Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 116 (2018)

Existence of Simple Extractors

Xue Chen[*]
xchen@cs.utexas.edu
The University of Texas at Austin

David Zuckerman[*]
diz@cs.utexas.edu
The University of Texas at Austin

August 31, 2018

Abstract

We show that a small subset of seeds of any strong extractor also gives a strong extractor with similar parameters when the number of output bits is a constant. Specifically, if $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is a strong (k,ϵ)-extractor, then for at least 99% of choices of $\tilde{O}(n \cdot 2^m/\epsilon^2)$ seeds, Ext restricted to these seeds is a $(k, 3\epsilon)$-extractor. Note that the degree of this restricted extractor is essentially optimal for $m = O(1)$. By combining this with the Leftover Hash Lemma, we deduce that there are strong extractors outputting a constant number of bits with essentially optimal degree where each seed is a linear function, or even a Toeplitz matrix, or a simply-implementable hash function. Although linear extractors were known, such as the one by Trevisan [Tre01], it didn't have close to optimal degree (although it did output more bits), and it wasn't known that most sets of linear functions give extractors.

While a simple application of the basic probabilistic method shows the existence of ordinary strong extractors, this approach fails to show the existence of the restricted extractors we seek, or even linear extractors. We therefore adopt a more sophisticated approach, using chaining as used by Rudra and Wootters [RW14] and others, combined with the Beck-Fiala theorem from discrepancy theory.

[*] Supported by NSF Grant CCF-1526952, NSF Grant CCF-1705028, and a Simons Investigator Award (#409864, David Zuckerman). The authors began this research while attending the Pseudorandomness program at the Simons Institute for Theoretical Computer Science in Berkeley, CA.

ISSN 1433-8092

1 Introduction

Randomness plays a vital role in computer science, both in theory and in practice. For example, randomness is provably necessary for many tasks in distributed computing and cryptography. At the same time, random sources from the real world are often biased and defective. A randomness extractor is an efficient algorithm that converts a "weak random source" into an almost uniform distribution. As is standard, we model a weak random source as a probability distribution with min-entropy.

Definition 1.1 *The min-entropy of a random variable X is*

$$H_\infty(X) = \min_{x \in \mathsf{supp}(X)} \log_2 \frac{1}{\Pr[X = x]}.$$

It is impossible to construct a deterministic randomness extractor for all sources of min-entropy k [SV86], even if k is as large as $n-1$. Therefore a seeded extractor also takes as input an additional independent uniform random string, called a seed, to guarantee that the output is close to uniform [NZ96]. We first define the distance measure between probability distributions.

Definition 1.2 (Statistical distance) *For any $d \in \mathbb{N}^+$, let U_d denote the uniform distribution over $\{0,1\}^d$. For two random variables W and Z with the same support, let $\|W - Z\|$ denote the statistical (or variation) distance*

$$\|W - Z\| = \max_{T \subseteq \mathsf{supp}(W)} \Big| \Pr_{w \sim W}[w \in T] - \Pr_{z \sim Z}[z \in T] \Big|.$$

Next we define extractors and strong extractors.

Definition 1.3 (Extractors and strong extractors) *A function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is a (k,ϵ)-extractor if for every source X with min-entropy k and an independent uniform distribution Y on $\{0,1\}^t$,*

$$\|\mathsf{Ext}(X,Y) - U_m\| \leq \epsilon.$$

It is a strong (k,ϵ)-extractor if in addition, it satisfies $\Big\| \big(\mathsf{Ext}(X,Y), Y\big) - \big(U_m, Y\big) \Big\| \leq \epsilon$.

Extractors and its strong variant have been studied extensively in computer science and have found numerous applications in seemingly unrelated areas beyond their original motivation (see the survey of Shaltiel [Sha02]).

Most known extractors are sophisticated and complicated to implement in practice. This raises natural questions — are there simple constructions like linear transformations and even simpler ones of Toeplitz matrices? Are there extractors with good parameters and efficient implementations?

We explore these questions in the context of the extractor degree. We call 2^t the degree of Ext, because when Ext is viewed as a bipartite graph on $\{0,1\}^n \cup \{0,1\}^m$, its left degree is 2^t. Often the degree 2^t is of more interest than the seed length t. For example, in the list-decoding view [TZ04], it is the length of the code. It is well known that the optimal degree of (k,ϵ)-extractors is $\Theta(\frac{n-k}{\epsilon^2})$, where the upper bound is from the probabilistic method and the lower bound was shown by Radhakrishnan and Ta-Shma [RT00].

Minimizing the degree of an extractor is crucial for many applications. For example, when viewed as a sampler, the degree is equivalent to the number of samples. When we simulate a probabilistic algorithm using weak random sources, a straightforward approach is to run the algorithm on all outputs of an extractor (cycling over all 2^t seeds) and take the majority vote. At the same time, explicit constructions with an optimal degree, even for constant error [Zuc07], have a variety of applications in theoretical computer science such as hardness of inapproximability and constructing almost optimal two source extractors and Ramsey graphs [BDT17].

In this work, we present a probabilistic construction to improve the degree of any given strong extractor. We show that it improves the degree of any strong extractor to almost optimal while keeping almost the same parameters of min entropy and error, at least when outputting few bits. Given an extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, we sample a few seeds from $\{0,1\}^t$ and consider the new extractor, called a restricted extractor, constituted by these seeds.

Definition 1.4 *Given an extractor* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ *and a sequence of seeds* $(y_1, \ldots, y_D)$ *where each* $y_i \in \{0,1\}^t$*, we define the restricted extractor* $\mathsf{Ext}_{(y_1,\ldots,y_D)}$ *to be* Ext *restricted to the domain* $\{0,1\}^n \times [D]$ *where* $\mathsf{Ext}_{(y_1,\ldots,y_D)}(x,i) = \mathsf{Ext}(x, y_i)$*.*

Our main result is that given any strong (k,ϵ)-extractor Ext, most restrictions with a quasi-linear degree $\tilde{O}(\frac{n}{\epsilon^2})$ from Ext are strong $(k, 3\epsilon)$-extractors for a constant number of output bits, despite the degree of Ext.

Theorem 1.5 *There exists a universal constant C such that given any strong (k,ϵ)-extractor* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$*, for* $D = C \cdot \frac{n \cdot 2^m}{\epsilon^2} \cdot \log^2 \frac{n \cdot 2^m}{\epsilon}$ *random seeds* $y_1, \ldots, y_D \in \{0,1\}^t$*,* $\mathsf{Ext}_{(y_1,\ldots,y_D)}$ *is a strong $(k, 3\epsilon)$-extractor with probability 0.99.*

Note that extractors with $D = 2^m$ are trivial, as we can set $\mathsf{Ext}(x,y) = y$. We also observe that any restriction of this trivial extractor requires $D \geq (1-\epsilon)2^m$ to obtain error ϵ. Though the dependency 2^m is necessary for extractor restrictions, it may not be necessary for strong extractors.

For $m = O(1)$, Theorem 1.5 improves the upper bound on the degree to $\tilde{O}(\frac{n}{\epsilon^2})$ for a broad range of strong extractors including strong AC^0 extractors and sparse extractors, based on known constructions [GVW15, CL16b, BG13]. We focus on its applications to simple strong extractors of linear transformations and Toeplitz matrices and extremely efficient strong extractors. These results follow because almost universal hash families are good strong extractors (by the Leftover Hash Lemma [ILL89]) and the above functions form almost universal hash families.

Definition 1.6 (Carter and Wegman [CW79]) *Let H be a family of functions mapping $\{0,1\}^n$ to $\{0,1\}^m$. H is universal if*

$$\forall x, y \in \{0,1\}^n (x \neq y), \Pr_{h \sim H}[h(x) = h(y)] \leq 2^{-m}.$$

Moreover, we say H is almost universal if,

$$\forall x, y \in \{0,1\}^n (x \neq y), \Pr_{h \sim H}[h(x) = h(y)] \leq 2^{-m} + 2^{-n}.$$

1.1 Applications

We discuss the application of Theorem 1.5 to almost-universal hash families in this section. Plugging the strong extractors of all linear transformations and Toeplitz matrices [ILL89] in Theorem 1.5, our result indicates that most strong extractors constituted by a quasi-linear number $\tilde{O}(\frac{n}{\epsilon^2})$ of linear transformations or Toeplitz matrices keep nearly the same parameters of the min-entropy and error, for a constant number of output bits. We treat the subset $\{0, 1, 2, \ldots, 2^n - 1\}$ the same as $\{0,1\}^n$ and $\mathbb{F}_2^n$ in this work.

Corollary 1.7 *There exists a universal constant C such that for any integers n, m, k, and $\epsilon > 0$ with $k \geq m + 2\log\frac{1}{\epsilon}$,* $\mathsf{Ext}_{(A_1,\ldots,A_D)}$ *with* $D = C \cdot \frac{n \cdot 2^m}{\epsilon^2} \cdot \log^2 \frac{n \cdot 2^m}{\epsilon}$ *random matrices* $A_1, \ldots, A_D \in \mathbb{F}_2^{m \times n}$*, mapping from* $\mathbb{F}_2^n \times [D]$ *to* $\mathbb{F}_2^m$ *as* $\mathsf{Ext}_{(A_1,\ldots,A_D)}(x,i) = A_i \cdot x$*, is a strong $(k, 3\epsilon)$-extractor with probability 0.99.*

Moreover, the same holds for D random Toeplitz matrices $A_1, \ldots, A_D \in \mathbb{F}_2^{n \times m}$*.*

Then we consider strong extractors from almost universal hash families, which have efficient implementations and wide applications in practice. We describe a few examples of almost universal hash families with efficient implementations.

1. Linear Congruential Hash by Carter and Wegman [CW79]: for any n and m, let p be a prime $> 2^n$ and $\mathcal{H}_1 = \big\{h_{a,b} | a, b \in \{0, 1, \ldots, p-1\}\big\}$ be the hash family defined as $h_{a,b}(x) = \big((ax+b) \mod p\big) \mod 2^m$ for every $x \in \{0, 1, \ldots, 2^n - 1\}$.

2. Multiplicative Universal Hash by Dietzfelbinger et al. [DHKP97] and Woelfel [Woe99]: for any n and m, let $\mathcal{H}_2 = \big\{h_{a,b} | a \in \{1, 3, 5, \ldots, 2^n - 1\}, b \in \{0, 1, \ldots, 2^{n-m} - 1\big\}$ be the hash family mapping $\{0, 1, \ldots, 2^n - 1\}$ to $\{0, 1, \cdots, 2^m - 1\}$ that first calculates $(ax + b)$ modulo 2^n then takes the high order m bits as the hash value. In C-code, this hash function could be implemented as $h_{a,b}(x) = (a * x + b) >> (n - m)$ when $n = 64$.

3. Shift Register Hash by Vazirani [Vaz87]: let p be a prime such that 2 is a generator modulo p and $a^{(i)}$ denote the ith shift of a string $a \in \mathbb{F}_2^n$, i.e., $a^{(i)} = a_{i+1} a_{i+2} \cdots a_n a_1 \cdots a_i$. For $n = p - 1$ and any $m \leq n$, let $\mathcal{H}_3 = \big\{h_a | a \in \mathbb{F}_2^p\big\}$ be the hash family mapping $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ as $\big(\langle a, 1 \circ x\rangle, \langle a^{(1)}, 1 \circ x\rangle, \ldots, \langle a^{(m-1)}, 1 \circ x\rangle\big)$, where $\langle w, z\rangle$ denotes the inner product of $w, z \in \mathbb{F}_2^p$ in $\mathbb{F}_2$.

Because all these hash families are almost universal, by the Leftover Hash Lemma [ILL89], $\mathsf{Ext}(x, y) = h_y(x)$ is a strong extractor for all hash functions h_y in one family. Plugging these extractors in Theorem 1.5, we obtain strong extractors with almost optimal degree and efficient implementations for a constant number of output bits.

