Randomness Extractors Seminar **December 21, 2016**

Explicit Construction of 2-source Extractors - Overview

Leader: KM Chung **Notes: Chi-Ning Chou**

We are going to see that motivation and construction overview of 2-source extractors.

1 Motivation and Related Works

We've seen the explicit construction of seeded extractors from Vadhan et. al. and Trevisan. Now, it's natural to ask: Can we extract randomness without any uniform bits? As we know that there cannot exist extractor for single weak source, a tempting goal is to explicitly construct extractors for more than one independent weak sources. While 2 sources is the least we can hope for, we formally define the extractors we want as follows.

Definition 1 (2-source extractors)**.** *$Ext : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ is a 2-source extractor for min-entropy $k \in [n]$ with error $\epsilon > 0$ if for any (n,k)-sources X and Y,*

$$|Ext(X,Y) - U_m| \leq \epsilon. \tag{1}$$

For simplicity, here we consider the case where $m = 1$, *i.e.,* , an one bit 2-source extractor.

There's an immediate existence result using probabilistic argument that when $k \geq \log n + \log 1/\epsilon + 1$.

Theorem 2 (existence of good 2-source extractors)**.** *Exists 2-source extractors for with error ϵ for weak sources of min-entropy $k \geq \log n + \log 1/\epsilon + 1$.*

Proof. First, think of a 2-source extractor as an $N \times N$ matrix A, where the columns and the rows are corresponding to the range of X and Y respectively. Since considering flat k-source is sufficient, the 2-source extractor has error ϵ if any size $K \times K$ submatrix of A has $\frac{1}{2} \pm \epsilon$ fraction of 1's. Now, we pick the entries of A uniformly random from $\{0,1\}$. Compute the probability that a single $K \times K$ submatrix B does not has $\frac{1}{2} \pm \epsilon$ fraction of 1's. By Chernoff's bound, we have

$$\mathbb{P}[B \text{ does not has } \frac{1}{2} \pm \epsilon \text{ fraction of 1's}] \leq 2\exp^{-K^2\epsilon^2/4}. \tag{2}$$

Since there are at most $\binom{N}{K} \times \binom{N}{K} \leq (\frac{Ne}{K})^{2K}$ such sub-matrices of A, by union bound

$$\mathbb{P}[\forall \text{ size } K \times K \text{ sub-matrix } B \text{ does not has } \frac{1}{2} \pm \epsilon \text{ fraction of 1's}] \leq 2\exp^{2K\log(Ne/K) - K^2\epsilon^2/4}. \tag{3}$$

When taking $k \geq \log n + \log 1/\epsilon + 1$, $2\exp^{2K\log(Ne/K) - K^2\epsilon^2/4} < 1$. Thus, we know that there exists a 2-source extractors for with error ϵ for weak sources of min-entropy $k \geq \log n + \log 1/\epsilon + 1$. $\square$

Before this work [CZ15], the best we can achieve is listed as follows.

Reference	k_1	k_2
[CG88]	$> 0.5n$	$> 0.5n$
[Bou05]	$\geq 0.499n$	$\geq 0.499n$
[Raz05]	$\geq 0.5n$	$O(\log n)$

For more sources, there are some better results.

- [BIW06]: constantly many (n, k) sources with min-entropy $k = \delta n$
- [Rao09]: constantly many (n, k) sources with min-entropy $k = n^\gamma$
- [Li12]: 3 sources with min-entropy $n^{0.51}$
- [Li15b]: 3 sources with min-entropy $\log^C n$

The main result of this work is to achieve the log barrier as follows.

- [CZ15]: Explicit 2-source extractors for min-entropy $k \geq \log^C n$ outputting 1 bit.
- [Li15a]: Explicit 2-source extractors for min-entropy $k \geq \log^C n$ outputting $0.9k_1$ bits.

2 The Journey to 2-source Extractors

To motivate the final construction, here I follow the storyline in the original paper.

