An Exposition of Bourgain's 2-Source Extractor

Anup Rao*

March 28, 2007

Abstract

A construction of Bourgain [Bou05] gave the first 2-source extractor to break the min-entropy rate 1/2 barrier. In this note, we write an exposition of his result, giving a high level way to view his extractor construction.

We also include a proof of a generalization of Vazirani's XOR lemma that seems interesting in its own right, and an argument (due to Boaz Barak) that shows that any two source extractor with sufficiently small error must be *strong*.

Keywords: Extractors

*Department of Computer Science, University of Texas at Austin, arao@cs.utexas.edu. Supported in part by an MCD fellowship from UT Austin and NSF Grant CCR-0310960.

1 Introduction

The min-entropy of a distribution is k if

$$\max_{x \in \mathsf{Supp}(X)} \Pr[X = x] = 2^{-k}$$

We say that a function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ is a 2-source extractor for entropy k if given any 2 independent distributions (a.k.a. sources) (X, Y) with min-entropy k, $\mathsf{Ext}(X, Y)$ is close to being uniformly random. We say that the extractor is *strong* if it satisfies the properties:

$$\Pr_{y \leftarrow_R Y}[|\mathsf{Ext}(X, y) - U_m| > \epsilon] < \epsilon$$

$$\Pr_{x \leftarrow_R X}[|\mathsf{Ext}(x, Y) - U_m| > \epsilon] < \epsilon$$

with low ϵ for arbitrary independent min-entropy k distributions X, Y.

Another way to view 2-source extractors is as boolean matrices which look random in a strong sense: Every 2-source extractor for entropy k gives an $N \times N$ boolean matrix in which every $K \times K$ minor has roughly the same number of 1's and 0's, with $N = 2^n, K = 2^k$.

The probabilistic method shows that most functions are 2-source extractors requiring entropy that is just logarithmic in the total length of each of the sources, though explicit constructions of such functions are far from achieving this bound.

The question of finding explicit deterministic polynomial time computable functions that match the random construction is the subject of this note. This question was first considered by [CG88, SV86, Vaz85]. The classical Lindsey Lemma gives a 2-source extractor for sources on n bits with entropy $n/2$. No significant progress was made in improving the entropy requirements over this, until recently. In the last few years, sparked by new results in arithmetic combinatorics [BKT04], there were several results [BIW04, BKS+05, Raz05, Bou05, Rao06, BRSW06] on constructing extractors for a few independent sources.

Today, the 2 source extractor that requires the lowest amount of entropy in every source is due to Bourgain [Bou05], who showed how to get an extractor for 2 sources, when the sum of the min-entropies of both sources is large than $2n(1/2 - \alpha)$ for some universal constant α[1]. Bourgain's construction relies on bounds coming from arithmetic combinatorics.

The rest of this note is structured as follows: first we describe Bourgain's argument. Then we give a proof of a generalization of Vazirani's XOR lemma (which appears to be folklore), that can be used to improve the output length of Bourgain's extractor. At the end we include a simple argument due to Boaz Barak that shows that any two source extractor with small enough error must be *strong*.

2 Preliminaries

2.1 Notation

We will reserve the variable p to denote primes.

$\mathbb{F}_p$ will denote the field of size p.

$\mathbb{C}$ will denote the complex numbers.

U_m will denote the uniform distribution on m bits.

[1] A result of Raz [Raz05] gives an incomparable extractor for two sources when one source has min-entropy greater than $0.51n$ and the other has entropy $\mathrm{polylog}(n)$.

G will denote a finite abelian group.
We use the convention that $N = 2^n, M = 2^m$.
For two elements of a vector space x, y, we will use $x \cdot y$ to denote the dot product $\sum_i x_i y_i$.
For a complex number x, we will use $\overline{x}$ to represent its complex conjugate.

2.2 Basic Facts

In this section we set up some basic background. We state several facts without proof though all of them can be worked out easily.

2.2.1 Inner product and Norms

Let $f : G \to \mathbb{C}$ and $g : G \to \mathbb{C}$ be two functions from a finite abelian group G to the complex numbers.
We define the inner product $\langle f, g \rangle = \mathbb{E}_x f(x)\overline{g(x)}$.
The ℓ^p norm of f is defined to be $\|f\|_{\ell^p} = (\sum_{x \in G} |f(x)|^p)^{1/p}$.
The L^p norm of f is defined to be $\|f\|_{L^p} = (\mathbb{E}_x |f(x)|^p)^{1/p} = |G|^{-1/p}\|f\|_{\ell^p}$.
The ℓ^∞ norm is defined to be $\|f\|_{\ell^\infty} = \max_x |f(x)|$.
We have the following basic relations between the norms:

Fact 2.1. $\|f\|_{\ell^\infty} \geq (1/\sqrt{|G|})\|f\|_{\ell^2}$.

Fact 2.2. $\|f\|_{\ell^2} \geq (1/\sqrt{|G|})\|f\|_{\ell^1}$.

Fact 2.3 (Triangle Inequality)**.** $|\langle f, g \rangle| \leq \|f\|_{L^1}\|g\|_{\ell^\infty}$.

2.2.2 The Cauchy Schwartz Inequality

The Cauchy Schwartz inequality will play a central role in the proof.

Proposition 2.4 (Cauchy Schwartz)**.** *For any two functions f, g as above, $|\langle f, g \rangle| \leq \|f\|_{L^2}\|g\|_{L^2}$.*

2.2.3 Characters and Discrete Fourier Basis

Let $\mathbb{F}$ be any field. Let $\psi : G \to \mathbb{F}^*$ be a group homomorphism. Then we call ψ a *character*. We call ψ non-trivial if $\psi \neq 1$. Unless we explicitly state otherwise, in this paper all characters will map into the multiplicative group of $\mathbb{C}$.

Definition 2.5 (Bilinear maps)**.** We say a map $e : G \times G \to \mathbb{C}$ is *bilinear* if it is a homomorphism in each variable (for every ξ, both $e(\cdot, \xi)$ and $e(\xi, \cdot)$ are homomorphisms). We say that it is *non-degenerate* if for every ξ, $e(\xi, \cdot)$ and $e(\cdot, \xi)$ are both non-trivial. We say that it is *symmetric* if $e(x, y) = e(y, x)$ for every $x, y \in G$.

