbf{X})}\Big[\#(v,\boldsymbol{z})\Big] &= (1-\delta)^k\cdot\Big(\mathbf{X}(x^1)\cdot\#(v,x^1)+\cdots+\mathbf{X}(x^p)\cdot\#(v,x^p)\Big)\\ &= (1-\delta)^k\cdot\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathbf{X}}\Big[\#(v,\boldsymbol{x})\Big]\\ &= (1-\delta)^k\cdot\#(v,\mathbf{X})\\ &= (1-\delta)^k\cdot\binom{n}{k}\cdot\Big(\mathsf{D}_k(\mathbf{X})\Big)_v.\end{aligned}$$

The first equation is because for a given size-k subset $S\subseteq[0:n-1]$ of indices at which v matches x^i, all of the positions in S "survive" into a string $\boldsymbol{z}\sim\mathrm{Del}_\delta(x^i)$ with probability exactly $(1-\delta)^k$.

As a result, it suffices to estimate $\mathbf{E}[\#(v,\boldsymbol{z})]$ to additive accuracy $\pm\xi(1-\delta)^k\binom{n}{k}$ for every string $v\in\{0,1\}^k$. For any fixed string $v\in\{0,1\}^k$, by a standard Chernoff bound, using

$$M=O\left(\frac{k}{\xi^2(1-\delta)^{2k}}\right)$$

samples the empirical estimate of $\mathbf{E}[\#(v,\boldsymbol{z})]$ will have the desired additive $\xi(1-\delta)^k\binom{n}{k}$ accuracy except with failure probability $0.01/2^k$. The success probability of 0.99 follows from union bound.

The running time of the algorithm uses the following simple observation: given $z\in\{0,1\}^{n'}$ and $v\in\{0,1\}^k$, there is a $\mathrm{poly}(n',k)$-time procedure that computes $\#(v,z)$. The procedure works by straightforward dynamic programming: For each $j\in[0:k-1]$ and $i\in[0:n'-1]$, the algorithm maintains a count of the number $\#(v_0\ldots v_j,z_0\ldots z_i)$. This then implies that the running time of the overall algorithm is $M\cdot 2^k\cdot\mathrm{poly}(n)$. This finishes the proof of the lemma. □

We prove the following main technical lemma in Sections 9 and 9. Intuitively, this lemma says that if the total variation distance between $\mathbf{X}$ and $\mathbf{Y}$ is not too small, then for a suitable (not too large) value of k^*, the distance between the k^*-decks of $\mathbf{X}$ and $\mathbf{Y}$ also cannot be too small.

Lemma 9.0.6. *Let ℓ be a positive integer with $\ell\le\log n$. Let $\mathbf{X}$ and $\mathbf{Y}$ be two distributions, each supported over at most ℓ strings from $\{0,1\}^n$. Then there is a positive integer*

$$k^*=\sqrt{n}\cdot(\log n)^{O(\ell)} \tag{9.3}$$

such that

$$d_{\mathrm{TV}}(\mathbf{X},\mathbf{Y})\le\exp\left(\sqrt{n}\cdot(\log n)^{O(\ell)}\right)\cdot\|\mathsf{D}_{k^*}(\mathbf{X})-\mathsf{D}_{k^*}(\mathbf{Y})\|_\infty.$$

We now present our algorithm A and use Lemma 9.0.6 to prove Theorem 9.0.1:

Proof of Theorem 9.0.1. The bound (9.1) we aim for holds trivially when $\ell \geq \log n$. To see this, we first notice that when $\ell \geq \log n$, the sample complexity bound (9.1) we aim for is at least

$$\frac{\mathrm{poly}(\ell)}{\varepsilon^2} \cdot \left(\frac{1}{1-\delta}\right)^n. \tag{9.4}$$

With $(1/(1-\delta))^n$ samples from $\mathrm{Del}_\delta(\mathbf{X})$, we expect to see a full string of length n where no bits are deleted and we know that such a string is drawn directly from $\mathbf{X}$. This means that, with (9.4) many samples, we receive $\mathrm{poly}(\ell)/\varepsilon^2$ draws from $\mathbf{X}$ with high probability. When the latter happens, the empirical estimation $\tilde{\mathbf{X}}$ of $\mathbf{X}$ satisfies $d_{\mathrm{TV}}(\mathbf{X}, \tilde{\mathbf{X}}) \leq \varepsilon$ with high probability. This allows us to focus on the case when $\ell \leq \log n$ in the rest of the proof (so Lemma 9.0.6 applies).

Let ε be the total variation distance we aim for in Theorem 9.0.1. Let k^* be the parameter in (9.3). Let ξ be a parameter to be specified later. By Theorem 9.0.2, the algorithm A can first use

$$M^* = O\left(\frac{k^*}{\xi^2(1-\delta)^{2k^*}}\right) \tag{9.5}$$

samples to obtain an estimate Q of $\mathsf{D}_{k^*}(\mathbf{X})$ such that

$$\|Q - \mathsf{D}_{k^*}(\mathbf{X})\|_\infty \leq \xi, \tag{9.6}$$

and it succeeds in obtaining such an estimate with probability at least 0.99.

With Q in hand the algorithm A computes $\|Q - \mathsf{D}_{k^*}(\mathbf{Y})\|_\infty$ for every distribution $\mathbf{Y}$ supported on at most ℓ strings such that the probability of each string in $\mathbf{Y}$ is an integer multiple of ξ/ℓ. Finally the algorithm outputs the distribution $\mathbf{X}^*$ that minimizes the distance (breaking ties arbitrarily).

We show that when Q satisfies (9.6), $\mathbf{X}^*$ must be close to $\mathbf{X}$. We start with a simple observation that one can round $\mathbf{X}$ to get a distribution $\mathbf{X}'$ in which the probability of each string is an integer multiple of ξ/ℓ and $d_{\mathrm{TV}}(\mathbf{X}, \mathbf{X}') \leq \xi$. This can be done by rounding the probability of every string except one to the nearest multiple of ξ/ℓ and setting the last probability as required so that the total probability is 1. We have

$$\begin{aligned}\left\|Q - \mathsf{D}_{k^*}(\mathbf{X}')\right\|_\infty &\leq \left\|Q - \mathsf{D}_{k^*}(\mathbf{X})\right\|_\infty + \left\|\mathsf{D}_{k^*}(\mathbf{X}) - \mathsf{D}_{k^*}(\mathbf{X}')\right\|_\infty \\ &\leq \left\|Q - \mathsf{D}_{k^*}(\mathbf{X})\right\|_\infty + d_{\mathrm{TV}}(\mathbf{X}, \mathbf{X}') \leq 2\xi.\end{aligned}$$

By definition of $\mathbf{X}^*$ and $\mathbf{X}'$, we have $\|Q - \mathsf{D}_{k^*}(\mathbf{X}^*)\|_\infty \le \|Q - \mathsf{D}_{k^*}(\mathbf{X}')\|_\infty \le 2\xi$. As a result,

$$\left\|\mathsf{D}_{k^*}(\mathbf{X}) - \mathsf{D}_{k^*}(\mathbf{X}^*)\right\|_\infty \le \left\|Q - \mathsf{D}_{k^*}(\mathbf{X}^*)\right\|_\infty + \left\|Q - \mathsf{D}_{k^*}(\mathbf{X})\right\|_\infty \le 3\xi.$$

It follows from Lemma 9.0.6 that

$$d_{\mathrm{TV}}(\mathbf{X}, \mathbf{X}^*) \le 3\xi \cdot \exp\left(\sqrt{n} \cdot (\log n)^{O(\ell)}\right).$$

Finally we choose ξ so that the RHS becomes ε. The number of samples needed in (9.5) becomes

$$\left(\frac{1}{\varepsilon}\right)^2 \cdot \left(\frac{2}{1-\delta}\right)^{\sqrt{n} \cdot (\log n)^{O(\ell)}}.$$

This finishes the proof of Theorem 9.0.1. □

We use the following two lemmas to prove Lemma 9.0.6. They are proved in Section 9 and 9.

Lemma 9.0.7. *Let d, q, L and λ be positive integers satisfying*

$$d, q \le \log n \quad \text{and} \quad L, \lambda \le (\log n)^{O(\log n)}.$$

Let $\Delta : \binom{[0:n-1]}{d} \to \mathbb{R}$ be a function that is not identically zero and has an (L, q, λ)-group cover. Let $m = d(n-1)L^2$. Then there exists a d-variate polynomial ϕ with degree at most $O(\sqrt{m} \cdot \log^{4q+1} m)$ and $\|\phi\|_1 = \exp(O(\sqrt{m} \cdot \log^{4q+3} m))$ such that

$$\left|\sum_{0 \le t_1 < \cdots < t_d < n} \phi(t_1, \ldots, t_d) \cdot \Delta\Big(\{t_1, \ldots, t_d\}\Big)\right| \ge \frac{\|\Delta\|_\infty}{\exp(O(\sqrt{m} \cdot \log^{4q-1} m))}.$$

We note that the following lemma holds for any two distributions $\mathbf{X}, \mathbf{Y}$ over $\{0,1\}^n$ regardless of their support size.

Lemma 9.0.8. *Let $d, k \in [n]$ with $k \ge d$. Let $\mathbf{X}, \mathbf{Y}$ be distributions each supported over strings from $\{0,1\}^n$. Then for any string $c \in \{0,1\}^d$ and d-variate polynomial ϕ of degree at most $k - d$,*

$$\left|\sum_{0 \le t_1 < \cdots < t_d < n} \phi(t_1, \ldots, t_d) \cdot \Delta_{\mathbf{X},\mathbf{Y},c}\Big(\{t_1, \ldots, t_d\}\Big)\right| \le \|\phi\|_1 \cdot n^{O(k)} \cdot \|\mathsf{D}_k(\mathbf{X}) - \mathsf{D}_k(\mathbf{Y})\|_\infty.$$

Proof of Lemma 9.0.6. Let $\mathbf{X}$ and $\mathbf{Y}$ be two distributions each supported over at most ℓ strings from $\{0,1\}^n$. It then follows from Lemma 9.0.1 and Lemma 9.0.4 that there exists a string $c \in \{0,1\}^d$ with $d = \lfloor \log(2\ell) \rfloor$ such that $\Delta := \Delta_{\mathbf{X},\mathbf{Y},c}$ satisfies $\|\Delta\|_\infty \geq d_{\mathrm{TV}}(\mathbf{X},\mathbf{Y})/\ell^{O(\ell)}$ and has an (L,q,λ)-group cover for some $L \leq 2^{2d\ell}$, $q \leq \ell$, and $\lambda = \ell^{O(\ell)}$. As we assumed that $\ell \leq \log n$, both d and q are at most $\log n$ and $L, \lambda \leq \ell^{O(\ell)} \leq (\log n)^{O(\log n)}$ (so Lemma 9.0.7 applies).

