Randomness Extractors

Alex Block

Purdue University

April 25, 2016

Table of Contents

Preliminaries

- $N = 2^n$, $K = 2^k$, $M = 2^m$
- If X is a random variable, then $H_\infty(X)$ is the min-entropy of X.
  - $H_\infty(X) \geq k \iff \Pr[X = x_i] \leq 1/K \ \forall x_i \in X$
- $\mathcal{U}_n$ is the uniform distribution over $\{0,1\}^n$

Definition (Statistical Distance)

If X and Y are random variables over the same sample space Ω, then the Statistical Distance of X and Y is:

$$\mathrm{SD}(X,Y) = \frac{1}{2}\sum_{z\in\Omega} |X(z) - Y(z)|$$

- The SD is one measure of "closeness" between distributions.
- Intuitively, if you sample from a distribution X and there is a distribution Y such that $\mathrm{SD}(X,Y)$ is small, then you can sample from Y instead.

Preliminaries

- Let X be a distribution over $\{0,1\}^n$ and let $W \subseteq \{0,1\}^n$ of size T. Then, X is a T-flat over W if it is uniformly distributed over W.

Lemma (Convexity of high min-entropy sources) [Vad12, p. 173]

Let X be a distribution over $\{0,1\}^n$ with $H_\infty(X) \geq k$. Then $X = \sum \alpha_i X_i$ where $0 \leq \alpha_i \leq 1$, $\sum \alpha_i = 1$, and each X_i is a K-flat.

- Intuitively, this lemma says that instead of working with arbitrary high min-entropy sources, it suffices to work with arbitrary flat sources of the same min-entropy.

What is an Extractor?

Definition (Extractor)

Given a min-entropy source X of size N, a function $\text{Ext}: \{0,1\}^n \to \{0,1\}^m$ is said to be an extractor if $\text{Ext}(X)$ is *close* to $\mathcal{U}_m$.

- By close, we mean that the SD is small.
- Like any function, can think of an extractor as a matrix (in this case, an $n \times m$ matrix).
- Note that random matrices make good extractors at the cost of being large.
  - We would like constructions of extractors that are smaller than just a random matrix.
  - Until recently, explicit constructions were difficult and were still outperformed by random matrices.

Random Matrices are Good Extractors

Proposition (Extractors on K-flats) [Vad12, p. 175]

For every $n, k, m \in \mathbb{N}$, every $\epsilon > 0$, and every K-flat source X, choosing a random $\text{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ with $m = k - 2\log(1/\epsilon) - \mathcal{O}(1)$, then $\text{Ext}(X)$ will be ϵ-close to $\mathcal{U}_m$ with probability $1 - 2^{\Omega(K\epsilon^2)}$.

- Recall that every extractor can be represented as a matrix.
- This proposition tells us that random $n \times m$ matrices are good extractors over K-flats (with exponentially high probability).
- Recall our lemma:

Lemma (Convexity of high min-entropy sources) [Vad12, p. 173]

Let X be a distribution over $\{0,1\}^n$ with $H_\infty(X) \geq k$. Then $X = \sum \alpha_i X_i$ where $0 \leq \alpha_i \leq 1$, $\sum \alpha_i = 1$, and each X_i is a K-flat.

- Thus we can extend this proposition to apply to any arbitrary high min-entropy source.

Seeded Extractors

Definition (Seeded Extrator) [Vad12, p. 176]

A function $\text{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ϵ)-*extractor* if for every k-source X on $\{0,1\}^n$, $\text{Ext}(X,\mathcal{U}_d)$ is ϵ-close to $\mathcal{U}_m$.

- Seeded extractors require a small amount of uniform randomness (the seed).
  - Question: Where do we get the initial amount of randomness?
  - If we happen to have some randomness, then we can extract a large amount of randomness with these extractors.
- Note for these extractors, we do not reveal the seed.