Let $\mathbb{Z}_r$ denote the ring $\mathbb{Z}/(r)$. It is easy to check that if we let e be the map that maps $(x, y) \mapsto \exp(2\pi xy\imath/r)$, then e is a symmetric non-degenerate bilinear map. Let $G = H_1 \oplus H_2$ be the direct sum of two finite abelian groups. Let $e_1 : H_1 \times H_1 \to \mathbb{C}$ and $\epsilon_2 : H_2 \times H_2 \to \mathbb{C}$ be symmetric non-degenerate bilinear maps. Then it is easy to see that the map $(x_1 \oplus y_1, x_2 \oplus y_2) \mapsto e_1(x_1, x_2)e_2(y_1, y_2)$ is a symmetric non-degenerate bilinear map.

By the fundamental theorem of finitely generated abelian groups, every finitely generated abelian group is isomorphic to a direct sum of cyclic groups. Thus the previous discussion gives that:

Fact 2.6. *For every abelian group G, there exists a symmetric non-degenerate bilinear $e : G \times G \to \mathbb{C}$.*

It can be shown that the characters of a finite abelian group G themselves form a finite abelian group $G^\wedge$ (called the *dual* group of G), where the group operation is pointwise multiplication. Now fix *any* symmetric, non-degenerate, bilinear map e. For every $x \in G$, let e_x denote the character $e(x, \cdot)$. The map $x \mapsto e_x$ can then be shown to be an isomorphism from G to $G^\wedge$.

Fact 2.7 (Orthogonality)**.** *For any two characters e_x, e_y, we have that* $\langle e_x, e_y \rangle = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases}$.

We define the fourier transform of f (with respect to the above e) to be the function $\hat{f} : G \to \mathbb{C}$ to be: $\hat{f}(\xi) = \langle f, e_\xi \rangle$. Then it is easy to check that this is a linear, invertible operation on the space of all such functions. We get that:

Fact 2.8 (Parseval)**.** $\|f\|_{L^2} = \|\hat{f}\|_{\ell^2}$.

Proposition 2.9. $\|f\|_{\ell^1} \leq |G|^{3/2} \|\hat{f}\|_{\ell^\infty}$.

Proof.

$$\begin{aligned}
&\|f\|_{\ell^1} \\
&\leq \sqrt{|G|} \|f\|_{\ell^2} \\
&= |G| \|f\|_{L^2} \\
&= |G| \|\hat{f}\|_{\ell^2} && \text{by Parseval(Fact 2.8)} \\
&\leq |G|^{3/2} \|\hat{f}\|_{\ell^\infty}
\end{aligned}$$

□

Fact 2.10 (Fourier Inversion)**.** $f(x) = |G| \hat{\hat{f}}(-x) = \sum_{\xi \in G} \hat{f}(\xi) e_\xi(x)$.

Fact 2.11 (Preservation of Inner Product)**.** $\langle f, g \rangle = |G| \langle \hat{f}, \hat{g} \rangle$.

By the *additive characters* of a vector space over a finite field, we mean the characters of the additive group of the vector space. In our applications for 2-source extractors, the characters will always be additive characters of some such vector space. The following proposition is easy to check:

Proposition 2.12. *Let $\mathbb{F}^l$ be a vector space over a finite field $\mathbb{F}$. Let ψ be any non-trivial additive character of $\mathbb{F}$. Then the map $e(x, y) = \psi(x \cdot y) = \psi(\sum_i x_i y_i)$ is symmetric, non-degenerate and bilinear.*

2.2.4 Distributions as Functions

Note that we can view every distribution on the group G as a function that maps every group element to the probability that the element shows up. Thus we will often view distributions as real valued functions in the natural way: $X(x) = \Pr[X = x]$.

Fact 2.13. *Let X be any random variable over G. Then $H_\infty(X) \geq k$ simply means that $\|X\|_{\ell^\infty} \leq 2^{-k}$ and implies that $\|X\|_{\ell^2} \leq 2^{-k/2}$.*

Fact 2.14. *Let X be any random variable over G, then $\mathbb{E}_X(f(X)) = |G| \langle X, f \rangle$.*

Fact 2.15. *If X is a distribution, $\hat{X}(0) = 1/|G|$.*

Let U denote the uniform distribution. Then note that $|G|U$ is simply the trivial character e_0. Thus:

Fact 2.16. $\hat{U}(\xi) = \begin{cases} 1/|G| & \xi = 0 \\ 0 & \xi \neq 0 \end{cases}$.

2.2.5 Line Point Incidences

Let $\mathbb{F}$ be a finite field.

We will call a subset $\ell \subset \mathbb{F} \times \mathbb{F}$ a line if there exist two elements $a, b \in \mathbb{F}$ s.t. the elements of ℓ are exactly the elements of the form $(x, ax + b)$ for all $x \in \mathbb{F}$.

Let $P \subseteq \mathbb{F} \times \mathbb{F}$ be a set of points and L be a set of lines. We say that a point (x, y) has an incidence with a line ℓ if $(x, y) \in \ell$. A natural question to ask is how many incidences can we generate with just K lines and K points. Bourgain, Katz and Tao [BKT04] proved a bound on the number of incidences for special fields when the number of lines and points is high enough. Konyagin [Kon03] improved the bound to eliminate the need for K to be large.

Theorem 2.17 (Line Point Incidences)**.** *[BKT04, Kon03] There exists universal constants $\beta, \alpha > 0$ such that for any prime field $\mathbb{F}_p$, if L, P are sets of K lines and K points respectively, with $K \leq p^{2-\beta_0}$, the number of incidences $I(L, P)$ is at most $O(K^{3/2-\alpha})$.*

An interesting thing to note is that the theorem above does not hold for pseudolines (sets with small pairwise intersections) over finite fields, though a similar theorem does hold over the reals.

When the field is of size 2^p for a prime p a weaker version of the line point incidences theorem holds.