Let $m = d(n-1)L^2$ and ϕ be the polynomial given in Lemma 9.0.7. Let $k^* = \deg(\phi)+d$ (we set $k = k^*$ in Lemma 9.0.8; the choice of k^* ensures that $\deg(\phi) \leq k^*-d$ as required in Lemma 9.0.8) with

$$k^* = O(\sqrt{m} \cdot \log^{4q+1} m) = \sqrt{n} \cdot (\log n)^{O(\ell)}.$$

Combining Lemma 9.0.7 and Lemma 9.0.8, we have

$$\frac{\|\Delta\|_\infty}{\exp(\sqrt{n} \cdot (\log n)^{O(\ell)})} \leq \exp\left(\sqrt{n} \cdot (\log n)^{O(\ell)}\right) \cdot n^{\sqrt{n} \cdot (\log n)^{O(\ell)}} \cdot \|\mathsf{D}_{k^*}(\mathbf{X}) - \mathsf{D}_{k^*}(\mathbf{Y})\|_\infty.$$

The lemma follows from the fact that $\|\Delta\|_\infty \geq d_{\mathrm{TV}}(\mathbf{X},\mathbf{Y})/\ell^{O(\ell)}$. □

Proof of Lemma 9.0.7

We defer this proof to [4].

Proof of Lemma 9.0.8

Let $\mathbf{X}$ and $\mathbf{Y}$ be two distributions each supported over strings from $\{0,1\}^n$.

Given $0 \leq j_1 < \cdots < j_d \leq k-1$, we use $g_{j_1,\ldots,j_d}$ to denote the following d-variate polynomial,

$$g_{j_1,\ldots,j_d}(t_1,\ldots,t_d) := \binom{t_1}{j_1} \cdot \binom{t_2 - t_1 - 1}{j_2 - j_1 - 1} \cdots \binom{t_d - t_{d-1} - 1}{j_d - j_{d-1} - 1} \cdot \binom{n - t_d - 1}{k - j_d - 1}. \quad (9.7)$$

To see the relevance of this polynomial to the k-deck, we note that given any $0 \leq t_1 < \cdots < t_d < n$ the quantity $g_{j_1,\ldots,j_d}(t_1,\ldots,t_d)$ is the number of ways to pick k indices from $[0:n-1]$ such that each t_i is the (j_i+1)th smallest index picked.

We first show that the following sum

$$\sum_{0 \leq t_1 < \cdots < t_d < n} g_{j_1,\ldots,j_d}(t_1,\ldots,t_d) \cdot \mathrm{restrict}\Big(\mathbf{X}, \{t_1,\ldots,t_d\}, c\Big) \quad (9.8)$$

can be written as a low-weight linear combination of entries of $\mathsf{D}_k(\mathbf{X})$.

Lemma 9.0.9. *For any $0 \le j_1 < \cdots < j_d \le k-1$ and any $c \in \{0,1\}^d$, the sum (9.8) can be written as a linear combination of entries of $\mathsf{D}_k(\mathbf{X})$ in which each coefficient is either 0 or $\binom{n}{k}$.*

Proof. Recall the combinatorial interpretation of $g_{j_1,\dots,j_d}(t_1,\dots,t_d)$ given after (9.7). We see that if we divide the sum in (9.8) by $\binom{n}{k}$, the result is precisely the probability that $(\boldsymbol{z}_{j_1},\dots,\boldsymbol{z}_{j_d}) = c$ when we draw $\boldsymbol{x} \sim \mathbf{X}$, draw a size-$k$ subset $\mathbf{T}$ of $[0:n-1]$ uniformly at random, and then set $\boldsymbol{z} = \boldsymbol{x}_{\mathbf{T}}$.

The latter probability can also be expressed using entries of $\mathsf{D}_k(\mathbf{X})$ as

$$\sum_{\substack{z\in\{0,1\}^k \\ (z_{j_1},\dots,z_{j_d})=c}} \Big(\mathsf{D}_k(\mathbf{X})\Big)_z,$$

as $(\mathsf{D}_k(\mathbf{X}))_z$ is the probability of $\boldsymbol{x}_{\mathbf{T}} = z$ with $\boldsymbol{x}$ and $\mathbf{T}$ drawn as above. This finishes the proof. □

Next we show that, for every monomial $t_1^{r_1}\cdots t_d^{r_d}$ of degree $r_1+\cdots+r_d \le k-d$, there exists a low-weight linear combination of polynomials $g_{j_1,\dots,j_d}$ that agrees with $t_1^{r_1}\cdots t_d^{r_d}$ over $t_1,\dots,t_d$ that satisfy $0 \le t_1 < \cdots < t_d < n$.

Lemma 9.0.10. *For any nonnegative integers $r_1,\dots,r_d$ with $r_1+\cdots+r_d \le k-d$, we have that*

$$t_1^{r_1}\cdots t_d^{r_d} = \sum_{0\le j_1<\cdots<j_d<k} w_{j_1,\dots,j_d}\cdot g_{j_1,\dots,j_d}(t_1,\dots,t_d), \quad \textit{for all } 0 \le t_1 < \cdots < t_d < n,$$

where the coefficients $w_{j_1,\dots,j_d}$ satisfy $\sum |w_{j_1,\dots,j_d}| \le k^{O(k)}$.

Before proving Lemma 9.0.10, we use Lemma 9.0.9 and Lemma 9.0.10 to prove Lemma 9.0.8.

Proof of Lemma 9.0.8. Combining Lemma 9.0.9 and Lemma 9.0.10, we have that

$$\begin{aligned}
&\sum_{0\le t_1<\cdots<t_d<n} t_1^{r_1}\cdots t_d^{r_d}\cdot \mathrm{restrict}\Big(\mathbf{X},\{t_1,\dots,t_d\},c\Big) \\
&= \sum_{0\le j_1<\cdots<j_d<k} w_{j_1,\dots,j_d} \sum_{0\le t_1<\cdots<t_d<n} g_{j_1,\dots,j_d}(t_1,\dots,t_d)\cdot \mathrm{restrict}\Big(\mathbf{X},\{t_1,\dots,t_d\},c\Big)
\end{aligned}$$

can be written as a linear combination of entries of $\mathsf{D}_k(\mathbf{X})$ in which each coefficient has magnitude at most $k^{O(k)}\cdot\binom{n}{k} = n^{O(k)}$. As a result, we have

$$\left|\sum_{0\le t_1<\cdots<t_d<n} t_1^{r_1}\cdots t_d^{r_d}\cdot \Delta_{\mathbf{X},\mathbf{Y},c}\Big(\{t_1,\dots,t_d\}\Big)\right| \le n^{O(k)}\cdot\|\mathsf{D}_k(\mathbf{X})-\mathsf{D}_k(\mathbf{Y})\|_\infty.$$

This finishes the proof of the lemma. □

Finally we prove Lemma 9.0.10. We follow a three-step approach. We say that a *quasimonomial* is a polynomial of the form

$$t_1^{\alpha_1} \cdot (t_2 - t_1 - 1)^{\alpha_2} \cdot (t_3 - t_2 - 1)^{\alpha_3} \cdots (t_d - t_{d-1} - 1)^{\alpha_d}$$

for some nonnegative integers $\alpha_1, \ldots, \alpha_d$; the degree of this quasimonomial is $\alpha_1 + \cdots + \alpha_d$. And we say that a *PBC* (*Product of Binomial Coefficients*) is a polynomial of the form

$$\binom{t_1}{\beta_1}\binom{t_2 - t_1 - 1}{\beta_2} \cdots \binom{t_d - t_{d-1} - 1}{\beta_d}$$

for some nonnegative integers $\beta_1, \ldots, \beta_d$; the degree of this PBC is $\beta_1 + \cdots + \beta_d$. We observe that, compared to PBCs, the polynomials $g_{j_1,\ldots,j_d}$ have an extra binomial coefficient that involves t_d at the end. The three steps of our approach are as follows:

- **First step:** Express each d-variable monomial $t_1^{r_1} \cdots t_d^{r_d}$ with $r_1 + \cdots + r_d \leq k - d$ as a low-weight linear combination of quasimonomials of degree at most $k - d$.
- **Second step:** Express each quasimonomial of degree at most $k - d$ as a low-weight linear combination of PBCs of degree at most $k - d$.
- **Third step:** Finally, express each PBC of degree at most $k - d$ as a low-weight linear combination of polynomials $g_{j_1,\ldots,j_d}$.

We elaborate on each of these three steps below. For each step, we bound the sum of magnitudes of coefficients in the linear combination.

First step. Consider the change of variables: $s_1 = t_1, s_2 = t_2 - t_1 - 1, \ldots, s_d = t_d - t_{d-1} - 1$. Then

$$t_1^{r_1} t_2^{r_2} \cdots t_d^{r_d} = s_1^{r_1}(s_2 + s_1 + 1)^{r_2} \cdots (s_d + s_{d-1} + \cdots + s_1 + d - 1)^{r_d}.$$

Each monomial of $s_1, \ldots, s_d$ on the RHS corresponds to a quasimonomial of degree at most $r_1 + \cdots + r_d \leq k - d$ so this gives us an expression of $t_1^{r_1} \cdots t_d^{r_d}$ as a linear combination of quasimonomials of degree at most $k - d$. Moreover, the sum of magnitudes of the coefficients is bounded by

$$3^{r_2} \cdot 5^{r_3} \cdots (2d-1)^{r_d} \leq (2d-1)^{\sum_{i=2}^{d} r_i} \leq k^{O(k)}. \tag{9.9}$$

Second step. We start with a one-dimensional version of the second step.