Corollary 1.8 *Let $\mathcal{H}$ be any almost universal hash family mapping $\{0,1\}^n$ to $\{0,1\}^m$. There exists a universal constant C such that for any integer k and $\epsilon > 0$ with $k \geq m + 2\log\frac{1}{\epsilon}$, $\mathsf{Ext}_{(h_1,\ldots,h_D)}$ with $D = C \cdot \frac{n\cdot 2^m}{\epsilon^2} \cdot \log^2 \frac{n\cdot 2^m}{\epsilon}$ random hash functions $h_1, \ldots, h_D \sim \mathcal{H}$, defined as $\mathsf{Ext}_{(h_1,\ldots,h_D)}(x, i) = h_i(x)$, is a strong $(k, 3\epsilon)$-extractor with probability 0.99.*

1.2 Our techniques

For ease of exposition, let $m = O(1)$ in this discussion. Because there are $\binom{2^n}{2^k}$ subsets of size 2^k, the simple probabilistic method using the Chernoff bound and the union bound only provides an upper bound $O(\frac{(n-k)2^k}{\epsilon^2})$. Our upper bound is a combination of discrepancy theory and tools from high dimensional probability. Recently, Cheraghchi et al. [CGV13] and Wootters [Woo13] have successfully applied techniques from high dimension probability [LT91] such as symmetrization, Gaussianization, and chaining arguments to coding theory, which showed random linear codes have nearly the same list-decoding capacity as random codes. In a recent breakthrough, Rudra and Wootters [RW14] proved that a random puncturing of any list-decodable code has near-optimal list-decoding rate using a chaining argument. In this work, we apply the tools introduced by Rudelson and Vershynin [RV08], Cheraghchi et al. [CGV13], Wootters [Woo13], and Rudra and Wootters [RW14] to the setting of extractors. We use techniques from discrepancy theory to obtain an almost tight degree $\tilde{O}(\frac{n}{\epsilon^2})$, which shaves a factor of $\mathsf{poly}(k)$ in the chaining argument of Rudra and Wootters [RW14].

1.3 Implications for list decodable codes

Strong extractors outputting one bit are equivalent to list-decodable binary codes [TZ04], so our results may be reformulated in the coding theoretic setting. Restricting an extractor corresponds to puncturing a code.

Our main result implies that randomly puncturing any almost-balanced binary code can improve the rate to $\Omega(\frac{\epsilon^2}{\log^2 \frac{1}{\epsilon}})$ while keeping the list size at $O(\frac{1}{\epsilon^2})$ for radius $(\frac{1}{2} - \epsilon)$. We state the result for any alphabet $[q] = \{0, 1, \ldots, q-1\}$ and follow the notation of strong extractors by choosing $q = 2^m$ and letting 2^t denote the length of the encoding.

Definition 1.9 *Let $C : \{0,1\}^n \to [q]^{2^t}$ be a code with relative distance $\frac{1}{q} - \epsilon^2$. Given D coordinates $y_1, \ldots, y_D \in [2^t]$, let $C_{(y_1,\ldots,y_D)}$ be the punctured code of C from $\{0,1\}^n$ to $[q]^D$, i.e., $C_{(y_1,\ldots,y_D)}(x) = C(x)_{y_1} \circ C(x)_{y_2} \circ \cdots \circ C(x)_{y_D}$.*

From the Johnson bound, the list decoding size of C is $O(\frac{1}{\epsilon^2})$ for radius $\frac{1}{q} - O(\epsilon)$. Applying Theorem 1.5 to the setting of list decodable codes, we deduce that $D = O\left(\frac{nq \cdot \log^2 \frac{1}{\epsilon}}{\epsilon^2}\right)$ random coordinates $y_1, \ldots, y_D$ provide a punctured code with the same list decoding capacity up to a constant factor.

Corollary 1.10 *Given n, q, ϵ, there exists $D = O\left(\frac{nq \cdot \log^2 \frac{1}{\epsilon}}{\epsilon^2}\right)$ such that for any code $C : \{0,1\}^n \to [q]^{2^t}$ with relative distance $\frac{1}{q} - \epsilon^2$, the punctured code $C_{(y_1,\ldots,y_D)}$ for D random coordinates $y_1, \ldots, y_D \in [2^t]$ is list-decodable with radius $\frac{1}{q} - O(\epsilon)$ and size $O(\frac{1}{\epsilon^2})$, with probability at least 0.99.*

For small alphabets such as binary codes with $q = 2$, this improves the rate of punctured codes obtained by Rudra and Wootters [RW14] by a factor of $\log^3 \frac{1}{\epsilon}$.

1.4 Previous work.

In a seminal work, Impagliazzo, Levin, and Luby [ILL89] proved the Leftover Hash Lemma, i.e., all functions from an almost universal hash family constitute a strong extractor. In particular, this implies that all linear transformations and all Toeplitz matrices constitute strong extractors respectively.

Most previous research has focused on linear extractors, whose extractor functions are linear on the random source for every fixed seed. Because of their simplicity and various applications such as building blocks in extractors for structured sources [Li16, CL16a], there have been several constructions of linear extractors with small degree. The first non trivial progress was due to Trevisan [Tre01], who constructed the first linear extractor with a degree polynomial in n and ϵ (for entropy $k = n^{\Omega(1)}$). Based on Trevisan's work, Shaltiel and Umans [SU05] built linear extractors with almost linear degree for constant error. Later on, Guruswami, Umans, and Vadhan [GUV09] constructed almost optimal linear condensers and vertex-expansion expanders, which are variants of extractors and lead to an extractor with a degree $n \cdot \mathsf{poly}(k/\epsilon)$. However, the GUV extractor is not linear. Moreover, it is still open whether the degree of linear extractors could match the degree $\Theta(\frac{n-k}{\epsilon^2})$ of general extractors.

On the other hand, much less is known about extractors consisting of Toeplitz matrices. Prior to this work, even for $m = 2$ output bits, the best known upper bound on extractors with Toeplitz matrices was exponential in n by the Leftover Hash Lemma [ILL89] of all Toeplitz matrices.

At the same time, for many practical applications of cryptography, it is desirable to have an extractor with a small degree that runs fast and is easy to implement. In this work, we consider efficient extractors from almost universal hash families, which are easier to implement than error-correcting codes and expander graphs used in most known constructions of extractors. The most notable hash function is the Multiplicative Universal Hash introduced by Dietzfelbinger et al. [DHKP97] (described in Section 1.1), which runs significantly fast in practice and is extremely easy to implement [Tho]. Prior to this work, the best known upper bounds on the degree of extractors from almost universal hash families were the Leftover Hash Lemma [ILL89], which are exponential in the length of the random source n.

This theme of simple extractors also has been studied under another two models of computation: the local computation model and constant-depth circuits AC^0 by Bogdanov and Guo [BG13], Goldreich et al. [GVW15], and Cheng and Li [CL16b]. Other work on improving the parameters of extractors focus on the error and number of output bits. Raz et al. [RRV99] showed how to reduce the error and enlarge the number of output bits of any given extractor by sacrificing the degree.

Organization. We introduce basic notation and tools in Section 2. In Section 3, we provide a proof overview of Theorem 1.5. Then we prove Theorem 1.5 for extractors and a lower bound on the degree of extractor restrictions. Next we show the version of Theorem 1.5 for strong extractors in Section 5. Finally, we discuss a few possible future directions in Section 6.

2 Preliminaries

We use $X \lesssim Y$ to denote the inequality $X \leq C \cdot Y$ for a universal constant C. For a subset S and integer k, we use $\binom{S}{k}$ to denote all subsets of size k in S. Given an integer n, we use $[n]$ to denote the set $\{1, 2, \cdots, n\}$. For a random variable X on $\{0,1\}^n$ and a function f from $\{0,1\}^n$, let $f(X)$ denote the random variable $f(x)$ on the image of f when $x \sim X$.

Given a subset $\Lambda \in \{0,1\}^n$, we consider the flat random source of the uniform distribution over Λ, whose min-entropy is $-\log_2 \frac{1}{|\Lambda|} = \log_2 |\Lambda|$. Because any random source with min-entropy k is a linear combination of flat random sources of min-entropy k, we focus on flat random sources in the rest of this work.

We state the Leftover Hash Lemma by Impagliazzo, Levin, and Luby [ILL89].

Lemma 2.1 *For any n and m, let $\mathcal{H}$ be a family of T hash functions $\{h_1, \cdots, h_T\}$ mapping $[2^n]$ to $[2^m]$ such that for any distinct x and y, $\Pr_{h \sim H}[h(x) = h(y)] \leq 2^{-n} + 2^{-m}$. Then $\mathsf{Ext} : \{0,1\}^n \times [T] \to \{0,1\}^m$ defined as $\mathsf{Ext}(x, y) = h_y(x)$ is a strong (k, ϵ)-extractor for any k and ϵ satisfying $k \geq m + 2\log \frac{1}{\epsilon}$.*

We always use $N(0,1)$ to denote the standard Gaussian random variable and use the following concentration bound on Gaussian random variables [LT91].

Lemma 2.2 *Given any n Gaussian random variables $G_1, \cdots, G_n$ (not necessarily independent) where each G_i has expectation 0 and variance σ_i^2,*

$$\mathbb{E}\left[\max_{i \in [n]} |G_i|\right] \lesssim \sqrt{\log n} \cdot \max_{i \in [n]} \{\sigma_i\}.$$

Let S be a subset of events, X a random variable, and f any function from $S \times \mathsf{supp}(X)$ to $\mathbb{R}^+$. We state the standard symmetrization and Gaussianization [LT91, RW14] that transform bounding $\max_{\Lambda \in S} \sum_{j=1}^{n} f(\Lambda, x_i)$ of n independent random variables $x_1, \cdots, x_n \sim X$ to a Gaussian process.

Theorem 2.3 *For any integer n and n independent random samples $x_1, \cdots, x_n$ from X,*

$$\mathop{\mathbb{E}}_{x_1 \sim X, \cdots, x_n \sim X}\left[\max_{\Lambda \in S} \sum_{j=1}^{n} f(\Lambda, x_i)\right] \leq \max_{\Lambda \in S} \mathop{\mathbb{E}}_{x}\left[\sum_{j=1}^{n} f(\Lambda, x_j)\right] + \sqrt{2\pi} \cdot \mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{g \sim N(0,1)^n}\left[\max_{\Lambda \in S}\left|\sum_{j=1}^{n} f(\Lambda, x_j) g_j\right|\right]\right].$$

The first term $\max_{\Lambda \in S} \mathbb{E}_x \left[\sum_{j=1}^n f(\Lambda, x_j)\right]$ is the largest expectation over all events Λ, and the second term is to bound the deviation of every event from its expectation simultaneously. For completeness, we provide a proof of Theorem 2.3 in Appendix A.

We state the Beck-Fiala theorem in the discrepancy theory [Cha00].

Theorem 2.4 *[Beck-Fiala theorem] Given a universe $[n]$ and a collection of subsets $S_1, \cdots, S_l$ such that each element $i \in [n]$ appears in at most d subsets, there exists an assignment $\chi : [n] \to \{\pm 1\}$ such that for each subset S_j, $|\sum_{i \in S_j} \chi(i)| < 2d$.*

3 Proof Overview

We sketch the proof of Theorem 1.5 in this section. For ease of exposition, we consider extractor restrictions and bound their degree in two steps. Given any (k, ϵ)-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, we first bound the degree of $\mathsf{Ext}_{(y_1,\cdots,y_D)}$ to fool one fixed test $T \subseteq \{0,1\}^m$. Then we generalize it to all statistical tests in $\{0,1\}^m$.

For any fixed test $T \subseteq \{0,1\}^m$, we consider the expected error of a random restricted extractor in the test T:

$$\mathop{\mathbb{E}}_{y_1,\cdots,y_D \sim \{0,1\}^t} \left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \sum_{i=1}^{D} \Pr_{x \sim \Lambda} \left[\mathsf{Ext}(x, y_i) \in T\right] \right]. \tag{1}$$

We apply a chaining argument to prove that $D = \tilde{O}(\frac{n}{\epsilon^2})$ suffices to bound (1) by $D(\frac{|T|}{2^m} + 2\epsilon)$, which improves the naive union bound over all flat sources Λ with min entropy k.

By Theorem 2.3, it suffices to consider the expectation of every Λ and bound the deviation of every Λ from its expectation simultaneously by a Gaussian process. The expectation of every Λ is at most $D(\frac{|T|}{2^m} + \epsilon)$ from the property of Ext. Then we bound the Gaussian process

$$\mathop{\mathbb{E}}_{g \sim N(0,1)^D} \left[\max_{\Lambda} \left| \sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] \cdot g_i \right| \right] \lesssim \sqrt{nD} \cdot \log D \text{ for any } y_1, \cdots, y_D.$$

To bound this, by Dudley's entropy Theorem, we plan to find a small covering by balls of vectors

$$\left\{ \big(\Pr[\mathsf{Ext}(\Lambda, y_i) \in T]\big)_{i \in [D]} \middle| \Lambda \in \binom{\{0,1\}^n}{2^k} \right\}$$

in $\mathbb{R}^D$ under the Euclidean norm. One natural method to construct the covering and bound its radius is Maurey's empirical method (namely the probabilistic method), which is also the chaining argument used by Rudra and Wootters [RW14]. In this proof, we provide an alternative method to bound the covering radius based on the Beck-Fiala Theorem 2.4, which shaves an extra factor of $\mathsf{poly}(k)$ in the previous method.