2.1 A wild start from strong seeded extractors

Recall that we had seen the strong seeded extractors $Ext : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ by Salil et. al. which require only length $O(\log n)$ seed. Suppose Ext works for k-source with error ϵ and output only one bit, i.e. $m = 1$. Given arbitrary k-source X we have

$$|(Ext(X, U_d), U_d) - (U_1, U_d)| \leq \epsilon. \tag{4}$$

Namely,

$$\mathbb{P}_{r \leftarrow U_d}[|(Ext(X, r), r) - (U_1, r)| > \sqrt{\epsilon}] \leq \sqrt{\epsilon}. \tag{5}$$

That is, there are $(1 - \sqrt{\epsilon})$ fraction of seeds make the output of Ext become $\sqrt{\epsilon}$-close to uniform. Thus, a naive idea is to enumerate all the possible seeds in $\{0,1\}^d$ as follows. For the convenience of notation, we use $1, 2, \cdots, D$ to denote the enumeration in $\{0,1\}^d$.

$$Z = (Ext(X, 1), Ext(X, 2), \cdots, Ext(X, D)). \tag{6}$$

Then using majority on Z to decide the one bit output. However, since we the bits in Z can be arbitrarily correlated, we cannot guarantee majority function to work. As a result, its nature to think about measuring the dependency among the bits of Z. Namely, introducing some independence among bits.

Definition 3 (pairwise independent). *For any $t > 0$, $\gamma > 0$, and a distribution $\mathcal{D}$ over n bits. $\mathcal{D}$ is t-wise independent if the restriction of any t bits on $\mathcal{D}$ is uniform. $\mathcal{D}$ is (t, γ)-wise independent if the restriction of any t bits of $\mathcal{D}$ is γ-close to uniform.*

From [Vio14], when there exists $D - D^{0.49}$ bits in Z are constant-wise independent, majority function can extract an almost uniform bit. As a result, we turn to find a way to make Z becoming pairwise independent.

2.2 Switch to non-malleable extractors

Definition 4 (non-malleable extractors). *For any $t \in [n]$, $nmExt : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (n, k, t)-non-malleable extractor with error $\epsilon > 0$ if for arbitrary functions $f_1, \cdots, f_t : \{0,1\}^d \to \{0,1\}^d$ with no fixing point and (n, k)-source X,*

$$\begin{aligned}&|(nmExt(X, U_d), U_d, nmExt(X, f_1(U_d)), \cdots, nmExt(X, f_t(U_d))) \\ &- (U_m, U_d, nmExt(X, f_1(U_d)), \cdots, nmExt(X, f_t(U_d)))| \leq \epsilon\end{aligned} \tag{7}$$

Theorem 5 ([CGL15]). *Exists poly-time constructible (t, k, ϵ)-non-malleable extractor nmExt with $k = O(t \log^2(n/\epsilon))$ and $d = O(t^2 \log^2(n/\epsilon))$.*

Now, take
$$Z = (nmExt(X, 1), nmExt(X, 2), \cdots, nmExt(X, D)). \tag{8}$$
Note that $D = 2^{O(t^2 \log^2(n/\epsilon))}$ which is super-polynomial right now. Also, with some similar argument as we did for strong seeded extractor, one can show that there exists a subset of bits in Z with size at least $(1 - O(\sqrt{\epsilon}))D$ such that these bits are $(t, O(t\sqrt{\epsilon}))$-wise independent uniform distribution. This is proved in Section 3 of [CZ15]. Formally, we can model Z as an abstract mathematical object as follows.

Definition 6 (restriction). *Let $x \in \{0,1\}^n$ and $S \subseteq [n]$, the restriction of x on S is the bits in x that corresponds to S. Denote it as x_S.*

Definition 7 (NOBF source). *A source Z on $\{0,1\}^D$ is a (q, t, γ) NOBF source if $\exists Q \subseteq [D]$ where $|Q| \leq q$ such that $Z_{\bar{Q}}$ is (t, γ)-wise independent.*

From the definition of NOBF source above, Z is a $(O(\sqrt{\epsilon})D, t, O(t\sqrt{\epsilon}))$ NOBF source.