Claim 9.0.11. *For each $r \geq 0$ and $t \geq 0$, we have*

$$t^r = \sum_{\beta=0}^{r} \left(\sum_{i=0}^{\beta} (-1)^{\beta-i} \cdot \binom{\beta}{i} \cdot i^r \right) \binom{t}{\beta}. \tag{9.10}$$

Proof. We can write $t^r = \sum_{\beta=0}^{r} v_\beta \binom{t}{\beta}$ with $v \in \mathbb{R}^{r+1}$ by changing bases in the space of polynomials in t. Let P be the $(r+1) \times (r+1)$ Pascal matrix with $P_{i,j} = \binom{i}{j}$, and define $u \in \mathbb{R}^{r+1}$ by $u_i = i^r$. Then $u = Pv$ so $v = P^{-1}u$. By Theorem 2 of [11], we have

$$v_\beta = \sum_{i=0}^{\beta} (-1)^{\beta-i} \cdot \binom{\beta}{i} \cdot i^r$$

as desired. □

We use Claim 9.0.11 d times to re-express each of $t_1^{\alpha_1}$, $(t_2 - t_1 - 1)^{\alpha_2}, \ldots, (t_d - t_{d-1} - 1)^{\alpha_d}$ as a linear combination of binomial coefficients. As a result, when $0 \leq t_1 < \cdots < t_d < n$ we have

$$\begin{aligned}
&t_1^{\alpha_1} \cdot (t_2 - t_1 - 1)^{\alpha_2} \cdots (t_d - t_{d-1} - 1)^{\alpha_d} \\
&= \left(\sum_{\beta_1=0}^{\alpha_1} \left(\sum_{i_1=0}^{\beta_1} (-1)^{\beta_1 - i_1} \binom{\beta_1}{i_1} i_1^{\alpha_1} \right) \binom{t_1}{\beta_1} \right) \\
&\qquad \cdot \left(\sum_{\beta_2=0}^{\alpha_2} \left(\sum_{i_2=0}^{\beta_2} (-1)^{\beta_2 - i_2} \binom{\beta_2}{i_2} i_2^{\alpha_2} \right) \binom{t_2 - t_1 - 1}{\beta_2} \right) \\
&\qquad \cdots \left(\sum_{\beta_d=0}^{\alpha_d} \left(\sum_{i_d=0}^{\beta_d} (-1)^{\beta_d - i_d} \binom{\beta_d}{i_d} i_d^{\alpha_d} \right) \binom{t_d - t_{d-1} - 1}{\beta_d} \right) \\
&= \sum_{\beta_1, \ldots, \beta_d} c_{\beta_1, \ldots, \beta_d} \cdot \binom{t_1}{\beta_1} \binom{t_2 - t_1 - 1}{\beta_2} \cdots \binom{t_d - t_{d-1} - 1}{\beta_d}
\end{aligned}$$

for coefficients $c_{\beta_1, \ldots, \beta_d}$ that we will proceed to bound. Note that the final sum is over $0 \leq \beta_i \leq \alpha_i$, so this is a linear combination of PBCs of degree at most $\alpha_1 + \cdots + \alpha_d \leq k - d$.

Now we bound the sum of magnitudes of coefficients. For $0 \leq \beta \leq \alpha$ we have

$$\left| \sum_{i=0}^{\beta} (-1)^{\beta-i} \cdot \binom{\beta}{i} \cdot i^\alpha \right| \leq \sum_{i=0}^{\beta} \beta^i i^\alpha \leq \sum_{i=0}^{\beta} (\beta i)^\alpha \leq \beta^{O(\alpha)},$$

which implies (using $\alpha_1 + \ldots + \alpha_d \le k - d \le k$)

$$\sum_{\beta_1,\ldots,\beta_d} |c_{\beta_1,\ldots,\beta_d}| \le \sum_{\beta_1,\ldots,\beta_d} \beta_1^{O(\alpha_1)} \cdots \beta_d^{O(\alpha_d)} \le \sum_{\beta_1,\ldots,\beta_d} k^{O(k)} = k^{O(k)}. \tag{9.11}$$

Third step: The next claim gives an expression of a PBC as a linear combination of $g_{j_1,\ldots,j_d}$'s.

Claim 9.0.12 (Third step: d-variable combinatorial identity)**.** *Given any $0 \le t_1 < \cdots < t_d < n$ and any nonnegative integers $\beta_1, \ldots, \beta_d$ with $\beta_1 + \cdots + \beta_d \le k - d$, we have*

$$\sum_{0 \le j_1 < \cdots < j_d < k} g_{j_1,\ldots,j_d}(t_1,\ldots,t_d) \cdot \binom{j_1}{\beta_1}\binom{j_2 - j_1 - 1}{\beta_2} \cdots \binom{j_d - j_{d-1} - 1}{\beta_d}$$
$$= \binom{t_1}{\beta_1}\binom{t_2 - t_1 - 1}{\beta_2} \cdots \binom{t_d - t_{d-1} - 1}{\beta_d} \cdot \binom{n - \beta_1 - \cdots - \beta_d - d}{k - \beta_1 - \cdots - \beta_d - d}.$$

Proof. Assume that $0 \le t_1 < \cdots < t_d < n$. We first consider the following combinatorial experiment with n balls numbered $[0 : n-1]$: (1) *Mark* k of the n balls, including balls $t_1, \ldots, t_d$; (2) *Color red* $\beta_1 + \cdots + \beta_d + d$ of the k marked balls, including balls $t_1, \ldots, t_d$, in such a way that for each $i \in [d]$, the $(\beta_1 + \cdots + \beta_i + i)$-th red one is t_i (i.e., there are β_1 red balls before t_1, β_2 red balls between t_1 and t_2, ..., and β_d red balls between t_{d-1} and t_d). Below we count the total number of outcomes of this experiment (including which balls are marked and which balls are colored red) in two different ways to obtain the desired identity.

In the first way, we note that at the end of this experiment there are β_1 balls that are *marked and red* within the t_1 balls $\{0, \ldots, t_1 - 1\}$ (and also t_1 is *marked and red*), and for each $i \in [2 : d]$ there are β_i balls that are *marked and red* within the $t_i - t_{i-1} - 1$ balls $\{t_{i-1} + 1, \ldots, t_i - 1\}$ (and also t_i is *marked and red*); and there are $k - \beta_1 - \cdots - \beta_d - d$ other balls that are marked within the $n - \beta_1 - \cdots - \beta_d - d$ other balls. So the total number of outcomes of the experiment is precisely

$$\binom{t_1}{\beta_1} \cdot \binom{t_2 - t_1 - 1}{\beta_2} \cdots \binom{t_d - t_{d-1} - 1}{\beta_d} \cdot \binom{n - \beta_1 - \cdots - \beta_d - d}{k - \beta_1 - \cdots - \beta_d - d}.$$

We can also count the number of outcomes of the experiment in a different way, by viewing the experiment as being carried out as follows: (a) For some numbers $0 \le j_1 < \cdots < j_d < k$, *mark* k balls such that for each $i \in [d]$, the $(j_i + 1)$-th marked ball is t_i; note that as mentioned earlier, $g_{j_1,\ldots,j_d}(t_1, \ldots, t_d)$ is precisely the number of ways to do this. (b) Color red β_1 of the j_1 marked balls before ball t_1 (and also color red ball t_1; there are $\binom{j_1}{\beta_1}$ ways to do this), and for each $i \in [2 : d]$, color red β_i

of the $j_i - j_{i-1} - 1$ marked balls that lie in $\{t_{i-1}+1, \ldots, t_i - 1\}$ (and also color red t_i; there are $\binom{j_i - j_{i-1} - 1}{\beta_i}$ ways to do this). From this perspective, the total number of outcomes is

$$\sum_{0 \le j_1 < \cdots < j_d < k} g_{j_1,\ldots,j_d}(t_1, \ldots, t_d) \cdot \binom{j_1}{\beta_1}\binom{j_2 - j_1 - 1}{\beta_2} \cdots \binom{j_d - j_{d-1} - 1}{\beta_d},$$

and we have established the identity as desired. □

Observe that when $\beta_1 + \cdots + \beta_d \le k - d$, we have

$$\left|\binom{j_1}{\beta_1}\binom{j_2 - j_1 - 1}{\beta_2} \cdots \binom{j_d - j_{d-1} - 1}{\beta_d}\right| \le k^{\beta_1 + \cdots + \beta_d} \le k^{k-d},$$

which implies that the sum of magnitudes of coefficients in the linear combination is

$$\frac{\sum_{0 \le j_1 < \cdots < j_d < k} \left|\binom{j_1}{\beta_1}\binom{j_2 - j_1 - 1}{\beta_2} \cdots \binom{j_d - j_{d-1} - 1}{\beta_d}\right|}{\binom{n - \beta_1 - \cdots - \beta_d - d}{k - \beta_1 - \cdots - \beta_d - d}} \le \sum_{0 \le j_1 < \cdots < j_d < k} k^{k-d} \le k^k. \tag{9.12}$$

We can now combine the three steps to prove Lemma 9.0.10.

Proof of Lemma 9.0.10. We express $t_1^{r_1} \cdots t_d^{r_d}$ as a linear combination of polynomials $g_{j_1,\ldots,j_d}$ via the three steps as described above, with coefficients $w_{j_1,\ldots,j_d}$.

The sum of magnitudes of coefficients in the linear combination used in the First, Second, and Third steps are bounded from above using inequalities (9.9), (9.11), and (9.12) respectively. These bounds give us

$$\sum_{0 \le j_1 < \cdots < j_d < k} |w_{j_1,\ldots,j_d}| \le k^{O(k)} \cdot k^{O(k)} \cdot k^k \le k^{O(k)}$$

as desired. This finishes the proof of the lemma. □

Chapter 10

Lower bounds

Our main result in this section is Theorem 10.0.1, given below, which establishes a lower bound on the sample complexity of population recovery under the deletion channel which is exponential in the population size for a wide range of population sizes:

Theorem 10.0.1. *Fix any constant deletion probability $\delta \in (0,1)$. Suppose that A is an algorithm which, when run on i.i.d. samples drawn from a distribution $\mathrm{Del}_\delta(\mathbf{X})$ with $|\mathrm{supp}(\mathbf{X})| \leq 2\ell$, outputs a hypothesis $\tilde{\mathbf{X}}$ which satisfies $d_{\mathrm{TV}}(\mathbf{X}, \tilde{\mathbf{X}}) \leq 0.49$ with probability at least 0.51. Then A must use*

$$\frac{\Omega\left(n/\ell^2\right)^{\frac{\ell+1}{2}}}{\ell^{\frac{3}{2}}}$$

many samples.

If the population size upper bound 2ℓ is a constant this gives a lower bound of $\Omega(n^{(\ell+1)/2})$ samples, and for any $\ell < n^{0.499}$ this gives a lower bound of $n^{\Omega(\ell)}$.

For the rest of this section fix $\delta \in (0,1)$ and let ρ denote $1-\delta$. The high-level idea of the proof is as follows: We show that there exist two distributions $\mathbf{X}, \mathbf{Y}$ over $\{0,1\}^n$ which have disjoint supports, each of size at most 2ℓ, but satisfy

$$d_{\mathrm{TV}}(\mathrm{Del}_\delta(\mathbf{X}), \mathrm{Del}_\delta(\mathbf{Y})) = O\left(\frac{\ell^2}{n}\right)^{\frac{\ell+1}{2}} \cdot \ell^{\frac{3}{2}} \cdot (1-\delta) \tag{10.1}$$

which clearly implies Theorem 10.0.1.