Seeded Extractors

Theorem (Nice Seeded Extractors Exist) [Vad12, p. 176-177]

For every $n \in \mathbb{N}$, $k \in [0, n]$, and $\epsilon > 0$, there exists a (k, ϵ)-extractor $\text{Ext}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m = (k + d) - 2\log(1/\epsilon) - \mathcal{O}(1)$ and $d = \log(n - k) + 2\log(1/\epsilon) + \mathcal{O}(1)$.

- Intuitively, if we have a source with k-bits of randomness and another uniform source of d-bits, then we can extract $(k + d) = m$-bits of randomness.
  - Larger d allows us to extract more randomness.
  - Potentially much greater than n bits of randomness.
  - $(n - k)$ term in d is to account for the "bad bits" of the input source.
- As long as you have enough pure randomness, a nice seeded extractor will always exist.
  - With a small amount of randomness, you can potentially generate very large amounts of randomness that are close to uniform.
  - Use output of one seeded extractor as a seed for another seeded extractor.

Deterministic Extractors

- What if we do not have a $\mathcal{U}_d$-source for a seed?
- Want extractors that rely only on min-entropy sources and not on any additional randomness.
- Deterministic Extractors are the solution.

Definition (Deterministic Extractor) [Gab11, p. 2]

Let $\mathcal{C}$ be a class of distributions over $\{0,1\}^n$. A function $\text{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ is a *deterministic ϵ-extractor* for $\mathcal{C}$ if for every distribution $X \in \mathcal{C}$ the distribution $\text{Ext}(X)$ is ϵ-close to $\mathcal{U}_m$ over $\{0,1\}^m$.

- We consider the case where $\mathcal{C}$ is the class of high min-entropy distributions.

1-Source Deterministic Extractors

- A 1-Source Extractor simply means that Ext only requires a single min-entropy source X
- Question: Is there a 1-Source extractor for all high min-entropy sources?
  - This is impossible!

1-Source Deterministic Extractors

- How bad is it? **We cannot even hope to extract a single bit!**
  - Suppose someone claims that they do have an extractor Ext that extracts one bit from every high min-entropy source of size N.
  - Consider the pre-images of 1 and 0 under this extractor.
  - One of these pre-images must be at least half the input size: $N/2 = 2^{(n-1)}$.
  - Consider a uniform distribution X over this pre-image.
  - Then $H_\infty(X) \geq (n-1)$ but $\text{Ext}(X)$ is a constant.
- Note that if the input source has some structure, then we can construct extractors for these types of sources.

2-Source Deterministic Extractors

- We want to develop extractors that have the minimal amount of assumptions on the input source.
- We have shown that this cannot be done with 1-Source Extractors.
- Suppose we are given two independent high min-entropy sources. Can we use both as input to an extractor?

2-Source Deterministic Extractors

- Yes we can! This gives us 2-Source Deterministic Extractors.

Definition (2-Source Extractor)

Given two independent min-entropy distribution X and Y such that $H_\infty(X) \geq k$ and $H_\infty(Y) \geq k$, let $\text{Ext}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$. If $\text{Ext}(X, Y)$ is close to $\mathcal{U}_m$, then we say Ext is a *2-source extractor for min-entropy k*.

- Ideally, we want k to be as small as possible compared to n.
  - There is a known lower bound: k must be at least $\text{polylog}(n)$.

History of 2-Source Extractors

1988 First construction by Chor and Goldreich with min-entropy requirement $k = n/2$ using Lindsey's Lemma. [CG88]

2005 Bourgain's Extractor was the first to break the $n/2$ barrier thanks to advances in arithmetic combinatorics, with $k = .499n$ (approximately) and outputting $m = \Omega(n)$ bits. [Bou05]

July 2015 Chattopadhyay and Zuckerman create an explicit 2-source extractor, with min-entropy requirements of $\log^C n$ for large enough constant C. This extractor outputs 1 bit. [CZ15]

August 2015 Xin Li improves upon Chattopadhyay and Zuckerman's construction by increasing the number of output bits to $k^{\Omega(1)}$. [Li15]

Bourgain's 2-Source Extractor

- Bourgain's Extractor broke the nearly 2 decade long barrier of $k = n/2$ min-entropy requirement.
- The heart of Bourgain's construction comes from using the Hadamard Extractor
  - The Hadamard Extractor works fairly well for many sources.
  - However, there are certain sources that when used for the Hadamard Extractor, it completely fails.
  - Bourgain realized that these cases were pathological and could be remedied by "fixing" the input source.
  - Bourgain described an encoding that, when done on any source with high enough min-entropy, the Hadamard Extractor with the encoded source would always succeed.