2.3 Sample subset of Z

Since the length of Z is D, which is super-polynomial, we can not expect to compute every bit of Z. As a result, we would like to sample polynomially many bits from Z while preserving the nice property of NOBF source. The idea is to use the other weak source Y to sample subset of bits from Z. By [Zuc97], we can get the following parameters.

Theorem 8 (reduce the length of Z). *There exist constants $\delta, c' > 0$ such that for every $n, t > 0$ there exists a poly-time constructible function reduce : $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{D'}$, $D' = n^{O(1)}$ such that for any independent (n, k)-sources with $k \geq c' t^4 \log^2 n$,*

$$\mathbb{P}_{y \leftarrow Y}[reduce(X, y) \text{ is a } (q, t, \gamma)\text{-NOBF source}] \geq 1 - n^{-\omega(1)}, \tag{9}$$

where $q = D^{1-\delta}$ and $\gamma = 1/D^{t+1}$.

2.4 Extract from NOBF source

Definition 9 (influence). *Let $f : \{0,1\}^D \to \{0,1\}$ be any boolean function, $\mathcal{D}$ be a distribution over $\{0,1\}^D$, and $Q \subseteq [D]$. Define $I_{Q,\mathcal{D}}(f)$, the influence of f on Q, as the probability of f being undetermined after fixing the variables outside Q to be sampled from distributions $\mathcal{D}$.*

- *Let $\mathcal{D}_{t,\gamma}$ be the set of all (t,γ)-wise independent distributions.*
- $I_{q,t,\gamma}(f) := \max_{\mathcal{D}\in\mathcal{D}_{t,\gamma},Q\subseteq[D],|Q|=q} I_{Q,\mathcal{D}}(f)$

Definition 10 (resilient function). *Let $f : \{0,1\}^n \to \{0,1\}$, $q \in [n]$, $t > 0$, $\gamma > 0$, and $\epsilon > 0$. We said f is (t,γ)-independent (q,ϵ)-resilient if $I_{q,t,\gamma}(f) \leq \epsilon$.*

Theorem 11 (using resilient function to extract NOBF source). *Let size m depth d circuit $f : \{0,1\}^n \to \{0,1\}$ be a function such that*

- $|\mathbb{E}_{x\sim U_n}[f(x)] - \frac{1}{2}| \leq \epsilon_1$.
- $I_q(f) \leq \epsilon_2$.

If $t \geq \log^{O(d)}(O(m)/\epsilon_3)$, then for any (t,γ)-wise independent distribution $\mathcal{D}$,

- $I_{q,t,\gamma}(f) \leq \epsilon_2 + \epsilon_3 + \gamma n^t$.
- $|\mathbb{E}_{x\sim\mathcal{D}}[f(x)] - \frac{1}{2}| \leq \epsilon_1 + \epsilon_3 + \gamma n^t$.

The proof idea behind this theorem is using results from [Bra10, Tal14] that AC^0 circuit has small (t,γ)-independent resilient and result from [AGM03] that for any (t,γ)-wise independent source is γn^t-close to a t-wise independent source.

Theorem 12 (existence of good resilient function). *For any $\delta > 0$ and large enough n, there exists a poly-time computable monotone Boolean function $f : \{0,1\}^n \to \{0,1\}$ satisfying:*

- *(small circuit) f is a depth 4 circuit in AC^0 of size $n^{O(1)}$.*
- *(small bias) $|\mathbb{E}_{x\sim U_n}[f(x)] - \frac{1}{2}| \leq \frac{1}{n^{\Omega(1)}} (= \epsilon_1)$.*
- *(small influence to size $n^{1-\delta}$) For any $q, t > 0$, $I_{q,t}(f) \leq \frac{q}{n^{1-\delta}}$ $(= \epsilon_2)$.*

The proof idea is to derandomize the Ajtai-Linial function, which is a randomized construction resilient function against coalition of size $O(n/\log^2 n)$.