For simplicity throughout this section we assume that n is odd, and we write m to denote $(n-1)/2$. The following notation will be useful: For $0 \leq i \leq 2\ell$ we write

e_{m+i} to denote the string $0^{m+i}10^{m-i}$. The two distributions $\mathbf{X}$ and $\mathbf{Y}$ that we consider will be supported on disjoint subsets of $\{e_{m+i}\}_{i\in[0:2\ell]}$ (and hence each distribution has support size at most $2\ell+1$, but in our proofs neither will have full support so their support size will be at most 2ℓ).

Notation and setup. For notational convenience, let $\mathrm{B}(r)$ denote the binomial distribution $\mathrm{Bin}(r,\rho)$.

Let S be a set of indices, π_S be a distribution over S, and $\{\mathbf{V}_i\}_{i\in S}$ be a set of random variables indexed by S. We write $\mathrm{Mix}(\pi_S;\{\mathbf{V}_i\}_{i\in S})$ to denote the mixture over $\{\mathbf{V}_i\}_{i\in S}$ with each $\mathbf{V}_i$ weighted by $\pi_S(i)$.

For conciseness we write $\mathbf{Z}_n$ to denote a random variable which is distributed according to the binomial distribution $\mathrm{B}(n)$. We recall the following convenient expression for the falling moments of the binomial distribution: for any $t=0,1,\ldots,$ we have

$$\mathbf{E}[\mathbf{Z}_n(\mathbf{Z}_n-1)\cdots(\mathbf{Z}_n-t)] = P_t(n), \qquad \text{where } P_t(n) = n(n-1)\cdots(n-t)\rho^{t+1}. \tag{10.2}$$

For completeness we include the derivation below:

$$\begin{aligned}
\mathbf{E}[\mathbf{Z}_n(\mathbf{Z}_n-1)\cdots(\mathbf{Z}_n-t)] &= \sum_{i=0}^{n} i(i-1)\cdots(i-t)\cdot\binom{n}{i}\rho^i(1-\rho)^{n-i}\\
&= \sum_{i=t+1}^{n}\frac{n!}{(n-i)!(i-t-1)!}\cdot\rho^i\cdot(1-\rho)^{n-i}\\
&= P_t(n)\cdot\sum_{j=0}^{n-t-1}\binom{n-t-1}{j}\rho^j(1-\rho)^{n-t-1-j}\\
&= P_t(n).
\end{aligned}$$

The key lemmas. The first main lemma makes precise the moment-matching property of π_S and π_T that we require:

Lemma 10.0.1 (Matching moments of mixtures of disjointly supported binomial distributions)**.** *Let $\ell \le O(\sqrt{n})$.*[1] *There are two disjoint subsets $S,T\subset[0:2\ell]$ and two distributions π_S,π_T supported on $\{e_{m+i}\}_{i\in S}$ and $\{e_{m+j}\}_{j\in T}$ respectively with the following property (which we call the "matching moment property"):*

Let $\tilde{\mathbf{D}}_S$ be a random variable whose distribution is the mixture of $\{\mathbf{Z}_{m+i}\}_{i\in S}$ in which distribution $\mathbf{Z}_{m+i}$ has mixing weight $\pi_S(e_{m+i})$, and likewise $\tilde{\mathbf{D}}_T$ be a random

[1] Note that if $\ell=\omega(\sqrt{n})$ then Theorem 10.0.1 holds trivially, so this assumption is without loss of generality.

variable whose distribution is the mixture of $\{\mathbf{Z}_{m+j}\}_{j\in T}$ *in which distribution* $\mathbf{Z}_{m+j}$ *has mixing weight* $\pi_T(e_{m+j})$*. Then the first* ℓ *moments of* $\tilde{\mathbf{D}}_S$ *and* $\tilde{\mathbf{D}}_T$ *match each other, i.e. for all* $t \in [\ell]$*, we have*

$$\mathbf{E}[(\tilde{\mathbf{D}}_S)^t] = \mathbf{E}[(\tilde{\mathbf{D}}_T)^t]. \tag{10.3}$$

The second main lemma (statement given in Lemma 10.1.1 below) gives the desired upper bound on total variation distance. To prove Theorem 10.0.1 it suffices to prove Lemmas 10.0.1 and 10.1.1.

Proof of Lemma 10.0.1

Proof. The proof is by a linear algebraic argument. Let $r = m + \ell$. Consider the mixtures

$$\tilde{\mathbf{D}}_S = \mathrm{Mix}(\{a_{|i|}\}_{i\in[-\ell:\ell]}; \{\mathbf{Z}_{r+i}\}_{i\in[-\ell:\ell]})$$

and

$$\tilde{\mathbf{D}}_T = \mathrm{Mix}(\{b_{|j|}\}_{j\in[-\ell:\ell]}; \{\mathbf{Z}_{r+j}\}_{j\in[-\ell:\ell]})$$

where all $a_i, b_i \in [0,1]$ and $\sum_{i=-\ell}^{\ell} a_{|i|} = \sum_{j=-\ell}^{\ell} b_{|j|} = 1$. Let $c_i = a_i - b_i$ for $0 \le i \le \ell$.

We will prove the existence of a non-trivial solution a_i, b_i (i.e., such that $a_i \neq b_i$ for some i) such that the following system holds:

$$\begin{aligned} \mathbf{E}[\tilde{\mathbf{D}}_S] &= \mathbf{E}[\tilde{\mathbf{D}}_T] \\ \mathbf{E}[\tilde{\mathbf{D}}_S(\tilde{\mathbf{D}}_S - 1)] &= \mathbf{E}[\tilde{\mathbf{D}}_T(\tilde{\mathbf{D}}_T - 1)] \\ &\cdots \\ \mathbf{E}[\tilde{\mathbf{D}}_S(\tilde{\mathbf{D}}_S - 1)\cdots(\tilde{\mathbf{D}}_S - \ell + 1)] &= \mathbf{E}[\tilde{\mathbf{D}}_T(\tilde{\mathbf{D}}_T - 1)\cdots(\tilde{\mathbf{D}}_T - \ell + 1)]. \end{aligned} \tag{10.4}$$

Observe that this is the same as requiring that $\mathbf{E}[\tilde{\mathbf{D}}_S^t] = \mathbf{E}[\tilde{\mathbf{D}}_T^t]$ for $t \le \ell$. In (10.4), we will be viewing $\mathbf{E}[Q(\tilde{\mathbf{D}}_S)]$ and $\mathbf{E}[Q(\tilde{\mathbf{D}}_T)]$ (for polynomials Q) as polynomials in n. We want the equations in (10.4) to hold as polynomial equalities.

Note that for $t \ge 0$, by (10.2) we can rewrite the condition $\mathbf{E}[\tilde{\mathbf{D}}_S(\mathbf{D}_S - 1)\cdots(\tilde{\mathbf{D}}_S - t)] = \mathbf{E}[\tilde{\mathbf{D}}_T(\tilde{\mathbf{D}}_T - 1)\cdots(\tilde{\mathbf{D}}_T - t)]$ as the condition

$$c_0 P_t(r) + \sum_{i=1}^{\ell} c_i \left(P_t(r+i) + P_t(r-i)\right) = 0, \tag{10.5}$$

viewing both sides as formal polynomials in r. Since $P_t(x)$ has degree $t+1$ in x, for $0 \le \ell \le t+1$ the coefficient of r^ℓ in the polynomial on the LHS of (10.5) is zero, and consequently we get a system of $t+2$ homogeneous linear equations in $c_0, c_1, \ldots, c_\ell$.

Naively, it seems that (10.4) gives us $2 + 3 + \cdots + \ell + 1 = \binom{\ell+2}{2} - 1$ many homogeneous linear equations, which is far more than the $\ell + 1$ variables $c_0, \ldots, c_\ell$ that are in play. At this point it is unclear that (10.4) necessarily has a nonzero solution in the c_i's. We will show that (10.4) is actually comprised of at most ℓ equations and hence it must have a nonzero solution.

Thus, to prove the existence of a non-trivial solution to (10.4) phrased in terms of $\tilde{\mathbf{D}}_S$ and $\tilde{\mathbf{D}}_T$, it suffices to prove the existence of a non-trivial solution to (10.4) phrased in terms of polynomial equalities.

To do this, we observe that equation (10.5) is also true when we replace r by $r+1$ and get the condition

$$c_0 P_t(r+1) + \sum_{i=1}^{\ell} c_i \left(P_t(r+1+i) + P_t(r+1-i)\right) = 0 \tag{10.6}$$

as a polynomial in r. (Note that the initial assumption $\ell \leq \Omega(n)$ still holds if we increase n.)

Observe that

$$P_t(r+1) = P_t(r) + \rho \cdot (t+1) P_{t-1}(r),$$

so if we subtract (10.5) from (10.6) and divide by $\rho(t+1)$, then we get the condition

$$c_0 P_{t-1}(r) + \sum_{i=1}^{\ell} c_i \left(P_{t-1}(r+i) + P_{t-1}(r-i)\right) = 0$$

as a polynomial in r. Since this is true for all t, then we have derived the condition $\mathbf{E}[\tilde{\mathbf{D}}_S(\tilde{\mathbf{D}}_S - 1)\cdots(\tilde{\mathbf{D}}_S - t + 1)] = 0$. It follows by induction that all of (10.4) follows from the condition $\mathbf{E}[\tilde{\mathbf{D}}_S(\tilde{\mathbf{D}}_S - 1)\cdots(\tilde{\mathbf{D}}_S - \ell + 1)] = 0$.

Thus we have reduced (10.4) to a system of $\ell + 1$ homogeneous linear equations over $\ell + 1$ variables, but the first equation (which comes from observing that the coefficient of r^ℓ in the LHS of (10.5) is 0) will be

$$2c_\ell + 2c_{\ell-1} + \cdots + 2c_1 + c_0 = 0 \tag{10.7}$$

and a second equation (which comes from observing that the coefficient of $r^{\ell-1}$ in the LHS of (10.5) is 0) will be

$$-\ell(\ell-1)c_\ell - \ell(\ell-1)c_{\ell-1} - \cdots - \ell(\ell-1)c_1 - (\ell/2)(\ell-1)c_0 = 0$$

because the coefficient of $r^{\ell-1}$ in $P_\ell(r)$ is $-(\ell/2)(\ell-1)$. This is just equation (10.7) times $-(\ell/2)(\ell-1)$. So there are actually at most ℓ distinct equations in $\ell + 1$ variables, and hence there is (at least) a line of non-trivial solutions in the c_i's.