Next we strengthen the guarantee in (1) to a high probability event:

$$\Pr_{y_1,\cdots,y_D} \left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \sum_{i=1}^{D} \left(\Pr_{x \sim \Lambda} \left[\mathsf{Ext}(x, y_i) \in T\right] - \frac{|T|}{2^m} \right) \le D \cdot 3\epsilon \right] \ge 1 - \delta \text{ after enlarging } D \text{ to } \tilde{O}\left(\frac{n \log \frac{1}{\delta}}{\epsilon^2}\right). \tag{2}$$

We will rewrite $\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \sum_{i=1}^{D} \left(\Pr_{x \sim \Lambda}\left[\mathsf{Ext}(x, y_i) \in T\right] - \frac{|T|}{2^m} \right)$ as the ℓ_∞-norm of a vector with dimension $\binom{2^n}{2^k}$, where each coordinate corresponds to one subset Λ. Because this vector is a summation of D vectors corresponding to D independent random variables $y_1, \cdots, y_D$, we apply a vector-concentration bound in Banach spaces from Ledoux and Talagrand [LT91] to prove (2).

To finish the proof of Theorem 1.5 for extractors, we set δ in (2) to be less than $\frac{1}{2^{2^m}}$ and apply a union bound over all statistical tests in $\{0,1\}^m$. To prove Theorem 1.5 for strong extractors, we follow the same outline and tailor the chaining argument for strong extractors.

4 Extractors

We study restricted extractors in this section and prove Theorem 1.5 for extractors. The main result in this section is that most sequences of $\tilde{O}(\frac{n \cdot 2^m}{\epsilon^2})$ seeds from any given extractor constitute a restricted extractor with nearly the same parameters of min entropy and error. On the other hand, we show that for certain extractors, the degree of its restrictions is $\Omega(2^m)$ to guarantee any constant error.

We first consider the upper bound on the degree of restricted extractors for all entropy-k flat sources fooling one fixed statistical test.

Lemma 4.1 *Let* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ *be an* (k, ϵ)*-extractor and* $D = C \cdot \frac{n \cdot \log^2 \frac{n}{\epsilon}}{\epsilon^2}$ *for a universal constant* C*. Given any subset* $T \subseteq \{0,1\}^m$*, for* D *independently random seeds* $y_1, \cdots, y_D$ *in* $\{0,1\}^t$*,*

$$\mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\max_{\Lambda:|\Lambda|=2^k} \sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right]\right] \le D \cdot \left(\frac{|T|}{2^m} + 2\epsilon\right). \tag{3}$$

Proof. We symmetrize and Gaussianize the L.H.S. of (3) by Theorem 2.3:

$$\mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\max_{\Lambda:|\Lambda|=2^k} \sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right]\right] \tag{4}$$

$$\le \max_{\Lambda} \mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right]\right] + \sqrt{2\pi} \mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\mathop{\mathbb{E}}_{g \sim N(0,1)^D}\left[\max_{\Lambda}\left|\sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] \cdot g_i\right|\right]\right]. \tag{5}$$

Because Ext is an extractor for entropy k sources with error ϵ, the first term

$$\max_{|\Lambda|=2^k} \mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right]\right] \le D \cdot \left(\frac{|T|}{2^m} + \epsilon\right).$$

The technical result is a bound on the Gaussian process for any $y_1, \cdots, y_D$.

Claim 4.2 *For any* $y_1, \cdots, y_D$*,* $\mathop{\mathbb{E}}_{g \sim N(0,1)^D}\left[\max_{\Lambda:|\Lambda|=2^k}\left|\sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] \cdot g_i\right|\right] \le C_0 \cdot \sqrt{nD} \cdot \log D$ *for some universal constant* C_0*.*

We defer the proof of Claim 4.2 to Section 4.1.

By choosing the constant C large enough, for $D = C^{\frac{n \log^2 \frac{n}{\epsilon}}{\epsilon^2}}$, we can ensure that $C_0\sqrt{nD} \cdot \log D \leq \frac{\epsilon D}{5}$. This bounds (5) by $D \cdot (\frac{|T|}{2^m} + \epsilon) + \epsilon D$. □

Next, we show that a restricted extractor is good with high probability. To do this, we provide a concentration bound on $\max_{|\Lambda|=2^k} \left\{ \sum_{i=1}^D \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] \right\}$. We prove that a restricted extractor with D random seeds achieves the guarantee in Lemma 4.1 with probability $1-\delta$ after enlarging D by a factor of $\tilde{O}(\log \frac{1}{\delta})$.

Lemma 4.3 *For any $\delta > 0$, let $D = C' \cdot \frac{n \cdot \log \frac{1}{\delta}}{\epsilon^2} \cdot \log^2 \frac{n \cdot \log \frac{1}{\delta}}{\epsilon}$ for a universal constant C'. Given any (k, ϵ)-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ and any subset $T \subseteq \{0,1\}^m$, for D independently random seeds $y_1, \cdots, y_D$ in $\{0,1\}^t$,*

$$\Pr_{y_1,\cdots,y_D}\left[\max_{\Lambda:|\Lambda|=2^k} \left\{ \sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] - \frac{D \cdot |T|}{2^m} \right\} \leq D \cdot 3\epsilon \right] \geq 1 - \delta.$$

We defer the proof of Lemma 4.3 to Section 4.2. Finally we state the result about extractors.

Theorem 4.4 *Let $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ be a (k, ϵ)-extractor and $D = C \cdot \frac{n \cdot (\log \frac{1}{\delta} + 2^m)}{\epsilon^2} \cdot \log^2 \frac{n \cdot (\log \frac{1}{\delta} + 2^m)}{\epsilon}$ for a universal constant C. For a random sequence $(y_1, \cdots, y_D)$ where each $y_i \sim \{0,1\}^t$, the restricted extractor $\mathsf{Ext}_{(y_1,\cdots,y_D)}$ is a $(k, 2\epsilon)$-extractor with probability $1 - \delta$.*

Proof. We choose the error probability to be $\frac{\delta}{2^{2^m}}$ in Lemma 4.3 and apply a union bound over all possible statistical tests T in $\{0,1\}^m$. □

For extractors, we show that 2^m dependence in the degree is necessary.

Claim 4.5 *There exists a $(k = 1, \epsilon = 0)$-extractor Ext such that for any constant $\epsilon' \leq 1/2$ and $k' > 0$, any restriction $\mathsf{Ext}_{(y_1,\cdots,y_D)}$ requires $D = \Omega(2^m)$ to be an (k', ϵ')-extractor.*

Proof. Let us consider the extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^m$ defined as $\mathsf{Ext}(x, y) = y$. From the definition, it is an $(k = 1, \epsilon = 0)$-extractor. On the other hand, $\mathsf{Ext}_{(}y_1, \cdots, y_D)$ is an $(k', 0.5)$-extractor only if $D \geq 0.5 \cdot 2^m$. □

However, this lower bound may not be necessary for *strong* extractors.

4.1 Proof of Claim 4.2

Given $y_1, \cdots, y_D$ and T, for any subset Λ, we use $\vec{p}(\Lambda)$ to denote the vector $\Big(\Pr\left[\mathsf{Ext}(\Lambda, y_1) \in T\right], \cdots, \Pr\left[\mathsf{Ext}(\Lambda, y_D) \in T\right] \Big)$. Let $t = 10$ be a fixed parameter in this proof.

We rewrite the Gaussian process

$$\mathop{\mathbb{E}}_{g \sim N(0,1)^D}\left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left| \sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] \cdot g_i \right| \right] = \mathop{\mathbb{E}}_{g \sim N(0,1)^D}\left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left| \langle \vec{p}(\Lambda), g \rangle \right| \right].$$

We construct a sequence of subsets $\mathcal{F}_{t-1}, \mathcal{F}_t, \cdots, \mathcal{F}_k$ of vectors in $\mathbb{R}^D$ and a sequence of maps $\pi_j : \binom{\{0,1\}^n}{2^k} \to \mathcal{F}_j$ for each j from $t-1$ to k. We first set $\mathcal{F}_k$ to be the subset of all vectors in the Gaussian process, i.e., $\mathcal{F}_k = \left\{\vec{p}(\Lambda) \middle| \Lambda \in \binom{\{0,1\}^m}{2^k}\right\}$ and $\pi_k(\Lambda) = \vec{p}(\Lambda)$. For convenience, we set $\mathcal{F}_{t-1} = \{\vec{0}\}$ and $\pi_{t-1}(\Lambda) = \vec{0}$ for any $\Lambda \in \binom{\{0,1\}^n}{2^k}$ and specify $\mathcal{F}_j$ and π_j for $j \in [t, k-1]$ later. For any $\vec{p}(\Lambda)$ in the Gaussian process, we use the equation $\vec{p}(\Lambda) = \sum_{j=k}^{t} \pi_j(\Lambda) - \pi_{j-1}(\Lambda)$ to rewrite it:

$$\mathop{\mathbb{E}}_{g \sim N(0,1)^n}\left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left|\langle \vec{p}(\Lambda), g\rangle\right|\right] = \mathop{\mathbb{E}}_{g}\left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left|\left\langle \sum_{j=k}^{t} \pi_j(\Lambda) - \pi_{j-1}(\Lambda), g\right\rangle\right|\right] \tag{6}$$

$$= \mathop{\mathbb{E}}_{g}\left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \sum_{j=k}^{t} \left|\left\langle \pi_j(\Lambda) - \pi_{j-1}(\Lambda), g\right\rangle\right|\right] \tag{7}$$

$$= \sum_{j=k}^{t} \mathop{\mathbb{E}}_{g}\left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left|\left\langle \pi_j(\Lambda) - \pi_{j-1}(\Lambda), g\right\rangle\right|\right] \tag{8}$$

$$\lesssim \sum_{j=k}^{t} \sqrt{\log(|\mathcal{F}_j| \cdot |\mathcal{F}_{j-1}|)} \cdot \max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left\|\pi_j(\Lambda) - \pi_{j-1}(\Lambda)\right\|_2. \tag{9}$$

Here (9) follows from the union bound over Gaussian random variables — Lemma 2.2. In the rest of this proof, we provide upper bounds on $\left\|\pi_j(\Lambda) - \pi_{j-1}(\Lambda)\right\|_2$ and $|\mathcal{F}_j|$ to finish the calculation of (9).

Two upper bounds for $\left\|\pi_j(\Lambda) - \pi_{j-1}(\Lambda)\right\|_2$. We provide two methods to bound $\left\|\pi_j(\Lambda) - \pi_{j-1}(\Lambda)\right\|_2$ in (9). In this proof, for any map π_j and Λ, we always choose the map $\pi_j(\Lambda) = p(\Lambda')$ for some subset Λ'.

Claim 4.6 *Given $|\Lambda_0| \geq D^2$, there always exists $\Lambda_1 \subseteq \Lambda_0$ with size $|\Lambda_1| \in \left[|\Lambda_0|/2 - 2D, |\Lambda_0|/2 + 2D\right]$ such that*

$$\|\vec{p}(\Lambda_0) - \vec{p}(\Lambda_1)\|_2 \leq 6D^{1.5}/|\Lambda_0|.$$

Proof. We plan to use the Beck-Fiala Theorem 2.4 to find Λ_1 given Λ_0. Let the ground set be Λ_0 and the collection of subsets be $S_i = \left\{\alpha \in \Lambda_0 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}$ for each $i \in [D]$ and $S_{D+1} = \Lambda_0$. Because the degree is at most $D+1$, Theorem 2.4 implies an assignment $\chi : \Lambda_0 \to \{\pm 1\}$ satisfying that for each S_i, $|\sum_{\alpha \in S_i} \chi(\alpha)| < 2(D+1)$. We set $\Lambda_1 = \left\{\alpha \in \Lambda_0 \middle| \chi(\alpha) = 1\right\}$.

Because $|\sum_{\alpha \in \Lambda_0} \chi(\alpha)| < 2(D+1)$, $\left|\Lambda_1 - \frac{|\Lambda_0|}{2}\right| < (D+1) \leq 2D$. At the same time, for each $i \in [D]$ and S_i, $\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right| - \frac{|S_i|}{2} < (D+1)$.