- The construction of Ajtai-Linial function was a conjunction of randomly chosen tribe functions, which partition $[n]$ well (similar to design).
  - (For bounding influence) Any small subset (of size $n^{1-\delta}$) of $[n]$ has small intersection with most partitions.
  - (For bounding bias) The partitions are pairwise pseudorandom: the intersection of any two blocks is bounded.
- Idea: Using strong seeded extractor to build the pseudorandom collection of partitions.

2.5 Wrapping up

From the above journey, we summarize our construction into three steps:

- Step 1: Using non-malleable extractors to reduce X to NOBF source Z.
  - Section 3 of [CZ15].
  - Z is a $(\sqrt{\epsilon}D, t, 5t\sqrt{\epsilon})$-NOBF source for any $t > 0$ and $D = 2^{O(t^2 \log^2(n/\epsilon))}$.
- Step 2: Using Y to sample poly-length NOBF source Z'.
  - Section 3 of [CZ15].
  - Z' is a $((D')^{1-\delta}, t, 1/(D')^{t+1})$-NOBF source for any $t > 0$ and $D' = n^{O(1)}$ with probability $1 - n^{-\omega(1)}$.
- Step 3: Explicitly constructing good resilient function to extract from Z'.
  - Section 4 and 5 of [CZ15].
  - For any $\delta > 0$, exists function $f : \{0,1\}^{D'} \to \{0,1\} \in AC^0$ that on input an (q, t, γ)-NOBF source, f has at most $\frac{1}{(D')^{\Omega(1)}}$ bias and is (t, γ)-wise $(q, \frac{q}{(D')^{1-\delta/2}})$ resilient.

Finally, we can prove the following main theorem with the results from the above three steps.

Theorem 13. *main theorem $\exists C > 0$ such that for all $n \in \mathbb{N}$, there exists a poly-time computable construction of 2-source extractor $2Ext : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with error $n^{-\Omega(1)}$ for weak sources with min-entropy at least $\log^C n$.*

Lemma 14 (extracting NOBF source with resilient function)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be (t, γ)-wise (q, ϵ_1)-resilient and on any (q, t)-wise independent uniform distribution, the bias is at most ϵ_2. Then, f is an extractor for (q, t, γ)-NOBF source with error $\epsilon_1 + \epsilon_2$.*

proof of Lemma 14. Let X be a (q, t, γ)-NOBF source. Exists $Q \subseteq [n]$ where $|Q| \leq q$ and $X_{\bar{Q}}$ enjoys a (q, t)-wise independent uniform distribution, say $\mathcal{D}_1$. Next, denote E as the event that function f is determined on fixing bits in $\bar{Q}$ by sampling from distribution $\mathcal{D}_1$. Since f is (t, γ)-wise (q, ϵ_1)-resilient,

$$\mathbb{P}_X[E] \geq 1 - \epsilon_1. \tag{10}$$

In addition, let $\mathcal{D}$ be an (q, t)-wise independent uniform distribution on $[n]$ that matches $\mathcal{D}_1$ on $\bar{Q}$, as f has bias at most ϵ_2 on inputing any (q, t)-wise independent uniform distribution, we have

$$|\mathbb{P}_{x\sim\mathcal{D}}[f(x) = 1] - \frac{1}{2}| \leq \epsilon_2. \tag{11}$$

Next, decompose $\mathbb{P}_{x\sim\mathcal{D}}$ by conditioning on E and $\bar{E}$.

$$\mathbb{P}_{x\sim D}[f(x) = 1] = \mathbb{P}[E] \cdot \mathbb{P}_{x\sim D}[f(x) = 1|E] + \mathbb{P}[\bar{E}] \cdot \mathbb{P}_{x\sim D}[f(x) = 1|\bar{E}]. \tag{12}$$