Given a satisfying assignment to the c_i's, for each i with $c_i = 0$ we set $a_i = b_i = 0$. If $c_i > 0$, then we set $a_i = c_i$ and $b_i = 0$. If $c_i < 0$, then we set $a_i = 0$ and $b_i = -c_i$. Note that

$$0 = 2c_\ell + 2c_{\ell-1} + \cdots + 2c_1 + c_0 = 2a_\ell + 2a_{\ell-1} + \cdots + 2a_1 + a_0 - (2b_\ell + 2b_{\ell-1} + \cdots + 2b_1 + b_0)$$

and that by homogeneity, we can scale all the c_i's by any multiplicative constant and still get a valid solution to (10.4). We scale the c_i's so that $2a_\ell + 2a_{\ell-1} + \cdots + 2a_1 + a_0 = 1$. The above equation implies that $2b_\ell + 2b_{\ell-1} + \cdots + 2b_1 + b_0 = 1$ as well.

This results in the coefficients a_i and b_i satisfying (10.4) and $\tilde{\mathbf{D}}_S$ and $\tilde{\mathbf{D}}_T$ being valid distributions that are disjointly supported. Since the c_i's were non-trivial, then at least one coefficient c_i is non-zero and by equation (10.7), there exist coefficients c_j and c_k of opposite sign. Thus, both $\tilde{\mathbf{D}}_S$ and $\tilde{\mathbf{D}}_T$ have support sizes at most 2ℓ.

We take $\pi_S(e_{m+t}) = a_{|t-\ell|}$ and $\pi_T(e_{m+t}) = b_{|t-\ell|}$ to conclude the proof. □

We will use the following corollary of Lemma 10.0.1:

Corollary 10.0.2. *Let S, T, π_S, π_T be as in Lemma 10.0.1. Then for any polynomial p of degree at most ℓ, we have*

$$\sum_{i \in \mathbb{N}} \pi_S(e_{m+i}) p(m+i) = \sum_{j \in \mathbb{N}} \pi_S(e_{m+j}) p(m+j). \tag{10.8}$$

Proof. Equation (10.3) can be rewritten as

$$\sum_{i \in \mathbb{N}} \pi_S(e_{m+i}) \, \mathbf{E}[(\mathbf{Z}_{m+i})^t] = \sum_{j \in \mathbb{N}} \pi_S(e_{m+j}) \, \mathbf{E}[(\mathbf{Z}_{m+j})^t],$$

which holds for all $t \leq \ell$. This is equivalent to having equal falling moments, i.e. for all $t \in [\ell]$,

$$\sum_{i \in \mathbb{N}} \pi_S(e_{m+i}) \, \mathbf{E}[P_{t-1}(\mathbf{Z}_{m+i})] = \sum_{j \in \mathbb{N}} \pi_S(e_{m+j}) \, \mathbf{E}[P_{t-1}(\mathbf{Z}_{m+j})].$$

Indeed, for a random variable $\mathbf{Z}$, $\mathbf{E}[P_{t-1}(\mathbf{Z})]$ can be written as a linear combination of $1, \mathbf{E}[\mathbf{Z}], \mathbf{E}[\mathbf{Z}^2], \ldots, \mathbf{E}[\mathbf{Z}^t]$ and since $1, P_0(\mathbf{Z}), P_1(\mathbf{Z}), \ldots, P_{\ell-1}(\mathbf{Z})$ form a set of ℓ polynomials in $\mathbf{Z}$ with degrees $0, 1, 2, \ldots, \ell$, then they form a basis for polynomials in $\mathbf{Z}$ with degree at most ℓ.

By (10.2), this is in turn equivalent to having, for all $t \in [\ell]$,

$$\sum_{i \in \mathbb{N}} \pi_S(e_{m+i}) \cdot P_{t-1}(m+i) = \sum_{j \in \mathbb{N}} \pi_S(e_{m+j}) \cdot P_{t-1}(m+j),$$

which is in turn equivalent to (10.8) by the reasoning in the above paragraph. □

10.1 Total Variation Distance Upper Bound

We state Lemma 10.1.1 below. Informally, it says that if π_S, π_T have the matching moment property, then the variation distance between two corresponding mixtures of two-dimensional vector-valued random variables is small. (In the following, the notation $(\mathrm{B}(a), \mathrm{B}(b))$ stands for a vector-valued random variable in which the two coordinates are independently drawn from $\mathrm{B}(a)$ and $\mathrm{B}(b)$ respectively.)

Lemma 10.1.1. *Let $\mathbf{X}, \mathbf{Y}$ be two distributions with disjoint supports $\{e_{m+i}\}_{i\in S}$ and $\{e_{m+j}\}_{j\in T}$ respectively, where $S \cup T \subset [0 : 2\ell]$, with the matching moment property from Lemma 10.0.1 above. Then*

$$d_{\mathrm{TV}}(\mathrm{Del}_\delta(\mathbf{X}), \mathrm{Del}_\delta(\mathbf{Y})) \leq O\left(\frac{\ell^2}{n}\right)^{\frac{\ell+1}{2}} \cdot \ell^{\frac{3}{2}} \cdot (1-\delta). \tag{10.9}$$

Setup and useful results. Our proof of Lemma 10.1.1 is based on "moment-matching" results for Poisson binomial distributions which were proved by Roos [82] and subsequently used by Daskalakis and Papadimitriou [25]. Our approach is similar to the approach used in [25]. To state these results, recall that a *Poisson binomial distribution* (PBD) is a sum $\mathbf{U} = \mathbf{A}_1 + \cdots + \mathbf{A}_n$ of independent Bernoulli random variables (so each $\mathbf{A}_i$ is a random variable taking value 1 with some probability $p_i \in [0, 1]$ and taking value 0 with probability $1 - p_i$).

In [25], it is shown that if two PBDs satisfy some mild technical condition and have matching first ℓ moments, then they have total variation distance at most $2^{-\Omega(\ell)}$. We show that two mixtures of pairs of suitable binomially distributed variables that have matching first ℓ moments will have total variation distance at most $n^{-\Omega(\ell)}$.

We recall Theorem 1 of [25], which gives a "Krawtchouk expansion" for any Poisson binomial distribution. This provides an expression for the exact probability that the Poisson binomial distribution puts on any outcome in its support. (We state the theorem for PBDs which are a sum of n' many random variables, as when we apply it later it will be for such PBDs where $n' = m + \ell = (n-1)/2 + \ell$.)

Theorem 10.1.1 (Theorem 1 of [82], see also Theorem 7 of [25])**.** *Let $\mathbf{U} = \mathbf{A}_1 + \cdots + \mathbf{A}_{n'}$ be a Poisson binomial distribution in which each $\mathbf{A}_i$ takes value 1 with probability $p_i \in [0, 1]$. Then for all $r \in [n']$ and all $p \in [0, 1]$, we have*

$$\mathbf{Pr}[\mathbf{U} = r] = \sum_{t=0}^{n'} \alpha_t(p_1, \ldots, p_{n'}; p) \cdot \Delta^t B_{n'-t,p}(r), \tag{10.10}$$

where

- $\alpha_0(p_1, \ldots, p_{n'}; p) = 1$ *and for* $t \in [0 : n']$,

$$\alpha_t(p_1, \ldots, p_{n'}; p) := \sum_{1 \le u(1) < \cdots < u(t) \le n'} \prod_{r=1}^{t} (p_{u(r)} - p),$$

- *and for all* $t \in [0 : n']$,

$$\Delta^t B_{n'-t,p}(r) := \frac{(n'-t)!}{n'!} \cdot \frac{d^t}{dp^t} B_{n',p}(r),$$

where in the last expression $B_{n',p}(r)$ *denotes the value* $\binom{n'}{r} p^r (1-p)^{n'-r}$, *the probability that the binomial distribution* $\mathrm{Bin}(n', p)$ *puts on the outcome* r, *viewed as a function of* p.

We highlight the fact that $\Delta^t B_{n',p}(r)$ has no dependence on the parameters p_1, …, $p_{n'}$. This will be important for us later.

The following result, deduced from [82], is very useful in analyzing (10.10). It bounds each of the $n'+1$ summands in (10.10) which add up to $\mathbf{Pr}[\mathbf{U} = r]$.

Theorem 10.1.2. *Let* $(p_1, \ldots, p_{n'}) \in [0,1]^{n'}$, $p \in [0,1]$, *and* $\alpha_t(\cdot,\cdot)$ *be as in the statement of Theorem 10.1.1. Define*

$$\theta(p_1, \ldots, p_{n'}; p) := \frac{2\sum_{i=1}^{n'} (p_i - p)^2 + (\sum_{i=1}^{n'} (p_i - p))^2}{2n'p^2(1-p)^2}. \tag{10.11}$$

For $t \in [n']$,

$$|\alpha_t(p_1, \ldots, p_{n'}; p)| \cdot \|\Delta^t B_{n'-t,p}(\cdot)\|_1 \le \sqrt{e} \cdot \theta(p_1, \ldots, p_{n'}; p)^{\frac{t}{2}} t^{\frac{1}{4}} \tag{10.12}$$

where $\|\Delta^t B_{n'-t,p}(\cdot)\|_1$ *denotes the* 1*-norm of* $\Delta^t B_{n'-t,p}(\cdot)$ *when viewed as an* $(n'+1)$ *- dimensional vector, i.e.*

$$\|\Delta^t B_{n'-t,p}(\cdot)\|_1 := \sum_{r=0}^{n'} \left|\Delta^t B_{n'-t,p}(r)\right|$$

Proof. Inequality (30) in [82] gives

$$|\alpha_t(p_1, \ldots, p_{n'}; p)| \le p^{\frac{t}{2}} (1-p)^{\frac{t}{2}} \theta(p_1, \ldots, p_{n'}; p)^{\frac{t}{2}} \left(\frac{n'}{n'-t}\right)^{\frac{n'-t}{2}}$$

for $t \in [n']$.

Inequality (38) in [82] gives

$$\|\Delta^t B_{n'-t,p}(\cdot)\|_1 \leq \sqrt{e} \cdot t^{\frac{1}{4}} \left(\frac{n'-t}{n'}\right)^{\frac{n'-t}{2}} \left(\frac{t}{n'p(1-p)}\right)^{\frac{t}{2}}$$

for $t \in [n']$.