Conceptual Picture of Bourgain's Extractor

X and Y are sources over $\{0,1\}^n$ each with min-entropy k:

$$X, Y \xrightarrow{\text{Encoding}} X', Y' \xrightarrow{\text{Hadamard}} \mathcal{U}_m$$

Bourgain's 2-Source Extractor

Theorem (Bourgain's Extractor)

There exists a universal constant $\gamma > 0$ and a polynomial time computable function $\mathrm{Bou}\colon (\{0,1\}^n)^2 \to \{0,1\}^m$ such that if X and Y are two independent $(n, (1/2-\gamma)n)$ sources, then $\mathop{\mathbb{E}}_{y\sim Y}[\|\mathrm{Bou}(X,y) - \mathcal{U}_m\|_{\ell 1}] < \epsilon$, with $\epsilon = 2^{-\Omega(n)}$ and $m = \Omega(n)$.

- Bourgain's Extractor tells us the following:
  - Suppose we are given two independent sources over $\{0,1\}^n$ with min-entropy *less than* $n/2$.
  - Then there is a function that extracts m bits of randomness.
  - The *average* difference of the output distribution and $\mathcal{U}_m$ is exponentially small in n.

The Hadamard Extractor

- The Hadamard Extractor is the core of Bourgain's Extractor.

Definition (The Hadamard Function)

Let $\mathbb{F}$ be a finite field. Define Had: $\mathbb{F}^l \times \mathbb{F}^l \to \mathbb{F}$ as the "dot-product" function $\text{Had}(x, y) = x \cdot y$.

- Note that this "dot-product" is more general; it is interpreted as a field-level multiplication.
- If we wish to use the Hadamard Extractor to extract m bits from n bit sources, we choose $\mathbb{F}$ such that $|\mathbb{F}| = M$ and l so that $|\mathbb{F}|^l = N$.
  - This leads us to a theorem:

The Hadamard Extractor

Theorem (The Hadamard Extractor)

For every constant $\delta > 0$, there exists a polynomial time algorithm $\mathrm{Had}\colon (\{0,1\}^n)^2 \to \{0,1\}^m$ such that if X, Y are independent sources of size n with min-entropy $(1/2 + \delta)n$, then $\mathop{\mathbb{E}}_{y \sim Y}[\|\mathrm{Had}(X, y) - \mathcal{U}_m\|_{\ell^1}] < \epsilon$ with $m = \Omega(n)$ and $\epsilon = 2^{-\Omega(n)}$.

- If we are given two independent sources with min-entropy slightly above $n/2$, then on average the Hadamard Extractor will be close to uniform.
  - This closeness is exponentially small in n.
- One may ask: why not just use the Hadamard Extractor by itself?

The Hadamard Extractor

- The Hadamard Extractor by itself is not good enough.
  - There are weird (pathological as Bourgain described it) cases where Had fails.
  - For example, if l is large, take X and Y such that both have min-entropy exactly $n/2$ such that X picks the first half of its field elements randomly and sets the rest to 0 and Y picks the second half of its field elements randomly and sets the rest to 0. Then the output of the dot product function is always 0.
  - As another example, take $l = 1$. Then the output must have at least n bits of entropy to generate a uniformly random point in $\mathbb{F}$ (so we have not achieved anything).
- Can we fix this so we can use the Hadamard Extractor for any source?