Theorem 2.18 (Line Point Incidences)**.** *[BKT04, Kon03] There exists a universal constant $\beta > 0$ such that for any field $\mathbb{F}_{2^p}$ of size 2^p for prime p, if L, P are sets of K lines and K points respectively with $2^{(1-\beta)p} \leq K \leq 2^{(1-\beta)p}$, the number of incidences $I(L, P)$ is at most $O(K^{3/2-\alpha})$.*

3 Bourgain's Extractor

In this section we describe Bourgain's construction. We start by revisiting the argument for why the hadamard matrix gives a good 2 source extractor for higher min-entropy.

3.1 Review: The Hadamard Extractor

First let us recall how to extract from two sources when the min-entropy is high. For a finite field $\mathbb{F}$, let $\mathsf{Had} : \mathbb{F}^l \times \mathbb{F}^l \to \mathbb{F}$ be the dot product function, $\mathsf{Had}(x, y) = x \cdot y$.

Let us review the following theorem.

Theorem 3.1. *[CG88, Vaz85] For every constant $\delta > 0$, there exists a polynomial time algorithm $\mathsf{Had} : (\{0,1\}^n)^2 \to \{0,1\}^m$ s.t. if X, Y are independent $(n, (1/2+\delta)n)$ sources, $\mathbb{E}_Y[\|\mathsf{Had}(X, Y) - U_m\|_{\ell^1}|] < \epsilon$ with $m = \Omega(n)$ and $\epsilon = 2^{-\Omega(n)}$.*

Proof. For a convenient l, we treat both inputs as elements of $\mathbb{F}^l$ (so $|\mathbb{F}|^l = N$) and then use the dot product function as described above.

We can view the random variable X as a function $X : \mathbb{F}^l \to [0, 1]$, which for each element of $\mathbb{F}^l$ assigns the probability of taking on that element. We will prove the theorem by using the XOR

lemma (Section 4). To use the lemma, we need to bound $\mathsf{bias}_\psi(X,Y) = |\mathbb{E}[\psi(\mathsf{Had}(X,Y))]|$ for every non-trivial character ψ.

Fix such a character ψ and let $e(x,y)$ be the symmetric non-degenerate bilinear map $e(x,y) = \psi(x \cdot y)$ (Proposition 2.12). Recall that e_x denotes the character $e(x,\cdot)$. Below we will use Fourier analysis according to e.

Note that

$$\mathsf{bias}_\psi(X,Y) = \left| \sum_{y \in \mathbb{F}^l} Y(y) \sum_{x \in \mathbb{F}^l} X(x)\psi(x \cdot y) \right| \tag{1}$$

Now observe that $\sum_{x \in \mathbb{F}^l} X(x)\psi(x \cdot y) = |\mathbb{F}|^l \langle e_y, X \rangle = |\mathbb{F}|^l \overline{\hat{X}(y)}$. Thus we get that

$$\mathsf{bias}_\psi(X,Y) = |\mathbb{F}|^l \left| \sum_{y \in \mathbb{F}^l} Y(y)\overline{\hat{X}(y)} \right| = |\mathbb{F}|^{2l} |\langle Y, \hat{X} \rangle|$$

Using the Cauchy Schwartz inequality and the fact that $\|f\|_{\ell^2}^2 = |\mathbb{F}|^l \|f\|_{L^2}^2$ for every $f : \mathbb{F}^l \to \mathbb{C}$, we obtain the bound:

$$\begin{aligned}
\mathsf{bias}_\psi(X,Y)^2 &\leq |\mathbb{F}|^{4l} \|Y\|_{L^2}^2 \|\hat{X}\|_{L^2}^2 \\
&= |\mathbb{F}|^{2l} \|Y\|_{\ell^2}^2 \|\hat{X}\|_{\ell^2}^2 \\
&= |\mathbb{F}|^{2l} \|Y\|_{\ell^2}^2 \|X\|_{L^2}^2 && \text{by Parseval(Fact 2.8)} \\
&= |\mathbb{F}|^{l} \|Y\|_{\ell^2}^2 \|X\|_{\ell^2}^2 \\
&\leq 2^n 2^{-k_1} 2^{-k_2}
\end{aligned}$$

Where the last inequality is obtained by Fact 2.13, assuming X, Y have min-entropy k_1, k_2. Thus, as long as $k_1 + k_2 > n$, the bias is less than 1.

Set l so that $N^{1/l} = M = |\mathbb{F}|$. By the XOR lemma Lemma 4.2 we get m bits which are $2^{(n-k_1-k_2+m)/2}$ close to uniform. The fact that the extractor is strong follows from Theorem 5.1.

Remark 3.2. If we use the more general XOR lemma Lemma 4.1, we can even afford to have $l = 1$. The final extractor function would then be $\sigma(\mathsf{Had}(X,Y))$.

■

One question we might ask is: is this error bound just an artifact of the proof? Does the Hadamard extractor actually perform better than this bound suggests? If $l = 1$, the answer is clearly no, since the output must have at least n bits of entropy to generate a uniformly random point of $\mathbb{F}$. If l is large the answer is still no; there exist sources X, Y with entropy exactly $n/2$ for which the above extractor does badly. For example let X be the source which picks the first half of its field elements randomly and sets the rest to 0. Let Y be the source that picks the second half of its field elements randomly and sets the rest to 0. Then each source has entropy rate exactly $1/2$, but the dot product function always outputs 0.

3.2 Bourgain's Extractor

A key observation of Bourgain's is that the counterexample that we exhibited for the Hadamard extractor is just a pathalogical case. He shows that although the Hadamard function doesn't extract from any sources with lower entropy, there are essentially very few counterexamples for which it fails. He then demonstrates how to encode any general source in a way that ensures that it is not a counterexample for the Hadamard function. Thus his extractor is obtained by first encoding each source in some way and then applying the Hadamard function.

For instance, consider our counterexamples from the last subsection. The counterexamples were essentially subspaces of the original space. In particular, each source was *closed under addition*, i.e. the entropy of the source $X + X$ obtained by taking two independent samples of X and summing them is exactly the same as the entropy of X. We will argue that when the source *grows with addition* (we will define exactly what we mean by this), the Hadamard extractor does not fail.

Our proof of Bourgain's theorem will be obtained in the following steps:

- First we will argue that for sources which grow with addition, the Hadamard extractor succeeds.
- Then we will show how to encode any source with sufficiently high entropy in a way that makes it grow with addition.

