Extractors for polynomial sources over $\mathbb{F}_2$

Eshan Chattopadhyay Jesse Goodman Mohit Gurumukhani

Extractors

Definition (**Extractor**). *A function* $\text{Ext} : \{0,1\}^n \to \{0,1\}^m$ *is called an* ε*-extractor for a class* $\mathcal{C}$ *of distributions over* $\{0,1\}^n$ *if for all* $\mathbf{X} \in \mathcal{C}$,

$$|\text{Ext}(\mathbf{X}) - \mathbf{U}_m| \leq \varepsilon,$$

where $|\cdot|$ denotes statistical distance and $\mathbf{U}_m$ is the uniform distribution over m bits.

For extractors to exist for a class $\mathcal{C}$ of sources, we typically require each source in $\mathcal{C}$ to have some min entropy. For a source $\mathbf{X}$ with support Ω, we define it's min-entropy $H_\infty(\mathbf{X}) = -\log(\max_{x\in\Omega} \Pr(\mathbf{X} = x))$.

Algebraic sources

Definition (**Polynomial sources**). *A polynomial source* $\mathbf{X} \sim \mathbb{F}_2^n$ *is associated with a degree* d *polynomial map* $P = (p_1, \ldots, p_n)$ *where for* $1 \leq i \leq n, p_i : \mathbb{F}_2^m \to \mathbb{F}_2$. *Then,* $\mathbf{X} = P(U_m)$ *where* U_m *is the uniform distribution over* $\mathbb{F}_2^m$.

Definition (**Variety sources**). *A variety source* $\mathbf{X} \sim \mathbb{F}_2^n$ *has associated polynomials* $p_1, \ldots, p_m : \mathbb{F}_2^n \to \mathbb{F}_2$. $\mathbf{X}$ *is uniform over the variety generated by these degree* d *polynomials, i.e., it is uniform over the set* $V = \{x \in \mathbb{F}_2^n : \forall i \in [m] : p_i(x) = 0\}$.

Definition (**Polynomial NOBF sources**). *A polynomial NOBF source* $\mathbf{X} \sim \mathbb{F}_2^n$ *with* $H_\infty(\mathbf{X}) = k$ *is associated with* $n-k$ *degree* d *polynomials* $p_1, \ldots, p_{n-k}$ *where for each* $1 \leq i \leq n-k, p_i : \mathbb{F}_2^k \to \mathbb{F}_2$. *There are some* k *output bits in* $\mathbf{X}$ *that are independent and uniform, call them* $x_1, \ldots, x_k$. *The remaining* $n-k$ *bits output* $p_i(x_1, \ldots, x_k)$ *for* $1 \leq i \leq n-k$.

Observation. *A degree* d *polynomial NOBF source is both a degree* d *polynomial source as well as a degree* d *variety source.*

Related work

- Explicit extractors are known for polynomial sources over fields whose size is > 2.
- Explicit extractors with optimal dependence on min entropy are known for degree 1 polynomial sources over $\mathbb{F}_2$ (affine sources).
- Explicit extractors known for degree ≥ 2 variety sources over $\mathbb{F}_2$.
- **No explicit extractors** previously known for degree ≥ 2 polynomial sources over $\mathbb{F}_2$ even with min entropy $n - 1$!

Result 1: Explicit extractor

Theorem. *Let* $\varepsilon > 0$ *be a constant. For all* $d \in \mathbb{N}$, *there exists an explicit* ε*-extractor for degree* d *polynomial sources over* $\mathbb{F}_2^n$ *with min-entropy* $k \geq n - \frac{\sqrt{\log n}}{(d \log\log n)^{d/2}}$.

Idea 1: Input reduction

Remark. *In the definition of polynomial sources, the number of inputs to the source,* m, *is unbounded. Hence, its unclear whether even existentially, extractors for polynomial sources exist.*

We get around this with our key technical lemma that also proved to be useful for the explicit construction:

Lemma. *Let* $\text{Ext} : \mathbb{F}_2^n \to \mathbb{F}_2^t$ *be an extractor for the class of degree* d *polynomial sources with min-entropy* k *and* $O(k)$ *inputs. Then,* Ext *is also an extractor for the class of degree* d *polynomial sources with min-entropy* $\Omega(k)$ *and arbitrary inputs.*

Idea 2: Brute force extraction

Equipped with input reduction lemma and the fact that a random function is an optimal extractor for this class, we can already construct a non-trivial extractor:

Claim. *There exists an explicit extractor* $\text{Ext} : \mathbb{F}_2^n \to \mathbb{F}_2$ *for all constant degree* $d \in \mathbb{N}$ *polynomial sources with min-entropy* $k \geq n - O(\log\log n)$.

Proof sketch. Let $\mathbf{X}$ be d polynomial source with n outputs and $H_\infty(\mathbf{X}) \geq n - g$ where $g = O(\log\log n)$. Let $t = 2g$ be a length t prefix of $\mathbf{X}$ and let this source be $\mathbf{X}_{pre}$. Then, $H_\infty(\mathbf{X}_{pre}) \geq t - g = t/2$. We apply input reduction lemma over $\mathbf{X}_{pre}$ to infer it suffices to construct an extractor for min-entropy $t/2$ degree d polynomial sources with $O(t)$ inputs and t outputs. As random function over t bits will be an extractor for such sources, we exhaustively try all the 2^{2^t} functions from t bits to 1 bits as our candidate extractor. To verify, we brute force over all $2^{O(t)^d \cdot t}$ degree d polynomial sources with $O(t)$ inputs and t outputs and for each of them, check if it has enough min-entropy. If it does, check if our candidate extractor works. We will eventually find a candidate extractor that will work for all such sources, and we output that function as our extractor. □

Idea 3: Improved brute force extraction

We utilize the previously well known lemma that for a large enough r, a function from r-wise independent distribution will be a good extractor. So, in our brute force search, we only consider our candidate extractors to be functions from a r-wise independent hash function family. Doing this results in improved dependence on min entropy.

Result 2: Negative result against sumset extractors

A sumset extractor is a function that extracts from sumset sources, defined as:

Definition (**Sumset sources**). *A* (k,k) *sumset source* $\mathbf{X}$ *is such that* $\mathbf{X} = \mathbf{A} + \mathbf{B}$, *where* $\mathbf{A}, \mathbf{B}$ *are independent distributions on* $\{0,1\}^n$ *with* $H_\infty(\mathbf{A}) \geq k, H_\infty(\mathbf{B}) \geq k$.

Sumset extractors are the most powerful general purpose extractors available: using reductions, they can extract from many other well studied models of weak sources such as affine sources, class of two independent sources, sources generated by branching programs, sources generated by AC^0 circuits and many more. Recently, explicit sumset extractors with optimal dependence on min entropy were constructed. We show a strong negative result against them:

Theorem. *Sumset extractors cannot even disperse from degree* 2 *polynomial NOBF sources with min-entropy* $n - O\left(\frac{n}{\log\log n}\right)$.

As polynomial NOBF sources are a subclass of polynomial sources and variety sources, this lower bound applies to both of them! This result also shows that there are classes where simple constructions of extractors beat optimal sumset extractors: the generalized inner product function is an extractor for variety sources with a much lower min entropy requirement than optimal sumset extractors.

Open problems

1. Construct extractors for polynomial / polynomial NOBF sources with better min-entropy dependence than what we constructed here.
2. Construct extractors for polynomial sources with degree $\text{poly}(\log n)$. Such an extractor will also extract from sources sampled by $\text{AC}^0[\oplus]$ circuits, a model for which no non-trivial extractors are known.