To finish the proof, we prove $\big|\Pr[\mathsf{Ext}(\Lambda_0, y_i) \in T] - \Pr[\mathsf{Ext}(\Lambda_1, y_i) \in T]\big| \leq \frac{6D}{|\Lambda_0|}$.

$$\begin{aligned}
&\big|\Pr[\mathsf{Ext}(\Lambda_0, y_i) \in T] - \Pr[\mathsf{Ext}(\Lambda_1, y_i) \in T]\big| \\
=&\left|\frac{|S_i|}{|\Lambda_0|} - \frac{\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|}{|\Lambda_1|}\right| \\
\leq&\left|\frac{|S_i|}{|\Lambda_0|} - \frac{\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|}{|\Lambda_0|/2}\right| + \left|\frac{\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|}{|\Lambda_0|/2} - \frac{\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|}{|\Lambda_1|}\right| \\
<&\frac{2(D+1)}{|\Lambda_0|} + \left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right| \cdot \frac{\Big||\Lambda_0|/2 - |\Lambda_1|\Big|}{|\Lambda_0|/2 \cdot |\Lambda_1|} \\
<&\frac{2(D+1)}{|\Lambda_0|} + \frac{(D+1)}{|\Lambda_0|/2} \leq \frac{6D}{|\Lambda_0|}.
\end{aligned}$$

From the definition of $\vec{p}(\Lambda_0) = \Big(\Pr\big[\mathsf{Ext}(\Lambda_0, y_1) \in T\big], \cdots, \Pr\big[\mathsf{Ext}(\Lambda_0, y_D) \in T\big]\Big)$, this implies $\|\vec{p}(\Lambda_0) - \vec{p}(\Lambda_1)\|_2 \leq 6D^{1.5}/|\Lambda_0|$. □

Next we provide an alternative bound for Λ_0 with a small size using the probabilistic method.

Claim 4.7 *Given any* Λ_0 *of size at least* 100*, there always exists* $\Lambda_1 \subseteq \Lambda_0$ *with size* $|\Lambda_1| \in \big[|\Lambda_0|/2 - \sqrt{|\Lambda_0|}, |\Lambda_0|/2 + \sqrt{|\Lambda_0|}\big]$ *such that*

$$\|\vec{p}(\Lambda_0) - \vec{p}(\Lambda_1)\|_2 \leq 6\sqrt{D/|\Lambda_0|}.$$

Proof. We first show the existence of Λ_1 with the following two properties:

1. $|\Lambda_1| \in \big[|\Lambda_0|/2 - \sqrt{|\Lambda_0|}, |\Lambda_0|/2 + \sqrt{|\Lambda_0|}\big]$.
2. $\sum_{i \in [D]} \left(\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right| - \left|\left\{\alpha \in \Lambda_0 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|/2\right)^2 \leq D \cdot |\Lambda_0|$.

We pick each element $\alpha \in \Lambda_0$ to Λ_1 randomly and independently with probability 1/2. For the first property, $\mathop{\mathbb{E}}_{\Lambda_1}\Big[\big(|\Lambda_1| - |\Lambda_0|/2\big)^2\Big] = |\Lambda_0|/4$ implies it holds with probability at least $3/4$.

At the same time,

$$\mathop{\mathbb{E}}_{\Lambda_1}\left[\sum_{i \in [D]} \left(\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right| - \left|\left\{\alpha \in \Lambda_0 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|/2\right)^2\right] = \sum_{i \in [D]} \left|\left\{\alpha \in \Lambda_0 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|/4$$

implies the second property holds with probability at least $3/4$. Therefore there exists Λ_1 satisfying both properties.

Now let us bound $\|\vec{p}(\Lambda_0) - \vec{p}(\Lambda_1)\|_2$:

$$\sum_{i\in[D]} (\Pr[\mathsf{Ext}(\Lambda_0, y_i) \in T] - \Pr[\mathsf{Ext}(\Lambda_1, y_i) \in T])^2$$

$$\leq 2\sum_{i\in[D]} \left(\frac{\left|\left\{\alpha \in \Lambda_0 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|}{|\Lambda_0|} - \frac{\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|}{|\Lambda_0|/2} \right)^2$$

$$+ 2\sum_{i\in[D]} \left(\frac{\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|}{|\Lambda_0|/2} - \frac{\left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|}{|\Lambda_1|} \right)^2$$

$$\leq \frac{8}{|\Lambda_0|^2} \sum_{i\in[D]} \left(\left|\left\{\alpha \in \Lambda_0 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|/2 - \left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right| \right)^2$$

$$+2\sum_{i\in[D]} \left|\left\{\alpha \in \Lambda_1 \middle| \mathsf{Ext}(\alpha, y_i) \in T\right\}\right|^2 \cdot \left(\frac{|\Lambda_1| - |\Lambda_0|/2}{|\Lambda_1| \cdot |\Lambda_0|/2} \right)^2$$

$$\leq \frac{8D}{|\Lambda_0|} + 2\sum_{i\in[D]} \frac{|\Lambda_0|}{|\Lambda_0|^2/4} \leq 16D/|\Lambda_0|.$$

□

Constructions of $\mathcal{F}_j$. We construct $\mathcal{F}_{k-1}, \cdots, \mathcal{F}_t$ to fit in with Claim 4.6 and 4.7. We define two parameters $s(j)_l$ and $s(j)_u$ on the order of 2^j for each $\mathcal{F}_j$ such that

$$\mathcal{F}_j = \left\{ \vec{p}(\Lambda) \middle| \Lambda \in \binom{\{0,1\}^n}{s(j)_l} \cup \binom{\{0,1\}^n}{s(j)_l + 1} \cup \cdots \cup \binom{\{0,1\}^n}{s(j)_u} \right\}.$$

We start with $s(k)_l = s(k)_u = 2^k$ and define $s(j)_l$ and $s(j)_u$ from $j = k-1$ to t.

1. $j > 2\log D + 8$: we define $s(j)_l = \frac{s(j+1)_l}{2} - 2D$ and $s(j)_u = \frac{s(j+1)_u}{2} + 2D$. In this proof, we bound $2^j - 4D \leq s(j)_l \leq s(j)_u \leq 2^j + 4D$.

2. $j \leq 2\log D + 8$: we define $s(j)_l = \frac{s(j+1)_l}{2} - \sqrt{s(j+1)_l}$ and $s(j)_u = \frac{s(j+1)_u}{2} + \sqrt{s(j+1)_u}$. We bound $0.8 \cdot 2^j \leq s(j)_l \leq s(j)_u \leq 1.4 \cdot 2^j$ by induction. The base case is $j > 2\log D + 8$, which is proved above. Because $2D$ is always less than $\sqrt{s(j+1)_l}$ for $j > 2\log D + 8$,

$$\frac{s(j)_l}{2^j} = \prod_{i=k-1}^{j} \frac{2s(i)_l}{s(i+1)_l} \geq \prod_{i=k-1}^{j} (1 - \frac{2}{\sqrt{s(i+1)_l}}) \geq 1 - \sum_{i=k-1}^{j} \frac{2}{\sqrt{s(i+1)_l}}.$$

By induction, $\sum_{i=k-1}^{j} \frac{2}{\sqrt{s(i+1)_l}} \leq \sum_{i=k-1}^{t} \frac{2}{\sqrt{0.8 \cdot 2^j}} \leq 0.2$ given $t = 10$. Similarly,

$$\frac{s(j)_u}{2^j} = \prod_{i=k-1}^{j} \frac{2s(i)_u}{s(i+1)_u} \leq \prod_{i=k-1}^{j} (1 + \frac{2}{\sqrt{s(i+1)_u}}).$$

By induction, $\sum_{i=k-1}^{j} \frac{2}{\sqrt{s(i+1)_u}} \leq \sum_{i=k-1}^{t} \frac{2}{\sqrt{1.4 \cdot 2^j}} \leq 0.2$ given $t = 10$, which implies $\frac{s(j)_u}{2^j} \leq 1.4$.

Constructions of π_j. Next we define π_j from $j = k$ to $j = t$ by induction. The base case is $j = k$ such that $\pi_j(\Lambda) = \vec{p}(\Lambda)$ for any Λ of size 2^k. Given Λ and $\pi_j(\Lambda) \in \mathcal{F}_j$, we define $\pi_{j-1}(\Lambda)$ using Claim 4.6 or 4.7. From the definition of $\mathcal{F}_j$, $\pi_j(\Lambda) = \vec{p}(\Lambda_j)$ for some Λ_j with size in $[s(j)_l, s(j)_u]$.

For $j > 2\log D + 8$, we apply Claim 4.6 on Λ_j to find Λ_{j-1} of size $|\Lambda_{j-1}| \in [|\Lambda_j| - 2D, |\Lambda_j| + 2D]$ satisfying $\|\vec{p}(\Lambda_j) - \vec{p}(\Lambda_{j-1})\|_2 \leq 6D^{1.5}/|\Lambda_j|$.

For $j \leq 2\log D+8$, we apply Claim 4.7 on Λ_j to find Λ_{j-1} of size $|\Lambda_{j-1}| \in [|\Lambda_j| - \sqrt{|\Lambda_j|}, |\Lambda_j| + \sqrt{|\Lambda_j|}]$ satisfying $\|\vec{p}(\Lambda_j) - \vec{p}(\Lambda_{j-1})\|_2 \leq 6\sqrt{D/|\Lambda_j|}$.

Thus $|\Lambda_{j-1}|$ is always in $[s(j-1)_l, s(j-1)_u]$, which indicates $\vec{p}(\Lambda_{j-1})$ is in $\mathcal{F}_{j-1}$. We set $\pi_{j-1}(\Lambda) = \vec{p}(\Lambda_{j-1})$.

To finish this proof, we plug $0.8 \cdot 2^j \leq s(j)_l \leq s(j)_u \leq 1.4 \cdot 2^j$ and $|\mathcal{F}_j| = \sum_{i=s(j)_l}^{s(j)_u} \binom{2^n}{i} \leq (s(j)_u - s(j)_l + 1) \cdot \binom{2^n}{s(j)_u} \leq 2^j \cdot 2^{n \cdot 2^j} \leq 2^{2n \cdot 2^j}$ into (9).

$$\begin{aligned}
&\sum_{j=k}^{t} \sqrt{\log(|\mathcal{F}_j| \cdot |\mathcal{F}_{j-1}|)} \cdot \max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left\|\pi_j(\Lambda) - \pi_{j-1}(\Lambda)\right\|_2 \\
\leq & \sum_{j=k}^{2\log D+9} \sqrt{4n \cdot 2^j} \cdot 6D^{1.5}/s(j)_l + \sum_{j=2\log D+8}^{t} \sqrt{4n \cdot 2^j} \cdot 6\sqrt{D/s(j)_l} \\
\lesssim & \sum_{j=k}^{2\log D+9} \sqrt{n \cdot 2^j} \cdot D^{1.5}/2^j + \sum_{j=2\log D+8}^{t} \sqrt{n \cdot 2^j} \cdot \sqrt{D/2^j} \\
\leq & \sum_{j=k}^{2\log D+9} \sqrt{nD} \cdot \frac{D}{2^{j/2}} + \sum_{j=2\log D+8}^{t} \sqrt{nD} \\
\lesssim & \log D \cdot \sqrt{nD}.
\end{aligned}$$

4.2 Larger degree with high confidence

We finish the proof of Lemma 4.3 in this section. Given $(y_1, \cdots, y_D) \in \{0,1\}^{t \times D}$ and T, we consider the error vector in $\mathbb{R}^{\binom{2^n}{2^k}}$:

$$\mathsf{Err}(y_1, \cdots, y_D) = \left(\sum_{i=1}^{D} \left(\Pr[\mathsf{Ext}(\Lambda, y_i) \in T] - \frac{T}{2^m} \right) \right)_{\Lambda \in \binom{\{0,1\}^n}{2^k}}.$$

Because $\max\limits_{\Lambda:|\Lambda|=2^k} \sum_{i=1}^{D} \left(\Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] - \frac{|T|}{2^m} \right) \leq \|\mathsf{Err}(y_1, \cdots, y_D)\|_\infty$, we will prove

$$\Pr_{y_1, \cdots, y_D}\left[\|\mathsf{Err}(y_1, \cdots, y_D)\|_\infty \geq 3\epsilon D\right] \leq \delta \text{ for } D = C' \cdot \frac{n \cdot \log \frac{1}{\delta}}{\epsilon^2} \cdot \log^2 \frac{n \cdot \log \frac{1}{\delta}}{\epsilon}. \tag{10}$$

Since $\mathsf{Err}(y_1, \cdots, y_D) = \mathsf{Err}(y_1, \emptyset, \cdots, \emptyset) + \mathsf{Err}(\emptyset, y_2, \emptyset, \cdots, \emptyset) + \cdots + \mathsf{Err}(\emptyset, \cdots, \emptyset, y_D)$, we plan to apply a concentration bound to prove (10).