By multiplying the above two inequalities together we get the desired result because $t \leq n'$. □

For conciseness we now let $\mathbf{D}_S$ denote

$$\mathrm{Mix}(\pi_S; ((\mathrm{Bin}(m+i,\rho), \mathrm{Bin}(m-i,\rho)))_{i\in S})$$

where in each component two-dimensional distribution the two distributions $\mathrm{Bin}(m+i,\rho)$ and $\mathrm{Bin}(m-i,\rho)$ are independent, and similarly we let $\mathbf{D}_T$ denote

$$\mathrm{Mix}(\pi_T; ((\mathrm{Bin}(m+j,\rho), \mathrm{Bin}(m-j,\rho)))_{j\in T}).$$

In the rest of the proof we will argue that

$$d_{\mathrm{TV}}(\mathbf{D}_S, \mathbf{D}_T) \leq O\left(\frac{\ell^2}{n}\right)^{\frac{\ell+1}{2}} \cdot \ell^{\frac{3}{2}} \tag{10.13}$$

This establishes the claimed upper bound on $d_{\mathrm{TV}}(\mathrm{Del}_\delta(\mathbf{X}), \mathrm{Del}_\delta(\mathbf{Y}))$ given in (10.9). To see this, observe that for any outcome in $\mathrm{supp}(\mathbf{X})$ or $\mathrm{supp}(\mathbf{Y})$, with probability δ the one 1-coordinate is deleted under Del_δ (in which case the distributions resulting from $\mathrm{Del}_\delta(\mathbf{X})$ and $\mathrm{Del}_\delta(\mathbf{Y})$ are identical), and that with the remaining $1-\delta$ probability (when the one 1-coordinate is not deleted) there is an exact correspondence between $\mathrm{Del}_\delta(\mathbf{X})$ and $\mathbf{D}_S$ and between $\mathrm{Del}_\delta(\mathbf{Y})$ and $\mathbf{D}_T$.

For an index $c \leq n'$, let $v^{(c)}$ denote the n'-dimensional real vector whose first c values are ρ and whose remaining values are 0.

For $t, t' \in [0 : n']$ we define

$$C_{t,t'}(p) = \sum_{i\in\mathbb{N}} \pi_S(e_{m+i}) \cdot \alpha_t(v^{(m+i)}; p) \cdot \alpha_{t'}(v^{(m-i)}; p),$$
$$D_{t,t'}(p) = \sum_{j\in\mathbb{N}} \pi_T(e_{m+j}) \cdot \alpha_t(v^{(m+j)}; p) \cdot \alpha_{t'}(v^{(m-j)}; p).$$

The following lemma is crucial for us. Recall that $n' = m + \ell$.

Lemma 10.1.2. *Let π_S, π_T be as in the statement of Lemma 10.1.1. Then for any $p \in [0,1]$, the values $C_{t,t'}(p)$ and $D_{t,t'}(p)$ are identical for $t, t' \geq 0$ and $t + t' \leq \ell$.*

Proof. Let p be any value in $[0,1]$. If $t+t'=0$, then $t=t'=0$. Recalling that $\alpha_0(\cdot,\cdot)\equiv 1$ we have that

$$C_{0,0}(p)=\sum_{i\in\mathbb{N}}\pi_S(e_{m+i})=1=\sum_{j\in\mathbb{N}}\pi_T(e_{m+j})=D_{0,0}(p)$$

as desired.

For an integer k, let $\Gamma_k=\binom{m+k}{c}\binom{n'-m-k}{t-c}\binom{m-k}{c'}\binom{n'-m+k}{t'-c'}$.

For $t+t'\geq 1$, we observe that $\alpha_t(v^{(m+i)};p)\cdot\alpha_{t'}(v^{(m-i)};p)$ is composed of summands of the form $(\rho-p)^{c+c'}(-p)^{t+t'-c-c'}$ for $c\in[0,t], c'\in[0,t']$.

In particular, we have

$$C_{t,t'}(p)=\sum_{i\in\mathbb{N}}\pi_S(e_{m+i})\cdot\sum_{c=0}^{t}\sum_{c'=0}^{t'}\Gamma_i\left(\rho-p\right)^{c+c'}\left(-p\right)^{t+t'-c-c'},$$

in which each $\pi_S(e_{m+i})$ is multiplied by a polynomial in m of degree at most $t+t'\leq\ell$.

Similarly, we have

$$D_{t,t'}(p)=\sum_{j\in\mathbb{N}}\pi_T(e_{m+j})\cdot\sum_{c=0}^{t}\sum_{c'=0}^{t'}\Gamma_j\left(\rho-p\right)^{c+c'}\left(-p\right)^{t+t'-c-c'}$$

and by Corollary 10.0.2, we see that $C_{t,t'}(p)=D_{t,t'}(p)$. □

Now we proceed to prove Lemma 10.1.1. Our argument is similar to the proof of Theorem 3 in [25].

Let $p\in[0,1]$ and $r,s\in[0:n']$. We have

$$\begin{aligned}\mathbf{Pr}[\mathbf{D}_S=(r,s)]-\mathbf{Pr}[\mathbf{D}_T=(r,s)]&=\sum_{t,t'=0}^{n'}\Delta^t B_{n',p}(r)\cdot\Delta^{t'}B_{n',p}(s)\left(C_{t,t'}(p)-D_{t,t'}(p)\right)\\&=\sum_{t+t'>\ell}^{n'}\Delta^t B_{n',p}(r)\cdot\Delta^{t'}B_{n',p}(s)\left(C_{t,t'}(p)-D_{t,t'}(p)\right)\end{aligned}$$

where the two lines are by Theorem 10.1.1 and Lemma 10.1.2 respectively. As a result, for any $p\in[0,1]$ we have

$$\begin{aligned}d_{\mathrm{TV}}(\mathbf{D}_S,\mathbf{D}_T)&=\frac{1}{2}\sum_{r,s=0}^{n'}\left|\,\mathbf{Pr}[\mathbf{D}_S=(r,s)]-\mathbf{Pr}[\mathbf{D}_T=(r,s)]\right|\\&\leq\frac{1}{2}\sum_{t+t'>\ell}^{n'}\left|C_{t,t'}(p)-D_{t,t'}(p)\right|\cdot\|\Delta^t B_{n',p}(\cdot)\|_1\cdot\|\Delta^{t'}B_{n',p}(\cdot)\|_1\\&\leq\frac{1}{2}\sum_{t+t'>\ell}^{n'}\left(\left|C_{t,t'}(p)\right|+\left|D_{t,t'}(p)\right|\right)\cdot\|\Delta^t B_{n',p}(\cdot)\|_1\cdot\|\Delta^{t'}B_{n',p}(\cdot)\|_1\end{aligned}$$

and expanding out definitions gives

$$d_{\mathrm{TV}}(\mathbf{D}_S, \mathbf{D}_T) \le \frac{1}{2} \sum_{t+t'>\ell}^{n'} \left(\left| \sum_{i \in S} \pi_S(e_{m+i}) \cdot \alpha_t(v^{(m+i)}; p) \cdot \alpha_{t'}(v^{(m-i)}; p) \right| \right.$$
$$\left. + \left| \sum_{j \in T} \pi_T(e_{m+j}) \cdot \alpha_t(v^{(m+j)}; p) \cdot \alpha_{t'}(v^{(m-j)}; p) \right| \right) \cdot \|\Delta^t B_{n',p}(\cdot)\|_1 \cdot \|\Delta^{t'} B_{n',p}(\cdot)\|_1$$
$$\le \frac{1}{2} \sum_{i \in S} \pi_S(e_{m+i}) \sum_{t+t'>\ell}^{n'} \left| \alpha_t(v^{(m+i)}; p) \right| \left| \alpha_{t'}(v^{(m-i)}; p) \right| \cdot \|\Delta^t B_{n',p}(\cdot)\|_1 \cdot \|\Delta^{t'} B_{n',p}(\cdot)\|_1$$
$$+ \frac{1}{2} \sum_{j \in T} \pi_T(e_{m+j}) \sum_{t+t'>\ell}^{n'} \left| \alpha_t(v^{(m+j)}; p) \right| \left| \alpha_{t'}(v^{(m-j)}; p) \right| \cdot \|\Delta^t B_{n',p}(\cdot)\|_1 \|\Delta^{t'} B_{n',p}(\cdot)\|_1.$$

By Theorem 10.1.2,

$$d_{\mathrm{TV}}(\mathbf{D}_S, \mathbf{D}_T) \le \frac{e}{2} \sum_{i \in S} \pi_S(e_{m+i}) \sum_{t+t'>\ell}^{n'} \theta(v^{(m+i)}; \delta)^{\frac{t}{2}} \cdot \theta(v^{(m-i)}; \delta)^{\frac{t'}{2}} \cdot t^{\frac{1}{4}} t'^{\frac{1}{4}}$$
$$+ \frac{e}{2} \sum_{j \in T} \pi_T(e_{m+j}) \sum_{t+t'>\ell}^{n'} \theta(v^{(m+j)}; \delta)^{\frac{t}{2}} \cdot \theta(v^{(m-j)}; \delta)^{\frac{t'}{2}} \cdot t^{\frac{1}{4}} t'^{\frac{1}{4}}$$
$$\le \frac{e}{2\sqrt{2}} \sum_{i \in S} \pi_S(e_{m+i}) \sum_{t+t'>\ell}^{n'} \theta(v^{(m+i)}; \delta)^{\frac{t}{2}} \theta(v^{(m-i)}; \delta)^{\frac{t'}{2}} \sqrt{t+t'}$$
$$+ \frac{e}{2\sqrt{2}} \sum_{j \in T} \pi_T(e_{m+j}) \sum_{t+t'>\ell}^{n'} \theta(v^{(m+j)}; \delta)^{\frac{t}{2}} \theta(v^{(m-j)}; \delta)^{\frac{t'}{2}} \sqrt{t+t'}$$

where the second inequality can be deduced from the AM-GM inequality.

Fix any $i \in S, j \in T$. Let $p = \rho$ in Theorem 10.1.2. Since ρ is a constant in $(0, 1)$ we get that

$$\theta(v^{(m+i)}; \rho) = \frac{2(\ell - i)\rho^2 + (\ell - i)^2 \cdot \rho^4}{2(m + \ell)\rho^2(1 - \rho)^2} \le O\left(\frac{\ell^2}{n}\right).$$

and similarly

$$\theta(v^{(m-i)}; \rho), \theta(v^{(m \pm j)}; \rho) \le O\left(\ell^2/n\right)$$

because $\ell \le O(\sqrt{n})$ (and we may assume that $\ell \le O(\sqrt{n})$ since otherwise the total variation distance bound claimed in the lemma is trivial).