3.2.1 Hadamard succeeds when the sources grow with addition

To show that the Hadamard extractor succeeds, we were trying to bound the bias of the output distribution of the extractor $\mathsf{bias}_\psi(X, Y)$ Equation 1:

$$\mathsf{bias}_\psi(X, Y) = \left| \sum_{y \in \mathbb{F}^l} Y(y) \sum_{x \in \mathbb{F}^l} X(x)\psi(x \cdot y) \right| \tag{2}$$

Now for any source X, let $X - X$ be the source that samples a point by sampling two points independently according to X and subtracting them.

Lemma 3.3. $\mathsf{bias}_\psi(X, Y)^2 \leq \mathsf{bias}_\psi(X - X, Y)$

Proof.

$$\begin{aligned} \mathsf{bias}_\psi(X, Y) &= \left| \sum_{y \in \mathbb{F}^l} Y(y) \sum_{x \in \mathbb{F}^l} X(x)\psi(x \cdot y) \right| \\ &\leq \sum_{y \in \mathbb{F}^l} Y(y) \left| \sum_{x \in \mathbb{F}^l} X(x)\psi(x \cdot y) \right| \end{aligned}$$

Then by convexity,

$$\begin{aligned}\mathsf{bias}_\psi(X,Y)^2 &\le \sum_{y\in\mathbb{F}^l} Y(y)\left|\sum_{x\in\mathbb{F}^l} X(x)\psi(x\cdot y)\right|^2 \\ &= \left|\sum_{y\in\mathbb{F}^l} Y(y)\sum_{x_1,x_2\in\mathbb{F}^l} X(x_1)X(x_2)\psi(x_1\cdot y)\psi(-x_2\cdot y)\right| \\ &= \left|\sum_{y\in\mathbb{F}^l} Y(y)\sum_{x_1,x_2\in\mathbb{F}^l} X(x_1)X(x_2)\psi((x_1-x_2)\cdot y)\right|\end{aligned}$$

Now let X' denote the source $X - X$. Then by grouping terms, we see that the last expression is simply:

$$\begin{aligned}\mathsf{bias}_\psi(X,Y)^2 &\le \left|\sum_{y\in\mathbb{F}^l} Y(y)\sum_{x\in\mathbb{F}^l} X'(x)\psi(x\cdot y)\right| \\ &= \mathsf{bias}(X-X,Y)\end{aligned}$$

□

Notice the magic of this "squaring the sum" trick. By squaring the sum for the expectation via Cauchy Schwartz, starting with our original bound for the error of the extractor, we obtained a bound that behaves as if our original source was $X' = X - X$ instead of X! If X' has much higher entropy than X, we have made progress; we can follow the rest of the proof of Theorem 3.1 in the same way and obtain an error bound that is a bit worse (because we had to square the bias), but now assuming that our input source was X' instead of X.

Let us explore how else we might use this trick. For one thing, we see that we can easily compose this trick with itself. Applying the lemma again we obtain $\mathsf{bias}_\psi(X,Y)^4 \le \mathsf{bias}_\psi(X-X,Y)^2 \le \mathsf{bias}_\psi(X-X-X+X,Y) = \mathsf{bias}_\psi(2X-2X,Y)$.

Applying the lemma with respect to Y (by symmetry), we obtain $\mathsf{bias}_\psi(X,Y)^8 \le \mathsf{bias}_\psi(2X-2X,Y-Y)$.

In general, we obtain the following lemma:

Lemma 3.4. *There exists a polynomial time computable function* $\mathsf{Had}:\mathbb{F}^l\times\mathbb{F}^l\to\{0,1\}^m$ *s.t. given two independent sources* X,Y *taking values in* $\mathbb{F}^l$ *and constants* c_1,c_2 *with the property that the sources* $2^{c_1}X-2^{c_1}X$ *and* $2^{c_2}Y-2^{c_2}Y$ *have min-entropy* k_1,k_2*, then* $|\mathbb{E}[\psi(\mathsf{Had}(X,Y))]| \le (|\mathbb{F}^l|2^{-(k_1+k_2)})^{1/2^{c_1+c_2+2}}$ *for every non-trivial character* ψ*.*

Note that $X-X$ has at least as high min-entropy as X, thus if it is convenient we may simply ignore the subtraction part of the hypothesis; it is sufficient to have that $2^{c_1}X, 2^{c_2}Y$ have high min-entropy to apply the above lemma.

3.2.2 Encoding sources to give sources that grow with addition

Given Lemma 3.4 our goal will be to find a way to encode X,Y in such a way that the resulting sources grow with addition. Then we can apply the dot product function and use the lemma to prove that our extractor works. How can we encode a source in a way that guarantees that it grows with addition? Our main weapon to do this will be bounds on the number of line point incidences (Theorem 2.17 or

Theorem 2.18). We will force the adversary to pick a distribution on lines and a distribution on points with high entropy. Then we will argue that if our encoding produces a source which does not grow with addition, the adversary must have picked a set of points and a set of lines that violates the line point incidences theorem.

We will use the following corollary of Theorem 2.17, which is slightly stronger than a theorem due to Zuckerman [Zuc06]. We will follow his proof closely.

Corollary 3.5. *Let $\mathbb{F}$ and $K = 2^{(2+\alpha)k}$ be such that a line point incidences theorem holds for $\mathbb{F}, K$, with α the constant from Theorem 2.17. Suppose L, X are two independent sources, with min-entropy $2k, k$ with L picking an element of $\mathbb{F}^2$ and X picking an element of $\mathbb{F}$ independently. Then the distribution $(X, L(X))$ where $L(X)$ represents the evaluation of the L'th line at X is $2^{-\Omega(k)}$-close to a source with min-entropy $(1 + \alpha/2)2k$.*

Proof. Every source with min-entropy k is a convex combination of sources with min-entropy k and support of size exactly 2^k. So without loss of generality we assume that $\mathsf{supp}(L)$ is of size 2^{2k} and that $\mathsf{supp}(X)$ has size 2^k.

Suppose $(X, L(X))$ is ϵ-far from any source with min-entropy $(1 + \alpha/2)2k$ in terms of statistical distance. Then there must exist some set H of size at most $2^{(1+\alpha/2)2k}$ s.t. $\Pr[(X, L(X)) \in H] \geq \epsilon$.