Our main tool is a concentration inequality of Ledoux and Talagrand [LT91] for symmetric vectors. For convenience, we use the following version for any Banach space from Rudelson and Vershynin, which is stated as Theorem 3.8 in [RV08].

Theorem 4.8 *Given a Banach space with norm $\|\cdot\|$, let $Y_1, \cdots, Y_m$ be independent and symmetric random vectors taking values in it with $\|Y_j\| \leq r$ for all $j \in [m]$. There exists an absolute constant C_1 such that for any integers $l \geq q$, and any $t > 0$, the random variable $\|\sum_{j=1}^m Y_j\|$ satisfies*

$$\Pr_{Y_1,\cdots,Y_m}\left[\|\sum_{j=1}^m Y_j\| \geq 8q\,\mathbb{E}\left[\|\sum_{j=1}^m Y_j\|\right] + 2r\cdot l + t\right] \leq (\frac{C_1}{q})^l + 2\exp(-\frac{t^2}{256q\,\mathbb{E}[\|\sum_{j=1}^m Y_j\|]^2}).$$

To apply this theorem for *symmetric* random vectors, we symmetrize our goal $\mathsf{Err}(y_1, \cdots, y_D)$ as follows. Given a subset $T \subseteq \{0,1\}^m$ and $2D$ seeds $(y_1, \cdots, y_D)$ and $(z_1, \cdots, z_D)$, we define a vector $\Delta_{y,z}$ from $\binom{\{0,1\}^n}{2^k}$ to $\mathbb{R}$:

$$\Delta_{y,z}(\Lambda) = \sum_{i=1}^D \left(\Pr[\mathsf{Ext}(\Lambda, y_i) \in T] - \Pr[\mathsf{Ext}(\Lambda, z_i) \in T]\right)$$

We use the ℓ_∞ norm in this section:

$$\|\Delta_{y,z}\|_\infty = \max_{\Lambda\in\binom{\{0,1\}^n}{2^k}} |\Delta_{y,z}(\Lambda)|.$$

Next we use the following Lemma to bridge Theorem 4.8 for symmetric random vectors and our goal (10).

Lemma 4.9 *When we generate $y = (y_1, \cdots, y_D)$ and $z = (z_1, \cdots, z_D)$ independently from the uniform distribution of $\{0,1\}^{D\times t}$, for $\mathsf{Err}(y)$ over the random choices $y = (y_1, \cdots, y_D)$,*

$$\Pr_y\left[\|\mathsf{Err}(y)\|_\infty \geq 2\mathop{\mathbb{E}}_y\left[\|\mathsf{Err}(y)\|_\infty\right] + \delta\right] \leq 2\cdot\Pr_{y,z}\left[\|\Delta_{y,z}\|_\infty \geq \delta\right].$$

Proof. Let Z and Z' be independent identically distributed non-negative random variables. We use the following fact from [RV08]

$$\Pr\left[Z \geq 2\,\mathbb{E}[Z] + \delta\right] \leq 2\Pr\left[Z - Z' \geq \delta\right]. \tag{11}$$

The reason is that

$$\Pr_Z\left[Z \geq 2\,\mathbb{E}[Z] + \delta\right] \leq \Pr_{Z,Z'}\left[Z - Z' \geq \delta | Z' \leq 2\,\mathbb{E}[Z]\right] \leq \frac{\Pr\left[Z - Z' \geq \delta \wedge Z' \leq 2\,\mathbb{E}[Z]\right]}{\Pr_{Z'}\left[Z' \leq 2\,\mathbb{E}[Z']\right]} \leq \frac{\Pr\left[Z - Z' \geq \delta\right]}{1/2}.$$

By plugging $Z = \|\mathsf{Err}(y)\|_\infty$ in (11),

$$\begin{aligned}
&\Pr_y\left[\|\mathsf{Err}(y)\|_\infty \geq 2\mathop{\mathbb{E}}_y\left[\|\mathsf{Err}(y)\|_\infty\right] + \delta\right] \leq 2\Pr_{y,z}\left[\|\mathsf{Err}(y)\|_\infty - \|\mathsf{Err}(z)\|_\infty \geq \delta\right]\\
=&2\Pr_{y,z}\left[\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\left\{\left|\sum_{i\in[D]}\Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] - \frac{D\cdot|T|}{2^m}\right|\right\} - \max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\left\{\left|\sum_{i\in[D]}\Pr\left[\mathsf{Ext}(\Lambda, z_i) \in T\right] - \frac{D\cdot|T|}{2^m}\right|\right\} \geq \delta\right]
\end{aligned}$$

At the same time, for any $y = (y_1, \cdots, y_D)$ and $z = (z_1, \cdots, z_D)$

$$\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left\{ \left| \sum_{i\in[D]} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] - \frac{D \cdot |T|}{2^m} \right| \right\} - \max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left\{ \left| \sum_{i\in[D]} \Pr\left[\mathsf{Ext}(\Lambda, z_i) \in T\right] - \frac{D \cdot |T|}{2^m} \right| \right\}$$

$$\le \max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left\{ \left| \sum_{i\in[D]} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] - \frac{D \cdot |T|}{2^m} \right| - \left| \sum_{i\in[D]} \Pr\left[\mathsf{Ext}(\Lambda, z_i) \in T\right] - \frac{D \cdot |T|}{2^m} \right| \right\}$$

$$\le \max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left\{ \left| \left(\sum_{i\in[D]} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] - \frac{D \cdot |T|}{2^m} \right) - \left(\sum_{i\in[D]} \Pr\left[\mathsf{Ext}(\Lambda, z_i) \in T\right] - \frac{D \cdot |T|}{2^m} \right) \right| \right\}$$

$$= \max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left\{ \left| \sum_{i\in[D]} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] - \sum_{i\in[D]} \Pr\left[\mathsf{Ext}(\Lambda, z_i) \in T\right] \right| \right\} = \|\Delta_{y,z}\|_\infty$$

From the discussion above, we have

$$\Pr_y \left[\|\mathsf{Err}(y)\|_\infty \ge 2 \mathbb{E}_y \left[\|\mathsf{Err}(y)\|_\infty\right] + \delta \right] \le 2 \Pr_{y,z} \left[\|\Delta_{y,z}\|_\infty \ge \delta\right].$$

□

Proof of Lemma 4.3. From Lemma 4.9, it is enough to use Theorem 4.8 to show a concentration bound on $\Delta_{y,z}$. We first bound $\mathbb{E}[\|\mathsf{Err}_y\|_\infty]$ and $\mathbb{E}[\|\Delta_{y,z}\|_\infty]$. Notice that the proofs of Theorem 2.3 and Lemma 4.1 indicate

$$\begin{aligned}
\mathbb{E}\left[\|\mathsf{Err}_y\|_\infty\right] =& \mathbb{E}\left[\max_{\Lambda:|\Lambda|=2^k} \left| \sum_{i=1}^{D} \left(\Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] - \frac{|T|}{2^m} \right) \right| \right] \\
\le& \mathbb{E}\left[\max_{\Lambda:|\Lambda|=2^k} \left\{ \left| \sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y_i) \in T\right] - \mathbb{E}_{y'}\left[\sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y'_i) \in T\right] \right] \right| \right.\right. \\
&+ \left.\left. \left| \mathbb{E}_{y'}\left[\sum_{i=1}^{D} \Pr\left[\mathsf{Ext}(\Lambda, y'_i) \in T\right] \right] - \frac{D \cdot |T|}{2^m} \right| \right\} \right] \\
\le& \sqrt{2\pi}\, \mathbb{E}_y \left[\mathbb{E}_{g \sim N(0,1)^n} \left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \sum_{i=1}^{D} \Pr[\mathsf{Ext}(\Lambda, y_i) \in T] \cdot g_i \right] \right] + \epsilon \cdot D \quad \text{by Claim A.1} \\
\le& C_2 \sqrt{nD} \cdot \log D + \epsilon D. \quad \text{by Claim 4.2}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}[\|\Delta_{y,z}\|_\infty] =& \mathbb{E}_{y,z} \left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \left| \sum_{i=1}^{D} \left(\Pr[\mathsf{Ext}(\Lambda, y_i) \in T] - \Pr[\mathsf{Ext}(\Lambda, z_i) \in T] \right) \right| \right] \\
\le& \sqrt{2\pi}\, \mathbb{E}_y \left[\mathbb{E}_{g \sim N(0,1)^n} \left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \sum_{i=1}^{D} \Pr[\mathsf{Ext}(\Lambda, y_i) \in T] \cdot g_i \right] \right] \quad \text{by Claim A.1} \\
\le& C_2 \sqrt{nD} \cdot \log D \quad \text{by Claim 4.2}
\end{aligned}$$

Now we rewrite $\Delta_{y,z}(\Lambda) = \sum_{i=1}^{D} \left(\Pr[\mathsf{Ext}(\Lambda, y_i) \in T] - \Pr[\mathsf{Ext}(\Lambda, z_i) \in T]\right)$ as the summation of D *symmetric and independent* random variables $\Delta_{y_i,z_i}(\Lambda) = \left(\Pr[\mathsf{Ext}(\Lambda, y_i) \in T] - \Pr[\mathsf{Ext}(\Lambda, z_i) \in T]\right)$ for $\Lambda \in \binom{\{0,1\}^n}{2^k}$. Next we bound each term

$$r = \max_{y_i,z_i} \left\{ \|\Delta_{y_i,z_i}\|_\infty \right\} = \max_{y_i,z_i,\Lambda \in \binom{\{0,1\}^n}{2^k}} \left| \Pr[\mathsf{Ext}(\Lambda, y_i) \in T] - \Pr[\mathsf{Ext}(\Lambda, z_i) \in T] \right| \le 1.$$

We choose the parameters $q = 2C_1 = \Theta(1), l = \log \frac{1}{\delta} \le \epsilon D/10, t = \epsilon D/3$ and plug them in Theorem 4.8 to bound

$$\Pr\left[\|\Delta_{y,z}\|_\infty \ge 8q \cdot C_2\sqrt{nD} \cdot \log D + 2r \cdot l + t\right] \le 2^{-l} + 2e^{-\frac{t^2}{256q\,\mathbb{E}[\|\Delta_{y,z}\|_\infty]^2}} \le 3\delta$$

while $8q \cdot C_2\sqrt{nD} \cdot \log D + 2r \cdot l + t \le 0.8\epsilon D$.

Since $\mathbb{E}[\|\mathsf{Err}(y)\|_\infty] \le 1.1\epsilon D$, we have $\Pr[\|\mathsf{Err}(y)\|_\infty \ge 3 \cdot \epsilon D] \le 3\delta$. □

5 Strong Extractors

We extend our techniques to strong extractors in this section.

Theorem 5.1 *Let $D = C \cdot \frac{n \cdot 2^m}{\epsilon^2} \cdot (\log \frac{n}{\epsilon} + m)^2$ for a universal constant C and $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ be any strong (k, ϵ)-extractor. For D independently random seeds $y_1, \cdots, y_D$, $\mathsf{Ext}_{(y_1,\cdots,y_D)}$ is a strong extractor for entropy k sources with expected error 2ϵ.*

The proof of Theorem 5.1 follows the same outline of the proof of Lemma 4.1 with different parameters. We apply a chaining argument to bound the L^1 error of all entropy k sources Λ:

$$\max_{|\Lambda|=2^k} \left\{ \sum_{i=1}^{D} \left(\sum_{\alpha \in \{0,1\}^m} |\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - 2^{-m}| \right) \right\},$$

instead of bounding the error over all statistical tests in degree D strong extractors. For completeness, we provide the proof in Appendix B.

Similar to Lemma 4.3, when we enlarge the degree by a factor of $\tilde{O}(\log \frac{1}{\delta})$, we improve the guarantee to a high probability $1 - \delta$ instead of an expected error.