When conditioning on E, the behavior of x is only affected by the portion in $\bar{Q}$, i.e., the distribution is the same as X.

$$\mathbb{P}_{x\sim D}[f(x) = 1|E] = \mathbb{P}_{x\sim X}[f(x) = 1|E]. \tag{13}$$

The second term can be upper bounded by ϵ_1. Thus, we have

$$|\mathbb{P}_{x\sim D}[f(x)=1]-\mathbb{P}_{x\sim X}[f(x)=1]|\leq\epsilon_1. \tag{14}$$

Combine (11) and (14), we have

$$|\mathbb{P}_{x\sim X}[f(x)=1]-\frac{1}{2}|\leq\epsilon_1+\epsilon_2. \tag{15}$$

□

proof of Theorem 13. From step 1 and 2, we can construct function $reduce:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}^{D'}$ to reduce from two weak sources to a poly-length NOBF source; from step 3, we can construct good resilient function $bitExt:\{0,1\}^{D'}\to\{0,1\}$.

Now, define our two source extractor $2Ext:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$ as

$$2Ext(x,y)=bitExt(reduce(x,y)). \tag{16}$$

To prove the theorem, consider arbitrary (n,k)-sources X,Y where $k\geq C_1(\log n)^C$ for some constant C_1,C. Let $Z'=reduce(X,Y)$. From step 1 and 2, with $1-n^{-\omega(1)}$ probability over $y\leftarrow Y$ that $Z|Y=y$ is a $((D')^{1-\delta},t,1/(D')^{t+1})$-NOBF source. Denote this even as F. Plug-in $q=(D')^{1-\delta}$, f is then (t,γ)-wise $((D')^{1-\delta},(D')^{-\delta/2})$ resilient and has bias at most $n^{-\Omega(1)}$ on input a $((D')^{1-\delta},t,1/(D')^{t+1})$-NOBF source.

By, Lemma 14, f extracts $Z|F$ with error at most $n^{-\Omega(1)}+(D')^{-\delta/2}$. Combine with the probability of $\bar{F}$, we have

$$|(2Ext(X,Y),Y)-(U_1,Y)|\leq n^{-\omega(1)}+n^{-\Omega(1)}+(D')^{-\delta/2}. \tag{17}$$

□

References

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88(3):107–110, 2003.

[BIW06] Boaz Barak, Russell Impagliazzo, and Avi Wigderson. Extracting randomness using few independent sources. *SIAM Journal on Computing*, 36(4):1095–1118, 2006.

[Bou05] Jean Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(01):1–32, 2005.

[Bra10] Mark Braverman. Polylogarithmic independence fools ac 0 circuits. *Journal of the ACM (JACM)*, 57(5):28, 2010.

[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.

[CGL15] Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. *arXiv preprint arXiv:1505.00107*, 2015.

[CZ15] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 22, page 119, 2015.

[Li12] Xin Li. Design extractors, non-malleable condensers and privacy amplification. In *Proceedings of the forty-fourth annual ACM Symposium on Theory of Computing*, pages 837–854. ACM, 2012.

[Li15a] Xin Li. Improved constructions of two-source extractors. *arXiv preprint arXiv:1508.01115*, 2015.

[Li15b] Xin Li. Three-source extractors for polylogarithmic min-entropy. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 863–882. IEEE, 2015.

[Rao09] Anup Rao. Extractors for a constant number of polynomially small min-entropy independent sources. *SIAM Journal on Computing*, 39(1):168–194, 2009.

[Raz05] Ran Raz. Extractors with weak random seeds. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 11–20. ACM, 2005.

[Tal14] Avishay Tal. Tight bounds on the fourier spectrum of ac0. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 21, page 174, 2014.

[Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014.

[Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. *Random Structures and Algorithms*, 11(4):345–367, 1997.