Then we have

- A set of points H : $2^{2k+k\alpha}$ points
- A set of lines $\mathsf{supp}(L)$: 2^{2k} lines.

Now we get an incidence whenever $(X, L(X)) \in H$. Thus the number of incidences is at least

$$\Pr[(X, L(X)) \in H]|\mathsf{supp}(L)||\mathsf{supp}(X)| \geq \epsilon 2^{3k}$$

However, by the line point incidences theorem (Theorem 2.17), the number of incidences is at most $2^{(3/2-\alpha)(2k+k\alpha)} = 2^{3k+3k\alpha/2-2k\alpha-k\alpha^2} < 2^{3k(1-\alpha/2)} = 2^{-(3k\alpha/2)}2^{3k}$.

These two inequalities imply that $\epsilon < 2^{-(3k\alpha/2)}$.

□

Remark 3.6. The above proof would work even if L, X is a blockwise source with the appropriate min-entropy.

Given this corollary, we now describe several ways to encode a source so that it grows with addition. It suffices to understand any one of these encodings to complete the proof for the extractor.

Encoding 1: $x \mapsto (x, g^x)$ We treat the input x from the source as an element of $\mathbb{F}^*$ for a field in which a version of the line point incidences theorem holds. Then we encode it into an element of $\mathbb{F}^2$ as (x, g^x) where g is a generator of the multiplicative group $\mathbb{F}^*$. Now fix an adversarially chosen source X. Consider the source $\overline{X}$ obtained by performing the above encoding.

$\overline{X}$ is a distribution on points of the form (x, g^x) where $x \neq 0$. By doing a change of variables, we think of every such point as $(\log_g \overline{x}, \overline{x})$.

First consider the distribution of $2\overline{X}$. An element of $\mathsf{supp}(2\overline{X})$ is of the form $(\log_g(\overline{x}_1\overline{x}_2), \overline{x}_1 + \overline{x}_2)$ for some $\overline{x}_1, \overline{x}_2$ in the support of $\overline{X}$. Notice that for each a, b with $a = \overline{x}_1\overline{x}_2$ and $b = \overline{x}_1 + \overline{x}_2$, there are at most two possible values for $(\overline{x}_1, \overline{x}_2)$, since for the solutions for $\overline{x}_1$ must satisfy some quadratic equation in a, b. This means that the min-entropy of $2\overline{X}$ is at least $2k - 1$ since the probability of

getting a particular (a, b) is at most twice the probability of getting a single pair from $\overline{X}, \overline{X}$. By changing k, in the rest of this discussion we assume that the min-entropy of $2\overline{X}$ is $2k$.

Now for each $a, b \in \mathbb{F}$ with $a, b \neq 0$ define the line

$$\begin{aligned}\ell_{a,b} &= \{(ax, b + x) \in \mathbb{F}^2 | x \in \mathbb{F}\} \\ &= \{(x, x/a + b) \in \mathbb{F}^2 | x \in \mathbb{F}\}\end{aligned}$$

Every (a, b) in our encoding then determines the line $\ell_{a,b}$. Let $L = 2\overline{X}$ be a random variable that picks a line according to $2\overline{X}$.

Every element of $\mathsf{supp}(3\overline{X})$ is of the form $(\log_g(\overline{x}_1\overline{x}_2\overline{x}_3), \overline{x}_1 + \overline{x}_2 + \overline{x}_3)$ and determines the point $(\overline{x}_1\overline{x}_2\overline{x}_3, \overline{x}_1 + \overline{x}_2 + \overline{x}_3) \in \mathbb{F}^2$.

Now think of the distribution of $3\overline{X}$ as obtained by first sampling a line according to $2\overline{X}$ and then evaluating that line at an independent sample from $\overline{X}$ and outputting the resulting point. Then we see that we are in a position to apply Corollary 3.5 to get that the encoding does grow with addition.

Encoding 2: $x \mapsto (x, x^2)$ Again we treat x as an element of the multiplicative group of a field $\mathbb{F}^*$ with charactersitic not equal to 2 in which a version of the line point incidences theorem holds. Now fix an adversarially chosen source X. Let $\overline{X}$ denote the source obtained by encoding X in the above way.

First consider the distribution of $2\overline{X}$. An element of $\mathsf{supp}(2\overline{X})$ is of the form $(\overline{x}_1 + \overline{x}_2, \overline{x}_1^2 + \overline{x}_2^2)$ for some $\overline{x}_1, \overline{x}_2$ in the support of $\overline{X}$. Notice that for each a, b with $a = \overline{x}_1 + \overline{x}_2$ and $b = \overline{x}_1^2 + \overline{x}_2^2$, there are at most two possible values for $(\overline{x}_1, \overline{x}_2)$. This means that the min-entropy of $2\overline{X}$ is at least $2k - 1$ since the probability of getting a particular (a, b) is at most twice the probability of getting a single pair from $\overline{X}, \overline{X}$. By changing k, in the rest of this discussion we assume that the min-entropy of $2\overline{X}$ is $2k$.

Now for each $a, b \in \mathbb{F}$ with $a, b \neq 0$ define the line

$$\begin{aligned}\ell_{a,b} &= \{(2ax + b^2 - a, a + x) \in \mathbb{F}^2 | x \in \mathbb{F}\} \\ &= \{(x, x/(2a) + (1 - b^2/(2a))) \in \mathbb{F}^2 | x \in \mathbb{F}\}\end{aligned}$$

Every (a, b) in our encoding then determines a unique line $\ell_{a,b}$. Let $L = 2\overline{X}$ be a random variable that picks a line according to $2\overline{X}$.

Every element of $\mathsf{supp}(3\overline{X})$ is then of the form $(\overline{x}_1 + \overline{x}_2 + \overline{x}_3, \overline{x}_1^2 + \overline{x}_2^2 + \overline{x}_3^2)$ and determines the point

$$\begin{aligned}&((\overline{x}_1 + \overline{x}_2 + \overline{x}_3)^2 - (\overline{x}_1^2 + \overline{x}_2^2 + \overline{x}_3^2), \overline{x}_1 + \overline{x}_2 + \overline{x}_3) \\ = (2(\overline{x}_1 + \overline{x}_2)\overline{x}_3 &+ (\overline{x}_1 + \overline{x}_2)^2 - (\overline{x}_1 + \overline{x}_2), (\overline{x}_1 + \overline{x}_2) + \overline{x}_3) \\ &= (2a\overline{x}_3 + b^2 - a, a + \overline{x}_3)\end{aligned}$$

Now think of the distribution of $3\overline{X}$ as obtained by first sampling a line according to $2\overline{X}$ and then evaluating that line at an independent sample from $\overline{X}$ and outputting the resulting point. Then we see that we can apply Corollary 3.5 to get that the encoding does grow with addition.