Corollary 5.2 *For any $\delta > 0$, let $D = C \cdot \frac{n \cdot 2^m \log \frac{1}{\delta}}{\epsilon^2} \cdot \log^2 \frac{n \cdot 2^m \log \frac{1}{\delta}}{\epsilon}$ for a universal constant C. Given any strong (k, ϵ)-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, for D independently random seeds $y_1, \cdots, y_D$, $\mathsf{Ext}_{(y_1,\cdots,y_D)}$ is a strong $(k, 3\epsilon)$-extractor with probability at least $1 - \delta$.*

6 Conclusion and future work

We have shown that a quasi-linear number of random seeds from any given extractor constitute a restricted extractor with nearly the same parameters of min entropy and error, though outputting a constant number of bits. In particular, this implies the existence of quasi-linear degree strong extractors from Toeplitz matrices with a constant number of output bits.

Even for a large output, we show that a quasi-linear number of random seeds from any given extractor fool any *fixed* statistical test for all sources with fixed-entropy. A natural question is whether the 2^m dependence in the degree of restricted *strong* extractors is necessary or not. We discuss the bottleneck of our approach and possible future directions in the rest of this section.

Chaining argument on strong extractors. Let us consider the chaining argument of strong extractors on the L^1 error $\|\mathsf{Ext}(\Lambda, y_i) - U_m\|_1 = \sum_{\alpha \in \{0,1\}^m} |\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - 2^{-m}|$. We symmetrize and Gaussianize

$$\mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\max_{\Lambda \in \binom{\{0,1\}^n}{2^k}} \textstyle\sum_{i=1}^D \|\mathsf{Ext}(\Lambda, y_i) - U_m\|_1\right] \text{ to}$$

$$\max_{\Lambda} \mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\sum_{i=1}^D \|\mathsf{Ext}(\Lambda, y_i) - U_m\|_1\right] + \sqrt{2\pi} \mathop{\mathbb{E}}_{y_1,\cdots,y_D} \underbrace{\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda}\left|\left\langle \left(\|\mathsf{Ext}(\Lambda, y_i) - U_m\|_1\right)_{i=1,\cdots,D}, g\right\rangle\right|\right]}_{\text{A Gaussian process } P}.$$

In our proof of Theorem 5.1, we use the following relaxation to bound the distance of two vectors in the Gaussian process P corresponding to two subsets Λ and Λ':

$$\left\|\left(\|\mathsf{Ext}(\Lambda, y_i) - U_m\|_1\right)_{i=1,\cdots,D} - \left(\|\mathsf{Ext}(\Lambda', y_i) - U_m\|_1\right)_{i=1,\cdots,D}\right\|_2^2 \tag{12}$$

$$= \sum_{i=1}^D \left(\|\mathsf{Ext}(\Lambda, y_i) - U_m\|_1 - \|\mathsf{Ext}(\Lambda', y_i) - U_m\|_1\right)^2 \tag{13}$$

$$\leq \sum_{i=1}^D (\|\mathsf{Ext}(\Lambda, y_i) - \mathsf{Ext}(\Lambda', y_i)\|_1)^2. \tag{14}$$

The shortcoming of our approach is that the subset chaining argument provides a tight analysis on (14) but not (12).

We show that the Gaussian process under the distance (14) is $\Omega(\sqrt{2^m})$ from the Sudakov minoration. For example, let us consider the distance of the first coordinate $\|\mathsf{Ext}(\Lambda, y_1) - \mathsf{Ext}(\Lambda', y_1)\|_1$. Because of the existence of codes with constant rate and linear distance, there exists $l = \exp(2^m)$ subsets $T_1, \cdots, T_l$ in $\{0,1\}^m$ such that $|T_i \setminus T_j| = \Omega(2^m)$ for any $i \neq j$. Let $\Lambda_1, \cdots, \Lambda_l$ be the inverse images of $T_1, \cdots, T_l$ in $\mathsf{Ext}(\cdot, y_1)$. Then $\|\mathsf{Ext}(\Lambda_i, y_1) - \mathsf{Ext}(\Lambda_j, y_1)\|_1 = \Omega(1)$ for any $i \neq j$ from the distance of the code, which indicates the Gaussian process is $\Omega(2^m)$ from the Sudakov minoration for the distance (14).

One possible approach is to construct a chain of the Gaussian process P for specific strong extractors based on the seeds $y_1, \cdots, y_D$. For example, when Ext is the linear extractor of all linear transformation, $\mathsf{Ext}(x, y_1), \cdots, \mathsf{Ext}(x, y_D)$ correspond to D linear transformations $A_1(x), \cdots, A_D(x)$. A interesting direction is to build a chaining argument based on these D linear transformations $A_1, \cdots, A_D$ instead of using the chain of subsets of size $2^{k-1}, 2^{k-2}, \cdots, O(1)$ for any $A_1, \cdots, A_D$.

Explicit Extractors. Although our results indicate that $\tilde{O}(\frac{n}{\epsilon^2})$ random Toeplitz matrices or universal hash functions constitute a good extractor for a constant number of output bits, it is unclear how to derandomize the chaining argument to obtain an explicit one. One open question is to construct explicit extractors of the above forms with degree polynomial in n and ϵ.

Other combinatorial objects. Another question is to generalize our results to other combinatorial objects such as condensers and bipartite expanders. For example, bipartite graphs of low degree Reed-Solomon codes on $\mathbb{F}_q^k \cup (\mathbb{F}_q \times \mathbb{F}_q)$, where each vertex in $\mathbb{F}_q^k$ is a degree $(k-1)$ polynomial $p(x)$ with q neighbors $\left\{\left(\alpha, p(\alpha)\right) \middle| \alpha \in \mathbb{F}_q\right\}$, are unbalanced expanders with large left-degree q. One method to reduce the degree q is to sample D random $\beta_1, \cdots, \beta_D \in \mathbb{F}_q$ and restrict the right hand side to $\{\beta_1 \times \mathbb{F}_q, \cdots, \beta_D \times \mathbb{F}_q\}$. The open problem is how large D should be to guarantee it is still an unbalanced expander.

Acknowledgements

The authors would like to thank Eric Price for many helpful discussions about chaining arguments.

References

[BDT17] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. An efficient reduction from two-source to non-malleable extractors: achieving near-logarithmic min-entropy. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017*, pages 1185–1194, 2017. 2

[BG13] Andrej Bogdanov and Siyao Guo. Sparse extractor families for all the entropy. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, ITCS '13, pages 553–560, 2013. 2, 5

[CGV13] Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. Restricted isometry of fourier matrices and list decodability of random linear codes. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 432–442, 2013. 3

[Cha00] Bernard Chazelle. *The Discrepancy Method*. Cambridge University Press, 2000. 6

[CL16a] Eshan Chattopadhyay and Xin Li. Extractors for sumset sources. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 299–311, 2016. 4

[CL16b] Kuan Cheng and Xin Li. Randomness extraction in AC0 and with small locality. *CoRR*, abs/1602.01530, 2016. 2, 5

[CW79] J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions (extended abstract). *JOURNAL OF COMPUTER AND SYSTEM SCIENCES*, 18:143–154, 1979. 2, 3

[DHKP97] Martin Dietzfelbinger, Torben Hagerup, Jyrki Katajainen, and Martti Penttonen. A reliable randomized algorithm for the closest-pair problem. *J. Algorithms*, 25(1):19–51, October 1997. 3, 4

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from parvaresh–vardy codes. *J. ACM*, 56(4):20:1–20:34, July 2009. 4

[GVW15] Oded Goldreich, Emanuele Viola, and Avi Wigderson. On randomness extraction in ac0. In *In 30th Conference on Computetational Complexity*, CCC 2015, pages 601–668, 2015. 2, 5

[ILL89] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*, STOC '89, pages 12–24, New York, NY, USA, 1989. ACM. 2, 3, 4, 5

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 168–177, 2016. 4

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach spaces*. Springer, 1991. 3, 5, 7, 12

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. Comput. Syst. Sci.*, 52(1):43–52, 1996. 1

[RRV99] Ran Raz, Omer Reingold, and Salil Vadhan. Error reduction for extractors. In *Proceedings of the 40th Annual Symposium on the Foundations of Computer Science*, New York, NY, October 1999. IEEE. 5

[RT00] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM J. Discrete Math.*, 13(1):2–24, 2000. 1

[RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008. 3, 12, 13

[RW14] Atri Rudra and Mary Wootters. Every list-decodable code for high noise has abundant near-optimal rate puncturings. In *STOC*, 2014. 1, 3, 4, 5, 6

[Sha02] Ronen Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the European Association for Theoretical Computer Science,*, 77:67–95, 2002. 1

[SU05] Ronen Shaltiel and Christopher Umans. Simple extractors for all min-entropies and a new pseudorandom generator. *J. ACM*, 52(2):172–216, March 2005. 4

[SV86] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from slightly-random sources. *J. Comput. System Sci.*, 33:75–87, 1986. 1

[Tho] Mikkel Thorup. High speed hashing for integers and strings. `https://arxiv.org/abs/1504.06804`. Online; accessed March 6th, 2018. 4

[Tre01] Luca Trevisan. Extractors and pseudorandom generators. *J. ACM*, 48(4):860–879, July 2001. 1, 4

[TZ04] Amnon Ta-Shma and David Zuckerman. Extractor codes. *IEEE Transactions on Information Theory*, 50:3015–3025, 2004. 1, 3

[Vaz87] U. Vazirani. Efficiency considerations in using semi-random sources. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC ’87, pages 160–168, 1987. 3

[Woe99] Philipp Woelfel. Efficient strongly universal and optimally universal hashing. In *International Symposium on Mathematical Foundations of Computer Science, MFCS, 1999*, pages 262–272, 1999. 3

[Woo13] Mary Wootters. On the list decodability of random linear codes with large error rates. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC ’13, pages 853–860, New York, NY, USA, 2013. ACM. 3

[Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(1):103–128, 2007. 1

A Symmetrization and Gaussianization

We finish the proof of Theorem 2.3 in this section. We first symmetrize it by

$$\begin{aligned}
\mathop{\mathbb{E}}_{x_1,\cdots,x_n}\left[\max_{\Lambda}\sum_{j=1}^{n} f(\Lambda, x_j)\right] &= \mathop{\mathbb{E}}_{x_1,\cdots,x_n}\left[\max_{\Lambda}\left(\sum_{j=1}^{n} f(\Lambda, x_j) - \mathop{\mathbb{E}}_{x'_1,\cdots,x'_n}[\sum_{j=1}^{n} f(\Lambda, x'_j)] + \mathop{\mathbb{E}}_{x'_1,\cdots,x'_n}[\sum_{j=1}^{n} f(\Lambda, x'_j)]\right)\right] \\
&\le \max_{\Lambda}\mathop{\mathbb{E}}_{x'}\left[\sum_{j=1}^{n} f(\Lambda, x'_j)\right] + \mathop{\mathbb{E}}_{x}\left[\max_{\Lambda}\left|\sum_{j=1}^{n} f(\Lambda, x_j) - \mathop{\mathbb{E}}_{x'}[\sum_{j=1}^{n} f(\Lambda, x'_j)]\right|\right].
\end{aligned}$$

Then we apply Gaussianization on the second term.