By choosing sufficiently large n and appropriate constants, we can upper bound the RHS by some $\theta < 1/2$. This gives

$$d_{\mathrm{TV}}(\mathbf{D}_S, \mathbf{D}_T) \leq O\left(\sum_{t+t'>\ell}^{n'} \theta^{\frac{t+t'}{2}} \sqrt{t+t'}\right) \leq O\left(\sum_{i>\ell}^{n'} \theta^{\frac{i}{2}} i^{\frac{3}{2}}\right) \leq O(\ell+1)^{\frac{-1}{2}} \sum_{i>\ell} \theta^{\frac{i}{2}} i^2$$

where the second inequality comes from the fact that there are $i+1$ pairs of non-negative integers t, t' that sum to i, and the third inequality comes from the fact that $i^{\frac{3}{2}} \leq (\ell+1)^{\frac{-1}{2}} i^2$ when $i > \ell$.

Observe that

$$\sum_{i>\ell} x^i i^2 = x \cdot \frac{d}{dx}(x \cdot \frac{d}{dx} \sum_{i>\ell} x^i) = x \cdot \frac{d}{dx}(x \cdot \frac{d}{dx} \frac{x^{\ell+1}}{1-x})$$

for $0 < x < 1$, so

$$\sum_{i>\ell} x^i i^2 = \frac{x^{\ell+1}}{(1-x)^3} \cdot (\ell^2(1-x)^2 + 2\ell(1-x) + 1 + x) \leq O(\ell+1)^2 \cdot \frac{x^{\ell+1}}{(1-x)^3} \leq O(\ell+1)^2 \cdot x^{\ell+1}$$

for $0 < x < 1/2$. This means

$$d_{\mathrm{TV}}(\mathbf{D}_S, \mathbf{D}_T) \leq O(\ell+1)^{\frac{3}{2}} \theta^{\frac{\ell+1}{2}} \leq O\left(\frac{\ell^2}{n}\right)^{\frac{\ell+1}{2}} \cdot \ell^{\frac{3}{2}}$$

giving (10.13) as desired and concluding the proof of Lemma 10.1.1.

Bibliography

[1] Per Austrin and Subhash Khot. A simple deterministic reduction for the gap minimum distance of code problem. In *International Colloquium on Automata, Languages, and Programming*, pages 474–485. Springer, 2011. 2.2, 2.2.3

[2] Arturs Backurs, Piotr Indyk, Ilya P. Razenshteyn, and David P. Woodruff. Nearly-optimal bounds for sparse recovery in generic norms, with applications to k-median sketching. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 318–337, 2016. 2.2, 2.2, 3

[3] Frank Ban, Vijay Bhattiprolu, Karl Bringmann, Pavel Kolev, Euiwoong Lee, and David P. Woodruff. A ptas for ℓ_p-low rank approximation. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 747–766, 2019. 1

[4] Frank Ban, Xi Chen, Adam Freilich, Rocco A. Servedio, and Sandip Sinha. Beyond trace reconstruction: Population recovery from the deletion channel. In *Foundations of Computer Science (FOCS), 2019 IEEE 56th Annual Symposium on*, 2019. 1, 7.1, 9

[5] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326. ACM, 2012. 2.2, 2.2, 2.2.1, 5, 5.3, 5.3, 5.3, 5.3.1, 5.3.2, 5.3.3, 5.3.4

[6] Lucia Batman, Russell Impagliazzo, Cody Murray, and Ramamohan Paturi. Finding heavy hitters from lossy or noisy data. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 16th International Workshop, APPROX 2013, and 17th International Workshop, RANDOM 2013, Berkeley, CA, USA, August 21-23, 2013. Proceedings*, pages 347–362, 2013. 1.2

[7] T. Batu, S. Kannan, S. Khanna, and A. McGregor. Reconstructing strings from random traces. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004*, pages 910–918, 2004. 1.2

[8] Aditya Bhaskara and Aravindan Vijayaraghavan. Approximating matrix p-norms. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 497–511. SIAM, 2011. 2.2

[9] Vijay Bhattiprolu, Mrinalkanti Ghoshi, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Inapproximability of matrix $p \to q$ norms. *Electronic Colloquium on Computational Complexity (ECCC)*, 2018. TR18-037. 2.2, 5.4

[10] Peter Borwein, Tamás Erdélyi, and Géza Kós. Littlewood-type problems on $[0, 1]$. *Proceedings of the London Mathematical Society*, 3(79):22–46, 1999. 7.1, 7.1, 7.1

[11] Robert Brawer and Magnus Pirovino. The linear algebra of the pascal matrix. *Linear Algebra and its Applications*, 174:13–23, 1992. 9

[12] Karl Bringmann, Pavel Kolev, and David P. Woodruff. Approximation algorithms for ℓ_0-low rank approximation. In *NIPS*, 2017. To appear. `http://arxiv.org/abs/1710.11253`. 1.1

[13] J. Paul Brooks and José H. Dulá. The ℓ_1-norm best-fit hyperplane problem. *Appl. Math. Lett.*, 26(1):51–55, 2013. 1.1

[14] J. Paul Brooks, José H. Dulá, and Edward L Boone. A pure ℓ_1-norm principal component analysis. *Computational statistics & data analysis*, 61:83–98, 2013.

[15] J. Paul Brooks and Sapan Jot. Pcal1: An implementation in r of three methods for ℓ_1-norm principal component analysis. *Optimization Online preprint*, 2012. 1.1

[16] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011. 1.1

[17] Kai-Yang Chiang, Cho-Jui Hsieh, and Inderjit S Dhillon. Robust principal component analysis with side information. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2291–2299, 2016. 1.1

[18] Flavio Chierichetti, Sreenivas Gollapudi, Ravi Kumar, Silvio Lattanzi, Rina Panigrahy, and David P. Woodruff. Algorithms for ℓ_p low-rank approximation. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 806–814, 2017. 1.1, 2.1, 2.2

[19] Christian Choffrut and Juhani Karhumäki. Combinatorics of words. In *Handbook of Formal Languages, Volume I*, pages 329–438. Springer, 1997. 7.1

[20] Kenneth L. Clarkson and David P. Woodruff. Numerical linear algebra in the streaming model. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 02, 2009*, pages 205–214, 2014. 4.1

[21] Kenneth L Clarkson and David P Woodruff. Input sparsity and hardness for robust subspace approximation. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 310–329. IEEE, 2015. 1.1, 2.2, 4.1

[22] Chen Dan, Kristoffer Arnsfelt Hansen, He Jiang, Liwei Wang, and Yuchen Zhou. On low rank approximation of binary matrices. *CoRR*, http://arxiv.org/abs/1511.01699, 2015. 1.1, 1.1

[23] Anirban Dasgupta, Petros Drineas, Boulos Harb, Ravi Kumar, and Michael W. Mahoney. Sampling algorithms and coresets for ℓ_p regression. *SIAM J. Comput.*, 38(5):19, 2009. 4.2, 4.4

[24] C. Daskalakis, I. Diakonikolas, and R. A. Servedio. Learning Poisson Binomial Distributions. *Algorithmica*, 72(1):316–357, 2015. 7.1

[25] Constantinos Daskalakis and Christos Papadimitriou. Sparse covers for sums of indicators. *Probability Theory & Related Fields*, 162:679–705, 2015. 7.1, 10.1, 10.1.1, 10.1

[26] A. De, M. Saks, and S. Tang. Noisy population recovery in polynomial time. Technical Report TR-16-026, Electronic Colloquium on Computational Complexity, 2016. To appear in FOCS 2016. 1.2

[27] Anindya De, Ryan O'Donnell, and Rocco A. Servedio. Optimal mean-based algorithms for trace reconstruction. In *Proceedings of the 49th ACM Symposium on Theory of Computing (STOC)*, pages 1047–1056, 2017. 1.2, 7, 7.1, 7.1, 7.1

[28] Anindya De, Ryan O'Donnell, and Rocco A. Servedio. Sharp bounds for population recovery. *CoRR*, abs/1703.01474, 2017. 1.2

[29] Amit Deshpande and Kasturi R. Varadarajan. Sampling-based dimension reduction for subspace approximation. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 641–650, 2007. 1.1

[30] Miroslav Dudík and Leonard Schulman. Reconstruction from subsequences. *Journal of Combinatorial Theory, Series A*, 103(2):337–348, 2003. 7.1

[31] Z. Dvir, A. Rao, A. Wigderson, and A. Yehudayoff. Restriction access. In *Innovations in Theoretical Computer Science*, pages 19–33, 2012. 1.2

[32] Dan Feldman and Michael Langberg. A unified framework for approximating and clustering data. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 569–578, 2011. 1.1

[33] Dan Feldman, Morteza Monemizadeh, Christian Sohler, and David P. Woodruff. Coresets and sketches for high dimensional subspace approximation problems. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 630–649, 2010. 1.1

[34] Nicolas Gillis and Stephen A. Vavasis. On the complexity of robust PCA and ℓ_1-norm low-rank matrix approximation. *CoRR*, http://arxiv.org/abs/1509.09236, 2015. 1.1

[35] Alexandre Grothendieck. *Résumé de la théorie métrique des produits tensoriels topologiques*. Soc. de Matemática de São Paulo, 1956. 2.2

[36] Venkatesan Guruswami, Prasad Raghavendra, Rishi Saket, and Yi Wu. Bypassing UGC from some optimal geometric inapproximability results. *ACM Transactions on Algorithms (TALG)*, 12(1):6, 2016. Conference version in SODA '12. 2.2, 5, 5.4, 5.4, 5.4, 5.4, 5.4.4, 5.4, 5.4.5

[37] Aram W Harrow and Ashley Montanaro. Testing product states, quantum Merlin-Arthur games and tensor optimization. *Journal of the ACM (JACM)*, 60(1):3, 2013. 2.2

[38] Lisa Hartung, Nina Holden, and Yuval Peres. Trace reconstruction with varying deletion probabilities. In *Proceedings of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2018, New Orleans, LA, USA, January 8-9, 2018.*, pages 54–61, 2018. 1.2, 1, 7

[39] Julien M Hendrickx and Alex Olshevsky. Matrix p-norms are NP-hard to approximate if $p \neq 1, 2, \infty$. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2802–2812, 2010. 2.2

[40] N. Holden and R. Lyons. Lower bounds for trace reconstruction. *CoRR*, abs/1808.02336, 2018. 1.2

[41] Nina Holden, Robin Pemantle, and Yuval Peres. Subpolynomial trace reconstruction for random strings and arbitrary deletion probability. *CoRR*, abs/1801.04783, 2018. 1.2, 7