Conclusion By picking an appropriate constant γ, we obtain the following lemma:

Lemma 3.7. *There is a universal constant γ s.t. if X is any source that picks an element of $\mathbb{F}$ with min-entropy $(1/2 - \gamma)\log|\mathbb{F}|$, $3\overline{X}$ is $|\mathbb{F}|^{-\Omega(1)}$-close to a source with min-entropy $(1/2 + \gamma)\log|\mathbb{F}^2|$.*

3.2.3 Putting things together

Putting together the results from the two previous sections and applying Lemma 4.1, we obtain the theorem for Bourgain's extractor.

Theorem 3.8. *[Bou05] There exists a univeral constant $\gamma > 0$ and a polynomial time computable function* $\mathsf{Bou} : (\{0,1\}^n)^2 \to \{0,1\}^m$ *s.t. if X, Y are two independent $(n, (1/2-\gamma)n)$ sources, $\mathbb{E}_Y[\|\mathsf{Bou}(X,Y) - U_m\|_{\ell^1}] < \epsilon$, with $\epsilon = 2^{-\Omega(n)}, m = \Omega(n)$.*

4 XOR lemma for abelian groups

In this subsection we will prove a generalization of Vazirani's XOR lemma. This lemma appears to be folklore, but we were unable to find a proof written down anywhere.

Throughout this section we reserve G for a finite abelian group.

The lemma we will prove is the following:

Lemma 4.1 (XOR lemma for cyclic groups)**.** *For every cyclic group $G = \mathbb{Z}_N$ and every integer $M \leq N$, there is an efficiently comptable function $\sigma : \mathbb{Z}_N \to \mathbb{Z}_M = H$ with the following property: Let X be any random variable taking values in $\mathbb{Z}_N$ s.t. for every non-trivial character $\psi : \mathbb{Z}_N \to \mathbb{C}^*$, we have $|\mathbb{E}[\psi(X)]| < \epsilon$, then $\sigma(X)$ is $O(\epsilon \log N\sqrt{M}) + O(M/N)$ close to the uniform distribution.*

It turns out that it is easy to extend this result to work for *any* abelian group G, though it's hard to state the result for general abelian groups in a clean way. In this section we will discuss the proof of the above lemma and just make a few remarks about how to extend it to general abelian groups.

Before we move on to prove Lemma 4.1, let us first prove a special case of this lemma which is a generalization of Vazirani's XOR lemma. For the proof of this case below, we essentially follow the proof as in Goldreich's survey [Gol95].

Lemma 4.2. *X be a distribution on a finite abelian group G s.t. $|\mathbb{E}[\psi(X)]| \leq \epsilon$ for every non-trivial character ψ. Then X is $\epsilon\sqrt{|G|}$ close to the uniform distribution: $\|X - U\|_{\ell^1} \leq \epsilon\sqrt{|G|}$.*

Proof. By the hypothesis, for every non-trivial character ψ of G, $|\langle \psi, X\rangle| = (1/|G|)|\mathbb{E}[\psi(X)]| \leq \epsilon/|G|$. In other words, $\|\widehat{X - U}\|_{\ell^\infty} \leq \epsilon/|G|$. Then by Proposition 2.9, we get that $\|X - U\|_{\ell^1} \leq \epsilon\sqrt{|G|}$. □

In Lemma 4.2, given a bound of ϵ on the biases, the statistical distance blows up by a factor of $\sqrt{|G|}$. This is too much if ϵ is not small enough. Lemma 4.1 gives us the flexibility to tradeoff this blowup factor with the number of bits that we can claim are statistically close to uniform. As M is made smaller, the blowup factor is reduced, but we get "less" randomness. Our proof for the general case will work (more or less) by reducing to the case of Lemma 4.2.

Note that if σ is an onto homomorphism, for every non-trivial character ϕ of H, $\phi \circ \sigma$ is a non-trivial character of G. Thus the bounds on the biases of X give bounds on the biases of $\sigma(X)$ and we can reduce to the case of Lemma 4.2. The problem is that we cannot hope to find such a homomorphism σ for every M. For instance, if $G = \mathbb{Z}_p$ for p a large prime, G contains no non-trivial subgroup and so σ cannot be a homomorphism for $M = \lceil p/2 \rceil$. Instead, we will show that we can find a σ which *approximates* a homomorphism in the sense:

- For every non-trivial character ϕ of H, $\phi \circ \sigma$ is approximated by a *few* characters of G. Formally, this is captured by bounding $\|\widehat{\phi \circ \sigma}\|_{L^1}$ (observe that if σ is a homomorphism, this quantity is $1/|G|$).

- We'll ensure that $\sigma(U)$ is the close to the uniform distribution on H.

Then we will be able to use the bounds on the biases of X to give bounds on the biases of $\sigma(X) - \sigma(U)$, where U is the uniform distribution. This will allow us to apply Proposition 2.9 to conclude that X is a pseudorandom generator for σ, i.e. $\|\sigma(X) - \sigma(U)\|_{\ell^1}$ is small, which implies that $\sigma(X)$ is close to uniform, since $\sigma(U)$ is close to uniform.

The following lemma asserts that every ϵ-biased distribution is pseudorandom for any function σ that satisfies the first condition above.

Lemma 4.3. *Let G, H be finite abelian groups. Let X be a distribution on G with $|\mathbb{E}_X[\psi'(X)]| \leq \epsilon$ for every non-trivial character ψ' of G and let U be the uniform distribution on G. Let $\sigma : G \to H$ be a function such that for every character ϕ of H, we have that*

$$\|\widehat{\phi \circ \sigma}\|_{L^1} \leq \tau/|G|$$

Then $\|\sigma(X) - \sigma(U)\|_{\ell^1} < \tau\epsilon\sqrt{|H|}$.