Claim A.1 *Let $g = (g_1, \cdots, g_n)$ denote the Gaussian vector sampled from $N(0,1)^n$,*

$$\mathop{\mathbb{E}}_{x}\left[\max_{\Lambda}\left|\sum_{j=1}^{n} f(\Lambda, x_j) - \mathop{\mathbb{E}}_{x'}[\sum_{j=1}^{n} f(\Lambda, x'_j)]\right|\right] \le \sqrt{2\pi}\cdot\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda}\left|\sum_{j=1}^{n} f(\Lambda, x_j) g_j\right|\right]\right].$$

Proof. Let g denote a sequence of n independent Gaussian random variables. We first use the convexity of the $|\cdot|$ function to move $\underset{x'}{E}$ to the left hand side:

$$\begin{aligned}
\mathop{\mathbb{E}}_{x}\left[\max_{\Lambda}\left|\sum_{j=1}^{n} f(\Lambda,x_j) - \mathop{\mathbb{E}}_{x'}[\sum_{j=1}^{n} f(\Lambda,x'_j)]\right|\right] &\le \mathop{\mathbb{E}}_{x}\left[\max_{\Lambda}\mathop{\mathbb{E}}_{x'}\left|\sum_{j=1}^{n} f(\Lambda,x_j)-\sum_{j=1}^{n} f(\Lambda,x'_j)\right|\right]\\
&\left(\text{use } \max_{i} \mathop{\mathbb{E}}_{G}[G_i] \le \mathop{\mathbb{E}}_{G}[\max_{i} G_i] \text{ to move } \mathop{\mathbb{E}}_{x'} \text{ out}\right)\\
&\le \mathop{\mathbb{E}}_{x,x'}\left[\max_{\Lambda}\left|\sum_{j=1}^{n} f(\Lambda,x_j)-\sum_{j=1}^{n} f(\Lambda,x'_j)\right|\right]\\
&\left(\text{use the fact } \mathbb{E}[|g_j|]=\sqrt{2/\pi}\right)\\
&\le \sqrt{\pi/2}\mathop{\mathbb{E}}_{x,x'}\left[\max_{\Lambda}\left|\sum_{j=1}^{n}\big(f(\Lambda,x_j)-f(\Lambda,x'_j)\big)\cdot\mathop{\mathbb{E}}_{g_j}|g_j|\right|\right]\\
&\left(\text{use the convexity of } |\cdot| \text{ to move } \mathop{\mathbb{E}}_{g}\right)\\
&\le \sqrt{\pi/2}\mathop{\mathbb{E}}_{x,x'}\left[\max_{\Lambda}\mathop{\mathbb{E}}_{g}\left|\sum_{j=1}^{n}\big(f(\Lambda,x_j)-f(\Lambda,x'_j)\big)\cdot|g_j|\right|\right]\\
&\left(\text{use } \max_{i} \mathop{\mathbb{E}}_{G}[G_i] \le \mathop{\mathbb{E}}_{G}[\max_{i} G_i] \text{ to move } \mathop{\mathbb{E}}_{g} \text{ out}\right)\\
&\le \sqrt{\pi/2}\mathop{\mathbb{E}}_{x,x'}\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda}\left|\sum_{j=1}^{n}\big(f(\Lambda,x_j)-f(\Lambda,x'_j)\big)\cdot|g_j|\right|\right]\\
&\left(\text{use the symmetry of } f(\Lambda,x_j)-f(\Lambda,x'_j)\right)\\
&= \sqrt{\pi/2}\mathop{\mathbb{E}}_{g}\mathop{\mathbb{E}}_{x,x'}\left[\max_{\Lambda}\left|\sum_{j=1}^{n}\big(f(\Lambda,x_j)-f(\Lambda,x'_j)\big)\cdot g_j\right|\right]\\
&\left(\text{use the triangle inequality}\right)\\
&\le \sqrt{\pi/2}\mathop{\mathbb{E}}_{x,x'}\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda}\left|\sum_{j=1}^{n} f(\Lambda,x_j)g_j\right| + \max_{\Lambda}\left|-\sum_{j=1}^{n} f(\Lambda,x'_j)g_j\right|\right]\\
&\left(\text{use the symmetry of } g_j\right)\\
&\le \sqrt{2\pi}\mathop{\mathbb{E}}_{x}\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda}\left|\sum_{j=1}^{n} f(\Lambda,x_j)g_j\right|\right].
\end{aligned}$$

□

Remark A.2 *We use the independence between $x_1,\cdots,x_n$ in the third last step.*

B Strong Linear Extractors: Proof of Theorem 5.1

We finish the proof of Theorem 5.1 by showing D random seeds constitute a strong extractor:

$$\mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\sum_{i=1}^{D}\sum_{\alpha\in\{0,1\}^m}\left|\Pr_{x\sim\Lambda}[\mathsf{Ext}_0(x,y_i)=\alpha]-2^{-m}\right|\right]\le 4\epsilon D. \tag{15}$$

For convenience, we use $\Pr[\mathsf{Ext}_0(\Lambda,y_i)=\alpha]$ to denote $\Pr_{x\sim\Lambda}[\mathsf{Ext}_0(x,y_i)=\alpha]$ and $\mathsf{Err}_y(\Lambda)$ to denote the error of the seed y and subset Λ , i.e., $\mathsf{Err}_y(\Lambda)=\sum_{\alpha\in\{0,1\}^m}\left|\Pr[\mathsf{Ext}_0(\Lambda,y)=\alpha]-2^{-m}\right|$. We use these notations to rewrite (15) as

$$\mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\sum_{i=1}^{D}\mathsf{Err}_{y_i}(\Lambda)\right].$$

Then we symmetrize and Gaussianize it by Theorem 2.3:

$$\mathop{\mathbb{E}}_{y}\left[\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\sum_{i=1}^{D}\mathsf{Err}_{y_i}(\Lambda)\right]\le\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\mathop{\mathbb{E}}_{y}\left[\sum_{i=1}^{D}\mathsf{Err}_{y_i}(\Lambda)\right]+\sqrt{2\pi}\mathop{\mathbb{E}}_{y}\left[\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\left|\sum_{i=1}^{D}\mathsf{Err}_{y_i}(\Lambda)\cdot g_i\right|\right]\right]. \tag{16}$$

Because Ext_0 is a strong extractor, the first term $\mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\sum_{i=1}^{D}\mathsf{Err}_{y_i}(\Lambda)\right]$ is at most $2\epsilon D$ for any Λ of size 2^k.

To bound the second term in (16), we fix the seeds $y_1,\cdots,y_D$ and bound the Gaussian process.

Claim B.1 *For any seeds $y_1,\cdots,y_D$, $\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\left|\sum_{i\in[D]}\mathsf{Err}_{y_i}(\Lambda)\cdot g_i\right|\right]\le C_0(\log D+m)\sqrt{nD\cdot 2^m}$ for a constant C_0.*

We defer the proof of this claim to Section B.1. We finish the proof by bounding (15) as follows:

$$\begin{aligned}&\mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\sum_{i=1}^{D}\mathsf{Err}_{y_i}(\Lambda)\right]\\ \le&\sqrt{2\pi}\cdot\mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left\{\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\left|\sum_{i}\mathsf{Err}_{y_i}(\Lambda)\cdot g_i\right|\right]\right\}+\epsilon D\\ \le&\sqrt{2\pi}\cdot C_0(\log D+m)\sqrt{nD\cdot 2^m}+D\cdot\epsilon.\end{aligned}$$

We choose $D=10C_0^2\cdot\frac{n(\log\frac{n}{\epsilon}+m)^2\cdot 2^m}{\epsilon^2}$ such that

$$\mathop{\mathbb{E}}_{y_1,\cdots,y_D}\left[\max_{\Lambda\in\binom{\{0,1\}^n}{2^k}}\sum_{i=1}^{D}\mathsf{Err}_{y_i}(\Lambda)\right]\le 4\epsilon D.$$

This indicates the error of the strong linear extractor constituted by $A_1,\cdots,A_D$ is 2ϵ in statistical distance.

B.1 Proof of Claim B.1

We prove Claim B.1 in this section. We fix a parameter $t = 8$ in this proof.

Recall that $y_1, \cdots, y_D$ are fixed in this section, we use $\mathsf{Err}(\Lambda)$ to denote the vector $(\mathsf{Err}_{y_1}(\Lambda), \cdots, \mathsf{Err}_{y_D}(\Lambda))$. We rewrite the Gaussian process as

$$\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda \in \binom{2^n}{2^k}} \left|\sum_{i \in [D]} \mathsf{Err}_{y_i}(\Lambda) \cdot g_i\right|\right] = \mathop{\mathbb{E}}_{g}\left[\max_{\Lambda \in \binom{2^n}{2^k}} \left|\big\langle \mathsf{Err}(\Lambda), g\big\rangle\right|\right].$$

We define a sequence of subsets $\mathcal{F}_{t-1}, \mathcal{F}_t, \mathcal{F}_{t+1}, \cdots, \mathcal{F}_k$ of vectors in $\mathbb{R}^D$ where $\mathcal{F}_{t-1} = \{\vec{0}\}$, $|\mathcal{F}_i| = \mathsf{poly}(\binom{2^n}{2^i})$, and $\mathcal{F}_k = \big\{\mathsf{Err}(\Lambda)\big|\Lambda \in \binom{\{0,1\}^n}{2^k}\big\}$. For each i from t to k, we construct a map $\pi_i : \mathcal{F}_k \to \mathcal{F}_i$, except that π_k is the identity map and $\pi_{t-1}(v) = \vec{0}$ for any v. For any vector $v \in \mathcal{F}_k$,

$$v = \sum_{j=k}^{t} \pi_j(v) - \pi_{j-1}(v).$$

We plug these notations into the Gaussian process:

$$\begin{aligned}
\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda \in \binom{2^n}{2^k}} \left|\big\langle \mathsf{Err}(\Lambda), g\big\rangle\right|\right] &= \mathop{\mathbb{E}}_{g}\left[\max_{v \in \mathcal{F}_k} \left|\big\langle v, g\big\rangle\right|\right] &(17)\\
&= \mathop{\mathbb{E}}_{g}\left[\max_{v \in \mathcal{F}_k} \left|\big\langle \sum_{j=k}^{t} \pi_j(v) - \pi_{j-1}(v), g\big\rangle\right|\right] &(18)\\
&\le \mathop{\mathbb{E}}_{g}\left[\max_{v \in \mathcal{F}_k} \sum_{j=k}^{t} \left|\big\langle \pi_j(v) - \pi_{j-1}(v), g\big\rangle\right|\right] &(19)\\
&\le \sum_{j=k}^{t} \mathop{\mathbb{E}}_{g}\left[\max_{v \in \mathcal{F}_k} \left|\big\langle \pi_j(v) - \pi_{j-1}(v), g\big\rangle\right|\right] &(20)\\
&\lesssim \sum_{j=k}^{t} \sqrt{\log|\mathcal{F}_j| \cdot |\mathcal{F}_{j-1}|} \cdot \max_{v} \|\pi_j(v) - \pi_{j-1}(v)\|_2. &(21)
\end{aligned}$$

We first construct $\mathcal{F}_j$ from $j = k$ to $j = t$ then define their maps $\pi_{k-1}, \cdots, \pi_t$. To construct $\mathcal{F}_j$, we will specify two parameters $s(j)_l = s(j)_u = \Theta(2^j)$ for the size of Λ such that

$$\mathcal{F}_j = \left\{\mathsf{Err}(\Lambda)\big|\Lambda \in \binom{\{0,1\}^n}{s(j)_l} \cup \binom{\{0,1\}^n}{s(j)_l + 1} \cdots \cup \binom{\{0,1\}^n}{s(j)_u}\right\}.$$

Notice that the size of each subset $\mathcal{F}_j$ is bounded by

$$|\mathcal{F}_j| \le \binom{2^n}{s(j)_l} + \cdots + \binom{2^n}{s(j)_u}.$$

The base case is $s(k)_l = s(k)_u = 2^k$ and $\mathcal{F}_k = \big\{\mathsf{Err}(\Lambda)\big|\Lambda \in \binom{\{0,1\}^n}{2^k}\big\}$.

Construction of $\mathcal{F}_j$ for $j > 4\log D + m$: $s(j)_l = s(j+1)_l/2 - 2D$ and $s(j)_u = s(j+1)_u/2 + 2D$. We bound $s(j)_l \geq 2^j - 4D$ and $s(j)_u \leq 2^j + 4D$ for all $j > 4\log D + m$.

Construction of $\mathcal{F}_j$ for $j \leq 4\log D + m$: $s(j)_l = s(j+1)_l/2 - \sqrt{s(j+1)_l}$ and $s(j)_u = s(j+1)_u/2 + \sqrt{s(j+1)_u}$. We bound $s(j)_l \geq 0.8 \cdot 2^j$ because $s(t)_l/2^t = \prod_{j=k-1}^{t} \frac{2s(j)_l}{s(j+1)_l}$ is at least

$$(1 - \frac{2}{\sqrt{s(t+1)_l}}) \cdot (1 - \frac{2}{\sqrt{s(t+2)_l}}) \cdots (1 - \frac{2}{\sqrt{s(k)_l}}) \geq 1 - \sum_{j=t+1}^{k} \frac{2}{\sqrt{s(j)_l}} \geq 1 - \sum_{j=t+1}^{k} \frac{2}{\sqrt{0.8 \cdot 2^j}} \geq 0.8.$$

Similarly, we bound $s(j)_u \leq 1.4 \cdot 2^j$ because

$$(1 + \frac{2}{\sqrt{s(t+1)_u}}) \cdot (1 + \frac{2}{\sqrt{s(t+2)_u}}) \cdots (1 + \frac{2}{\sqrt{s(k)_u}}) \leq 1 + 2\sum_{j=t+1}^{k} \frac{2}{\sqrt{s(j)_u}} \leq 1 + 2\sum_{j=t+1}^{k} \frac{2}{\sqrt{1.4 \cdot 2^j}} \leq 1.4.$$

Construction of π_j: we construct the map π_j from $j = k-1$ to $j = t$ and bound $\|\pi_{j+1}(v) - \pi_j(v)\|_2$ for each $v \in \mathcal{F}_k$ in (21). We first use the Beck-Fiala Theorem in the discrepancy method to construct π_j with $j > 4\log D + m$ then use a randomized argument to construct π_j with $j \leq 4\log D + m$.