Encoding the Source Input

- Bourgain recognized that the Hadamard Extractor works very well for most sources.
  - He shows that there are essentially very few counter examples for which it fails.
- Bourgain then demonstrated how to encode any general source in away that ensures the Hadamard Extractor does not fail.
  - He showed that when input sources meet certain requirements, the Hadamard extractor will always succeed.
  - He showed also that you can encode any source with sufficient min-entropy and the resulting source will not fail for the Hadamard Extractor.

Encoding the Source Input

- In a sense, the Hadamard Extractor fails because of linearity in the input.
- Idea for encoding: introduce non-linearity and interpret the encoding as an element in a higher-order field.
- One such encoding:

$$x \mapsto (x, x^2)$$

Encoding the Source Input $[x \mapsto (x, x^2)]$

- Consider an adversarially chosen source X. Let $\overline{X}$ be the source obtained by the performing the encoding.
- Now consider the distribution $2\overline{X}$ obtained by taking two independent samples of $\overline{X}$ and summing the samples.
  - An element in $\text{supp}(2\overline{X})$ is of the form $(\overline{x}_1 + \overline{x}_2, \overline{x}_1^2 + \overline{x}_2^2)$.
  - This gives us that there are at most two possible values for $(\overline{x}_1, \overline{x}_2)$.
  - Each of these elements defines a line in $\mathbb{F}^2$.

Encoding the Source Input $[x \mapsto (x, x^2)]$

- Now consider the distribution $3\overline{X}$.
  - Every element of $3\overline{X}$ is of the form $(\overline{x}_1 + \overline{x}_2 + \overline{x}_3, \overline{x}_1^2 + \overline{x}_2^2 + \overline{x}_3^2)$
  - This determines a point in $\mathbb{F}^2$:
    $((\overline{x}_1 + \overline{x}_2 + \overline{x}_3)^2 - (\overline{x}_1^2 + \overline{x}_2^2 + \overline{x}_3^2), \overline{x}_1 + \overline{x}_2 + \overline{x}_3)$
- We think of sampling from the distribution $3\overline{X}$ as the following:
  - First, we sample a line from $2\overline{X}$, then we independently sample a point from $\overline{X}$ and evaluate that point on that line.
  - We output the resulting point.
- These distributions $2\overline{X}$, $\overline{X}$ allow the Hadamard Extractor to always succeed.
- Note this is just one example of an encoding!

Statement of Bourgain's Extractor

Theorem (Bourgain's Extractor)

There exists a universal constant $\gamma > 0$ and a polynomial time computable function $\mathrm{Bou}\colon (\{0,1\}^n)^2 \to \{0,1\}^m$ such that if X and Y are two independent $(n, (1/2-\gamma)n)$ sources, then $\mathop{\mathbb{E}}_{y \sim Y}[\|\mathrm{Bou}(X,y) - \mathcal{U}_m\|_{\ell^1}] < \epsilon$, with $\epsilon = 2^{-\Omega(n)}$ and $m = \Omega(n)$.

References I

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

[CG88] Benny Chor and Oded Goldreich.
Unbiased bits from sources of weak randomness and probabilistic communication complexity.
SIAM Journal on Computing, 17(2):230–261, 1988.

[CZ15] Eshan Chattopadhyay and David Zuckerman.
Explicit two-source extractors and resilient functions.
Technical Report TR15-119, Electronic Colloquium on Computational Complexity, 2015.

References II

[Gab11] Ariel Gabizon.
Deterministic Extraction from Weak Random Sources.
Springer, 2011.

[Li15] Xin Li.
Improved constructions of two-source extractors, 2015.
`http://arxiv.org/abs/1508.01115`.

[Rao07] Anup Rao.
An exposition of bourgain's 2-source extractor.
Electronic Colloquium on Computational Complexity (ECCC), 14, 2007.

[Vad12] Salil Vadhan.
Pseudorandomness.
Foundations and Trends in Theoretical Computer Science, 7(1-3):1–336, 2012.

Questions?

Any final questions?

Thank you!

Special thanks to Prof. Maji for helping me when I had questions!