Proof. First note that the assumption on X is equivalent to $\|\widehat{X - U}\|_{\ell^\infty} \leq \epsilon/|G|$. Let ϕ be any non-trivial character of H. Then

$$\begin{aligned}
&|\langle \phi, \sigma(X) - \sigma(U)\rangle| \\
&= (1/|H|)|\mathbb{E}_{\sigma(X)}[\phi(\sigma(X))] - \mathbb{E}_{\sigma(U)}[\phi(\sigma(U))]| \\
&= \frac{|G|}{|H|}|\langle \phi \circ \sigma, X - U\rangle| \\
&= \frac{|G|^2}{|H|}|\langle \widehat{\phi \circ \sigma}, \widehat{X - U}\rangle| && \text{by preservation of inner product (Fact 2.11)} \\
&\leq \frac{|G|^2}{|H|}\|\widehat{\phi \circ \sigma}\|_{L^1}\|\widehat{X - U}\|_{\ell^\infty} && \text{by the triangle inequality (Fact 2.3)} \\
&\leq \tau\epsilon/|H|
\end{aligned}$$

On the other hand, $\langle 1, \sigma(X) - \sigma(U)\rangle = 0$, since X and U are distributions. Thus, we have shown that $\|\widehat{\sigma(X) - \sigma(U)}\|_{\ell^\infty} \leq \tau\epsilon/|H|$, which by Proposition 2.9 implies that $\|\sigma(X) - \sigma(U)\|_{\ell^1} \leq \tau\epsilon\sqrt{|H|}$. □

Note that when σ is the identity function (or any surjective homomorphism onto a group H), $\tau = 1$. Thus Vazirani's XOR lemma corresponds exactly to the case of σ being the identity function.

Next we show that in the special when G is a cyclic group, we can find a σ which satisfies the hypothesis of Lemma 4.3 with small τ.

Lemma 4.4. *Let M, N be integers satisfying $N > M$. Let $\sigma : \mathbb{Z}_N \to \mathbb{Z}_M$ be the function $\sigma(x) = x \bmod M$. Then for every character ϕ of $\mathbb{Z}_M$, $\|\widehat{\phi \circ \sigma}\|_{L^1} \leq O(\log N)/N$*

Proof. Note that if M divides N, the statement is trivial, since σ is a homomorphism. Below we show that even in the general case, this expectation is small. Define the function $\rho(x) = \exp(2\pi\imath x)$. Then note that $\rho(a + b) = \rho(a)\rho(b)$.

First let ϕ be any character of $\mathbb{Z}_M$. Then $\phi(y) = \rho(wy/M)$ for some $w \in \mathbb{Z}_M$. Clearly, $\phi(\sigma(x)) = \rho(wx/M)$.

$$
\begin{aligned}
&\|\widehat{\phi \circ \sigma}\|_{L^1} \\
&= (1/N^2) \sum_{t \in \mathbb{Z}_N} \left| \sum_{x \in \mathbb{Z}_N} \rho(tx/N)\rho(-wx/M) \right| \\
&= (1/N^2) \sum_{t \in \mathbb{Z}_N} \left| \sum_{x \in \mathbb{Z}_N} \rho\left(\frac{x(tM - wN)}{NM} \right) \right|
\end{aligned}
$$

Recall that for any geometric sum $\sum_{i=0}^{N} br^i = \frac{br^N - b}{r-1}$, as long as $r \neq 1$. The inner sum in this expression is exactly such a geometric sum. Thus we get:

$$
\begin{aligned}
&\|\widehat{\phi \circ \sigma}\|_{L^1} \\
&\leq (1/N^2) \sum_{t \in \mathbb{Z}_N, t \neq wN/M} \left| \sum_{x \in \mathbb{Z}_N} \rho\left(\frac{x(tM - wN)}{NM} \right) \right| + 1/N \\
&= (1/N^2) \sum_{t \in \mathbb{Z}_N, t \neq wN/M} \left| \frac{\rho(\frac{N(tM-wN)}{NM}) - 1}{\rho(\frac{tM-wN}{NM}) - 1} \right| + 1/N && \text{by simplifying the geometric sum} \\
&\leq (1/N^2) \sum_{t \in \mathbb{Z}_N, t \neq wN/M} \left| \frac{2}{\rho(\frac{tM-wN}{NM}) - 1} \right| + 1/N && \text{since } \left| \rho\left(\frac{N(tM - wN)}{NM} \right) - 1 \right| \leq 2 \\
&\leq (1/N^2) \sum_{t \in \mathbb{Z}_N, t \neq wN/M} \left| \frac{2}{\rho(\frac{t-(wN/M)}{N}) - 1} \right| + 1/N
\end{aligned}
$$

Now write $wN/M = c + d$, where c is an integer, and $d \in [0, 1]$. Then, by doing a change of variable from t to $t - c$, we get that the above sum is

$$
(1/N^2) \sum_{t \in \mathbb{Z}_N, t \neq d} \left| \frac{2}{\rho(\frac{t-d}{N}) - 1} \right| + 1/N
$$

We will bound two parts of this sum separately. Let r be a constant with $0 < r < 1/4$. Now note that $|\rho(\frac{t-d}{N}) - 1| \geq \Omega(1)$ when $rN < t < (1-r)N$, since in this situation the quantity is the distance between two points on the unit circle which have an angle of at least $2\pi r$ between them.

When t is not in this region, $|\rho(\frac{t-d}{N}) - 1| \geq |\sin(2\pi(t-d)/N)|$, since the sin function gives the vertical distance between the two points. This is at least $(t-d)/100N$ for r small enough, since we have that $|\sin x| > |x|$ for $-\pi/2 < x < \pi/2$. Thus, choosing r appropriately, we can bound the sum:

$$
\begin{aligned}
&(1/N^2)\sum_{t\in\mathbb{Z}_N, t\neq d}\left|\frac{2}{\rho(\frac{t-d}{N})-1}\right| + 1/N \\
&= (1/N^2)\left(\sum_{t\neq d, t\in[rN,(1-r)N]}\left|\frac{2}{\rho(\frac{t-d}{N})-1}\right| + \sum_{t\neq d, t\notin[rN,(1-r)N]}\left|\frac{2}{\rho(\frac{t-d}{N})-1}\right|\right) + 1/N \\
&\leq (1/N^2)\left(\sum_{t\neq d, t\in[0,rN]}\frac{800N}{t-d} + \sum_{t\neq d, t\notin[rN,(1-r)N]} O(1)\right) + 1/N \\
&\leq (1/N^2)(O(N\log N) + O(N)) + 1/N
\end{aligned}
$$

Here the last inequality used the fact that $\sum_{i=1}^{n} 1/i = O(\log n)$. Overall this gives us a bound of $\tau \leq O(\log N/N)$.