Claim B.2 *Given $\Lambda \geq D^4$ and D seeds $y_1, \cdots, y_D$, there always exists $\Lambda' \subseteq \Lambda$ with size $|\Lambda'| \in \big[|\Lambda|/2 - 2D, |\Lambda|/2 + 2D\big]$ such that*

$$\|\mathsf{Err}(\Lambda) - \mathsf{Err}(\Lambda')\|_2 \leq 6D^{1.5} \cdot 2^m/|\Lambda|.$$

Proof. We plan to use the Beck-Fiala Theorem from the discrepancy method. We define the ground set $S = \Lambda$ and $m = 2^m \cdot D + 1$ subsets $T_1, \cdots, T_m$ to be

$$T_{(i-1)2^m+\alpha} = \big\{x \in \Lambda \big| \mathsf{Ext}(x, y_i) = \alpha\big\} \text{ for each } \alpha \in [0, \cdots, 2^m - 1] \text{ and } i \in [D]$$

and the last $T_m = S = \Lambda$. Notice that the degree of every element $x \in \Lambda$ is $D+1$.

From the Beck-Fiala Theorem, there always exists $\chi : \Lambda \to \{\pm 1\}$ such that

$$\text{for any } i \in [m], |\sum_{x \in T_i} \chi(x)| < 2D + 2.$$

We choose $\Lambda' = \{x | \chi(x) = 1\}$. From the guarantee of T_m, we know $|\Lambda'| \in \big[|\Lambda|/2 \pm (D+1)\big]$. Next we consider $\|\mathsf{Err}(\Lambda) - \mathsf{Err}(\Lambda')\|_2$.

We fix $\alpha \in \{0,1\}^m$ and $i \in [D]$ and bound $(\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha])^2$ as follows.

$$\begin{aligned}
&(\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha])^2 \\
\leq& 2\left(\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \frac{|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}|}{|\Lambda|/2}\right)^2 + 2\left(\frac{|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}|}{|\Lambda|/2} - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha]\right)^2 \\
\leq& 2\left(\frac{|\{x \in \Lambda | \mathsf{Ext}(x, y_i) = \alpha\}| - 2|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}|}{|\Lambda|}\right)^2 + 2\left(|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}| \cdot (\frac{1}{|\Lambda|/2} - \frac{1}{|\Lambda'|})\right)^2 \\
\leq& 2(\frac{3D}{|\Lambda|})^2 + 2\left(|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}| \cdot \frac{|\Lambda|/2 - |\Lambda|'}{|\Lambda|/2 \cdot |\Lambda'|}\right)^2 \\
\leq& \frac{18D^2}{|\Lambda|^2} + 2(\frac{2D}{|\Lambda|/2})^2 = 26D^2/|\Lambda|^2.
\end{aligned}$$

We bound $\|\mathsf{Err}(\Lambda) - \mathsf{Err}(\Lambda')\|_2^2$ using the above inequality.

$$\begin{aligned}
\|\mathsf{Err}(\Lambda) - \mathsf{Err}(\Lambda')\|_2^2 &= \sum_{i=1}^{D} \left(\sum_{\alpha \in \{0,1\}^m} |\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - 2^{-m}| - |\Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha] - 2^{-m}| \right)^2 \\
&\leq \sum_{i=1}^{D} \left(\sum_{\alpha \in \{0,1\}^m} |\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha]| \right)^2 \\
&\leq 2^m \sum_{i=1}^{D} \sum_{\alpha \in \{0,1\}^m} \big(\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha]\big)^2 \\
&\leq 26 D^3 \cdot 2^{2m} / |\Lambda|^2.
\end{aligned}$$

□

Claim B.3 *Given any Λ of size at least 100, there always exists $\Lambda' \subseteq \Lambda$ with size $|\Lambda'| \in \big[|\Lambda|/2 - \sqrt{|\Lambda|}, |\Lambda|/2 + \sqrt{|\Lambda|}\big]$ such that*

$$\|\mathsf{Err}(\Lambda) - \mathsf{Err}(\Lambda')\|_2 \leq 6\sqrt{D \cdot 2^m / |\Lambda|}.$$

Proof. We show the existence of Λ' by the probabilistic method of picking each element in Λ to Λ' with probability $1/2$. Because $\mathbb{E}[|\Lambda'|] = \frac{|\Lambda|}{2}$ and $\mathbb{E}[(|\Lambda'| - \frac{|\Lambda|}{2})^2] = \frac{|\Lambda|}{4}$, Λ' satisfies

$$|\Lambda'| \in \big[\frac{|\Lambda|}{2} - \sqrt{|\Lambda|}, \frac{|\Lambda|}{2} + \sqrt{|\Lambda|}\big] \text{ with probability at least } 3/4 \text{ from the Chebyshev inequality.} \tag{22}$$

Next we consider

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\Lambda'}\left[\sum_{i \in [D]} \sum_{\alpha \in \{0,1\}^m} \big(\big|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}\big| - \big|\{x \in \Lambda | \mathsf{Ext}(x, y_i) = \alpha\}\big|/2 \big)^2 \right] \\
= &\sum_{i \in [D]} \sum_{\alpha \in \{0,1\}^m} \mathop{\mathbb{E}}_{\Lambda'}\left[\big(\big|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}\big| - \big|\{x \in \Lambda | \mathsf{Ext}(x, y_i) = \alpha\}\big|/2 \big)^2 \right] \\
= &\sum_{i \in [D]} \sum_{\alpha \in \{0,1\}^m} \big|\{x \in \Lambda | \mathsf{Ext}(x, y_i) = \alpha\}\big|/4 = D \cdot |\Lambda|/4.
\end{aligned}$$

With probability $3/4$,

$$\sum_{i \in [D]} \sum_{\alpha \in \{0,1\}^m} \big(\big|\{x \in \Lambda | \mathsf{Ext}(x, y_i) = \alpha\}\big|/2 - \big|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}\big| \big)^2 \leq D \cdot |\Lambda|. \tag{23}$$

We set Λ' to be a subset satisfying equations (22) and (23) and consider $\|\mathsf{Err}(\Lambda) - \mathsf{Err}(\Lambda')\|_2$.
We fix $\alpha \in \{0,1\}^m$ and $i \in [D]$ and bound $(\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha])^2$ as follows.

$$
\begin{aligned}
&(\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha])^2 \\
\leq & 2\left(\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \frac{|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}|}{|\Lambda|/2}\right)^2 + 2\left(\frac{|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}|}{|\Lambda|/2} - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha]\right)^2 \\
\leq & 2\left(\frac{|\{x \in \Lambda | \mathsf{Ext}(x, y_i) = \alpha\}| - 2|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}|}{|\Lambda|}\right)^2 + 2\left(|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}| \cdot (\frac{1}{|\Lambda|/2} - \frac{1}{|\Lambda'|})\right)^2 \\
\leq & 8\frac{\big(\big|\{x \in \Lambda | \mathsf{Ext}(x, y_i) = \alpha\}|/2 - |\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}\big|\big)^2}{|\Lambda|^2} + 20\frac{|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}|}{|\Lambda|^2}, \qquad (*)
\end{aligned}
$$

where we use the property (22) in the last step to bound the second term. Next we bound $\|\mathsf{Err}(\Lambda) - \mathsf{Err}(\Lambda')\|_2^2$ base on the above inquality:

$$
\begin{aligned}
\|\mathsf{Err}(\Lambda) - \mathsf{Err}(\Lambda')\|_2^2 = & \sum_{i=1}^{D}\left(\sum_{\alpha \in \{0,1\}^m} |\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - 2^{-m}| - |\Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha] - 2^{-m}|\right)^2 \\
\leq & \sum_{i=1}^{D}\left(\sum_{\alpha \in \{0,1\}^m} |\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha]|\right)^2 \\
\leq & 2^m \sum_{i=1}^{D} \sum_{\alpha \in \{0,1\}^m} \big(\Pr[\mathsf{Ext}(\Lambda, y_i) = \alpha] - \Pr[\mathsf{Ext}(\Lambda', y_i) = \alpha]\big)^2 \quad \text{next apply } (*) \\
\leq & 8 \cdot 2^m \sum_{i,\alpha} \frac{\big(\big|\{x \in \Lambda | \mathsf{Ext}(x, y_i) = \alpha\}|/2 - |\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}\big|\big)^2}{|\Lambda|^2} \\
& + 20 \cdot 2^m \sum_{i,\alpha} \frac{|\{x \in \Lambda' | \mathsf{Ext}(x, y_i) = \alpha\}|}{|\Lambda|^2} \\
\leq & 8 \cdot 2^m \frac{2D \cdot |\Lambda|}{|\Lambda|^2} + 20 \cdot 2^m \frac{D \cdot |\Lambda|}{|\Lambda|^2} \leq \frac{36D \cdot 2^m}{|\Lambda|}.
\end{aligned}
$$

□

Now we define our map $\pi_j : \mathcal{F}_k \to \mathcal{F}_j$ from $j = k$ to t by induction. The base case π_k is the identity map. Then we define π_{j-1} given π_j.

For $j > 4\log D + m$, given $\Lambda \in \binom{\{0,1\}^n}{2^k}$, let $v = \pi_j\big(\mathsf{Err}(\Lambda)\big)$ be the vector in $\mathcal{F}_j$. From the definition of $\mathcal{F}_j$, there exists Λ_j of size between $[s(j)_l, s(j)_u]$ such that $v = \mathsf{Err}(\Lambda_j)$. Let Λ_{j-1} be the subset satisfying the guarantee in Claim B.2 for Λ_j. We set $\pi_{j-1}\big(\mathsf{Err}(\Lambda)\big) = \mathsf{Err}(\Lambda_{j-1})$.

Similarly, for $j \leq 4\log D + m$, given $u = \mathsf{Err}(\Lambda)$ and $\pi_j(u) = \mathsf{Err}(\Lambda_j)$ for Λ of size 2^k, we define $\pi_{j-1}(u) = \mathsf{Err}(\Lambda_{j-1})$ where Λ_{j-1} is the subset satisfying the guarantee in Claim B.3 for Λ_j.

To finish the calculation of (21), we bound $|\mathcal{F}_j|$ by

$$
|\mathcal{F}_j| \leq \binom{2^n}{s(j)_l} + \cdots + \binom{2^n}{s(j)_u} \leq 2 \cdot 2^j \cdot \binom{2^n}{1.8 \cdot 2^j} \leq 2^{2n2^j}.
$$

From the all discussion above, we bound the Gaussian process in (21) as

$$\begin{aligned}
\mathop{\mathbb{E}}_{g}\left[\max_{\Lambda\in\binom{2^n}{2^k}}\left|\left\langle|Pj(\Lambda)-2^{-m}\cdot\vec{1}|,g\right\rangle\right|\right] &\lesssim \sum_{j=k}^{t}\sqrt{\log|\mathcal{F}_j|\cdot|\mathcal{F}_{j-1}|}\cdot\max_{v}\|\pi_j(v)-\pi_{j-1}(v)\|_2\\
&\leq \sum_{j=k}^{4\log D+m}\sqrt{2n\cdot 2^j}\cdot(10D^{1.5}\cdot 2^{m-j})+\sum_{j=4\log D+m}^{t}\sqrt{2n\cdot 2^j}\cdot 10\sqrt{D\cdot 2^{m-j}}\\
&\lesssim \sum_{j=k}^{4\log D+m}\sqrt{n}\cdot\frac{(D^{1.5}\cdot 2^m)}{2^{j/2}}+\sum_{j=4\log D+m}^{t}\sqrt{2nD\cdot 2^m}\\
&\lesssim \sqrt{n}\cdot\frac{(D^{1.5}\cdot 2^m)}{D^2\cdot 2^{m/2}}+(4\log D+m)\cdot\sqrt{2nD\cdot 2^m}\\
&\lesssim (4\log D+m)\cdot\sqrt{2nD\cdot 2^m}.
\end{aligned}$$