[42] T. Holenstein, M. Mitzenmacher, R. Panigrahy, and U. Wieder. Trace reconstruction with constant deletion probability and related results. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008*, pages 389–398, 2008. 1.2, 7, 7.1, 7.1

[43] Russell Impagliazzo and Ramamohan Paturi. On the complexity of k-sat. *J. Comput. Syst. Sci.*, 62(2):367–375, 2001. 2.1

[44] Peng Jiang, Jiming Peng, Michael Heath, and Rui Yang. A clustering approach to constrained binary matrix factorization. In *Data Mining and Knowledge Discovery for Big Data*, pages 281–303. Springer, 2014. 1.1

[45] V. V. Kalashnik. Reconstruction of a word from its fragments. *Computational Mathematics and Computer Science (Vychislitel'naya matematika i vychislitel'naya tekhnika), Kharkov*, 4:56–57, 1973. 1.2, 7.1

[46] Daniel M. Kane, Jelani Nelson, and David P. Woodruff. An optimal algorithm for the distinct elements problem. In *Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2010, June 6-11, 2010, Indianapolis, Indiana, USA*, pages 41–52, 2010. 2.2, 4.2, 4.5

[47] Ravi Kannan and Santosh Vempala. Spectral algorithms. *Foundations and Trends in Theoretical Computer Science*, 4(3-4):157–288, 2009. 1.1

[48] Sampath Kannan and Andrew McGregor. More on reconstructing strings from random traces: Insertions and deletions. In *IEEE International Symposium on Information Theory*, pages 297–301, 2005. 1.2

[49] Qifa Ke and Takeo Kanade. Robust subspace computation using ℓ_1 norm. *Technical Report CMU-CS-03-172, Carnegie Mellon University, Pittsburgh, PA.*, 2003. 1.1

[50] Qifa Ke and Takeo Kanade. Robust ℓ_1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 739–746. IEEE, 2005. 1.1

[51] Subhash Khot. Hardness results for coloring 3-colorable 3-uniform hypergraphs. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 23–32. IEEE, 2002. 5.4

[52] Subhash A Khot and Nisheeth K Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 53–62. IEEE Computer Society, 2005. 5.3

[53] Eunwoo Kim, Minsik Lee, Chong-Ho Choi, Nojun Kwak, and Songhwai Oh. Efficient-norm-based low-rank matrix approximations for large-scale problems using alternating rectified gradient method. *IEEE transactions on neural networks and learning systems*, 26(2):237–251, 2015. 1.1

[54] Guy Kindler, Assaf Naor, and Gideon Schechtman. The UGC hardness threshold of the Lp Grothendieck problem. *Mathematics of Operations Research*, 35(2):267–283, 2010. Conference version in SODA '08. 5.4.3

[55] Ilia Krasikov and Yehuda Roditty. On a reconstruction problem for sequences,. *Journal of Combinatorial Theory, Series A*, 77(2):344–348, 1997. 7.1, 7.1, 7.1, 7.1

[56] Nojun Kwak. Principal component analysis based on ℓ_1-norm maximization. *IEEE transactions on pattern analysis and machine intelligence*, 30(9):1672–1680, 2008. 1.1

[57] Vladimir Levenshtein. Efficient reconstruction of sequences. *IEEE Transactions on Information Theory*, 47(1):2–22, 2001. 1.2

[58] Vladimir Levenshtein. Efficient reconstruction of sequences from their subsequences or supersequences. *Journal of Combinatorial Theory Series A*, 93(2):310–332, 2001. 1.2

[59] S. Lovett and J. Zhang. Improved Noisy Population Recovery, and Reverse Bonami-Beckner Inequality for Sparse Functions. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 137–142, 2015. 1.2

[60] Michael W. Mahoney. Randomized algorithms for matrices and data. *Foundations and Trends in Machine Learning*, 3(2):123–224, 2011. 1.1

[61] Pasin Manurangsi. Inapproximability of maximum biclique problems, minimum k-cut and densest at-least-k-subgraph from the small set expansion hypothesis. *Algorithms*, 11(1):10, 2018. 2.1, 5, 5.3

[62] Bennet Manvel, Aaron Meyerowitz, Allen Schwenk, Ken Smith, and Paul Stockmeyer. Reconstruction of sequences. *Discrete Mathematics*, 94(3):209–219, 1991. 7.1

[63] P. P. Markopoulos, S. Kundu, S. Chamadia, and D. A. Pados. Efficient ℓ_1-Norm Principal-Component Analysis via Bit Flipping. *ArXiv e-prints*, 2016. 1.1

[64] Panos P. Markopoulos, George N. Karystinos, and Dimitrios A. Pados. Some options for ℓ_1-subspace signal processing. In *ISWCS 2013, The Tenth International Symposium on Wireless Communication Systems, Ilmenau, TU Ilmenau, Germany, August 27-30, 2013*, pages 1–5, 2013.

[65] Panos P. Markopoulos, George N. Karystinos, and Dimitrios A. Pados. Optimal algorithms for ℓ_1-subspace signal processing. *IEEE Trans. Signal Processing*, 62(19):5046–5058, 2014. 1.1

[66] Andrew McGregor, Eric Price, and Sofya Vorotnikova. Trace reconstruction revisited. In *Proceedings of the 22nd Annual European Symposium on Algorithms*, pages 689–700, 2014. 1.2, 7.1, 7.1

[67] Deyu Meng, Zongben Xu, Lei Zhang, and Ji Zhao. A cyclic weighted median method for ℓ_1 low-rank matrix factorization with missing entries. In *AAAI*, volume 4, page 6, 2013. 1.1

[68] Xiangrui Meng and Michael W. Mahoney. Low-distortion subspace embeddings in input-sparsity time and applications to robust linear regression. In *Proceedings*

of the 45th Annual ACM Symposium on Theory of Computing, STOC 2013, Palo Alto, CA, USA, June 01 - 04, 2013, pages 91–100, 2013. 3

[69] Pauli Miettinen. Matrix decomposition methods for data mining: Computational complexity and algorithms. PhD Thesis, University of Helsinki, Finland, 2009. 1.1

[70] Ankur Moitra and Michael E. Saks. A polynomial time algorithm for lossy population recovery. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 110–116, 2013. 1.2

[71] Fedor Nazarov and Yuval Peres. Trace reconstruction with $\exp(o(n^{1/3}))$ samples. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017*, pages 1042–1046, 2017. 1.2, 7, 7.1, 7.1, 7.1

[72] Praneeth Netrapalli, UN Niranjan, Sujay Sanghavi, Animashree Anandkumar, and Prateek Jain. Non-convex robust pca. In *Advances in Neural Information Processing Systems*, pages 1107–1115, 2014. 1.1

[73] Feiping Nie, Jianjun Yuan, and Heng Huang. Optimal mean robust principal component analysis. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1062–1070, 2014. 1.1

[74] Lee Organick, Siena Dumas Ang, Yuan-Jyue Chen, Randolph Lopez, Sergey Yekhanin, Konstantin Makarychev, Miklos Z Racz, Govinda Kamath, Parikshit Gopalan, Bichlien Nguyen, et al. Random access in large-scale dna data storage. *Nature biotechnology*, 36(3):242, 2018. 1.2

[75] Young Woong Park and Diego Klabjan. Iteratively reweighted least squares algorithms for l1-norm principal component analysis. In *IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain*, pages 430–438, 2016. 1.1

[76] Yuval Peres and Alex Zhai. Average-case reconstruction for the deletion channel: Subpolynomially many traces suffice. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 228–239, 2017. 1.2, 7

[77] Yury Polyanskiy, Ananda Theertha Suresh, and Yihong Wu. Sample complexity of population recovery. In *Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017*, pages 1589–1618, 2017. 1.2

[78] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the Forty-second ACM Symposium on Theory of Computing*, STOC '10, pages 755–764, 2010. 2.1, 2.1, 5, 5.3

[79] Prasad Raghavendra, David Steurer, and Prasad Tetali. Approximations for the isoperimetric and spectral profile of graphs and related parameters. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 631–640. ACM, 2010. 5.3

[80] Prasad Raghavendra, David Steurer, and Madhur Tulsiani. Reductions between expansion problems. In *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*, pages 64–73. IEEE, 2012. 2.1, 5, 5.3

[81] Ilya P. Razenshteyn, Zhao Song, and David P. Woodruff. Weighted low rank approximations with provable guarantees. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 250–263, 2016. 2.2, 6.2

[82] B. Roos. Binomial approximation to the Poisson binomial distribution: The Krawtchouk expansion. *Theory Probab. Appl.*, 45:328–344, 2000. 7.1, 10.1, 10.1.1, 10.1, 10.1

[83] Alexander Scott. Reconstructing sequences. *Discrete Mathematics*, 175(1):231–238, 1997. 7.1, 7.1

[84] Bao-Hong Shen, Shuiwang Ji, and Jieping Ye. Mining discrete patterns via binary matrix factorization. In *KDD*, pages 757–766, 2009. 1.1

[85] Nariankadu D. Shyamalkumar and Kasturi R. Varadarajan. Efficient subspace approximation algorithms. *Discrete & Computational Geometry*, 47(1):44–63, 2012. 1.1

[86] Zhao Song, David P. Woodruff, and Peilin Zhong. Low rank approximation with entrywise l_1-norm error. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 688–701, 2017. 1.1, 2.1, 2.2

[87] Daureen Steinberg. Computation of matrix norms with applications to robust optimization. *Research thesis, Technion-Israel University of Technology*, 2005. 2.2

[88] David Steurer. Subexponential algorithms for d-to-1 two-prover games and for certifying almost perfect expansion. *Manuscript*, 2010. 5.3

[89] Krishnamurthy Viswanathan and Ram Swaminathan. Improved string reconstruction over insertion-deletion channels. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 399–408, 2008. 1.2

[90] A. Wigderson and A. Yehudayoff. Population recovery and partial identification. *Machine Learning*, 102(1):29–56, 2016. Preliminary version in FOCS 2012. 1.2

[91] David P. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1-2):1–157, 2014. 1.1, 2.2

[92] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, pages 2080–2088, 2009. 1.1

[93] S.M. Hossein Tabatabaei Yazdi, Ryan Gabrys, and Olgica Milenkovic. Portable and error-free DNA-based data storage. *Scientific Reports*, 7(1):5011, 2017. 1.2

[94] Huishuai Zhang, Yi Zhou, and Yingbin Liang. Analysis of robust pca via local incoherence. In *Advances in Neural Information Processing Systems*, pages 1819–1827, 2015. 1.1

[95] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi. Practical low-rank matrix approximation under robust ℓ_1-norm. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, pages 1410–1417, 2012. 1.1