□

On uniform input the distribution $\sigma(U)$ is quite close to uniform. Specifically, if $N = qM + r$, with q, r the quotient and remainder of N on dividing by N, we have that $\sigma(U)$ is $2r((q+1)/N - 1/M) = (2r/M)(M(q+1)/N - 1) = (2r/M)(M-r)/N = 2M/N$ close to the uniform distribution. Thus, overall we get that this σ turns any distribution which fools characters with bias at most ϵ into one that is $\epsilon \log N\sqrt{M} + O(M/N)$ close to uniform.

Now we discuss the situation for general abelian groups. The basic observation is that approximate homomorphisms can be combined to give a new approximate homomorphism:

Lemma 4.5. *Let $G = G_1 \oplus G_2$ and $H = H_1 \oplus H_2$ be finite abelian groups. Let $\sigma_1 : G_1 \to H_1$ and $\sigma_2 : G_2 \to H_2$ be two functions that satisfy the hypotheses of Lemma 4.3 with constants τ_1 and τ_2 respectively. Then the function $\sigma : G \to H$ defined as $\sigma(x \oplus y) = \sigma_1(x) \oplus \sigma_2(y)$ satisfies the hypotheses of the lemma with parameters $\tau_1\tau_2$.*

Given this lemma, it is clear how to get an xor lemma for every abelian group. Simply write the abelian group as a direct sum of cyclic groups. Then depending on how much randomness is needeed, we can compose several homomorphisms with approximate homorphisms to get a function σ that does the job.

5 Strong 2 source extractors vs low error 2 source extractors

In this subsection we give an argument due to Boaz Barak showing that every 2 source extractor which has sufficiently small error is in fact strong.

Theorem 5.1. *Let $\mathsf{IExt} : (\{0,1\}^n)^2 \to \{0,1\}^m$ be any two source extractor for min-entropy k with error ϵ. Then IExt is a strong two source extractor for min-entropy k' (strong with respect to both sources) with error $2^m(\epsilon + 2^{k-k'})$.*

Proof. Without loss of generality, we assume that X, Y have supports of size k'. Then we need to bound:

$$\sum_{y\in\mathsf{supp}(Y)} 2^{-k'}\|\mathsf{IExt}(X,y) - U_m\|_{\ell^1}$$

For any $z \in \{0,1\}^m$, define the set of bad y's for z

$$B_z = \{y : |\Pr[\mathsf{IExt}(X,y) = z] - 2^{-m}| \geq \epsilon\}$$

Claim 5.2. *For every z, $|B_z| < 2^k$*

Suppose not, then the flat distributions on B_z, X are two independent sources for which the extractor IExt fails. Now let $B = \cup_z B_z$. We see that $|B| < 2^k 2^m$. Thus,

$$\begin{aligned}
&\sum_{y \in \mathsf{supp}(Y)} 2^{-k'} \|\mathsf{IExt}(X,y) - U_m\|_{\ell^1} \\
&= \sum_{y \in \mathsf{supp}(Y) \cap B} 2^{-k'} \|\mathsf{IExt}(X,y)q - U_m\|_{\ell^1} + \sum_{y \in \mathsf{supp}(Y) \setminus B} 2^{-k'} \|\mathsf{IExt}(X,y) - U_m\|_{\ell^1} \\
&\leq 2^{-k'} 2^{k+m} + \epsilon 2^m \\
&= 2^m (2^{k-k'} + \epsilon)
\end{aligned}$$

■

6 Acknowledgements

I'd like to thank Avi Wigderson, who helped me understand the view of Bourgain's extractor presented here. Thanks to Boaz Barak for the argument showing that every two source extractor must be strong. Thanks to David Zuckerman and Swastik Kopparty for useful discussions.

References

[BIW04] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting Randomness Using Few Independent Sources. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 384–393, 2004.

[BKS+05] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating Independence: New Constructions of Condensers, Ramsey Graphs, Dispersers, and Extractors. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 1–10, 2005.

[BRSW06] B. Barak, A. Rao, R. Shaltiel, and A. Wigderson. 2 Source Dispersers for $n^{o(1)}$ Entropy and Ramsey Graphs beating the Frankl-Wilson Construction. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.

[Bou05] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1:1–32, 2005.

[BKT04] J. Bourgain, N. Katz, and T. Tao. A Sum-Product Estimate in Finite Fields, and Applications. *Geometric and Functional Analysis*, 14:27–57, 2004.

[CG88] B. Chor and O. Goldreich. Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[Gol95] O. Goldreich. Three XOR-Lemmas - An Exposition. *Electronic Colloquium on Computational Complexity (ECCC)*, 2(56), 1995.

[Kon03] S. Konyagin. A Sum-Product Estimate in Fields of Prime Order. Technical report, Arxiv, 2003. `http://arxiv.org/abs/math.NT/0304217`.

[Rao06] A. Rao. Extractors for a Constant Number of Polynomially Small Min-entropy Independent Sources. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.

[Raz05] R. Raz. Extractors with Weak Random Seeds. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 11–20, 2005.

[SV86] M. Santha and U. V. Vazirani. Generating Quasi-Random Sequences from Semi-Random Sources. *Journal of Computer and System Sciences*, 33:75–87, 1986.

[Vaz85] U. Vazirani. Towards a Strong Communication Complexity Theory or Generating Quasi-Random Sequences from Two Communicating Slightly-random Sources (Extended Abstract). In *Proceedings of the 17th Annual ACM Symposium on Theory of Computing*, pages 366–378, 1985.

[Zuc06] D. Zuckerman. Linear Degree Extractors and the Inapproximability of Max Clique and Chromatic Number. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.