6.S897 Algebra and Computation March 21, 2012

Dispersers and Extractors for Low-Degree Algebraic Sources

Instructor: Madhu Sudan (Lecturer: Eli Ben-Sasson) *Scribe: Elette Boyle*

Much of today is based on a paper of Ben-Sasson and Gabizon [BG11] on ECCC.

1 Motivation

Randomness has proven to be a useful tool in design of algorithms. For some tasks, such as polynomial identity testing, efficient randomized algorithms are known but there is no deterministic equivalent. (Another example of this used to be primality testing for integers, but as we saw in the last lecture an efficient deterministic algorithm now exists). This raises several questions: Is randomness needed to achieve certain tasks? Can we remove it? Exactly what does randomness do in computation?

These questions are the center of several lines of important research in theoretical computer science. This includes some of the following goals:

- Removing randomness (*derandomization*). That is, for some randomized algorithm, can you come up with an algorithm that doesn't require randomness? For some simple classes, like AC^0, this can always be done. In general, it is not always known.
- Distilling randomness (*extractors*). Perhaps we have a source of partial randomness, with only a guarantee that there is high entropy. However, many applications require *truly* random bits. An extractor is a fully deterministic construction that takes partly random bits and compresses them in a clever way to extract out pure randomness.

  An example application of extractors arises in the setting of cryptography, where you may have only an entropy source, or randomness on which partial information has been leaked, and you need *truly* random secret keys to ensure security of your cryptosystem/protocol.
- Constructing *pseudo-random objects*. For example, expander graphs are an "approximation" of random graphs. Because of this property, they are useful for many applications, such as the construction of error-correcting codes.

2 Extracting from Low-Degree Polynomials Sources

2.1 Low-Degree Polynomial Sources

Definition 2.1. A *source* is a distribution over some finite domain. In our case, we will consider distributions over $\mathbb{F}_q^n$.

For today, we will look at a particular class of "simple" sources that are generated by polynomials of low degree, and will show how to extract (true) randomness from such sources.

Given a collection of polynomials $f_1, ..., f_m$ (in our case, we will consider polynomials of low degree, say $\deg f_i \leq 2$), there are two natural associated sources:

1. The first is the uniform distribution over the set of common zeroes of the f_i. Namely,

$$X \text{ uniform on } \{x \in \mathbb{F}_q^r : f_1(x) = \cdots = f_m(x) = 0\}.$$

   This set has many nice mathematical properties: in particular, it forms an *algebraic variety* (in our case, a degree-2 variety). One thing to note is that we do not have a method of efficiently sampling from this distribution, for general choices of polynomials f_i. However, even so, it would still be very interesting (and probably useful in the long run) to be able to extract from it.

2. Today, we will be focusing on parameterized sources corresponding to the f_i. That is,

$$X = \{f_1(Z), ..., f_n(Z) : Z \text{ uniform on } \mathbb{F}_q^r\}. \tag{1}$$

   The output of this source is computed by evaluating a number of polynomials on a uniformly random input.

We remark that for $\deg f_i = 1\ \forall f_i$, these two sources are actually equivalent. This is because affine sources can be equivalently described as the image of a linear transformation or the kernel of one. This type of source is referred to as an *affine source.*

2.2 Extractors for Low-Degree Poly Sources

The most important property of a source (for us) is its *min-entropy.* Loosely speaking, the min-entropy of X corresponds to the size of $Supp(X)$.

Definition 2.2. X has *min-entropy* k if $\forall x \in \mathbb{F}_q^n$, $\Pr[X = x] \leq q^{-k}$.

Note that extracting from a source becomes harder and harder as the min-entropy of the source decreases. Often, people look at the *min-entropy rate*, where the min-entropy rate of a source X over $\mathbb{F}_q^n$ is defined to be the ratio (min-entropy(X))/n.

We will now focus on the low-degree polynomial sources that we introduced in the previous section.

Definition 2.3. We define an $[n, k, d]_q$*-source* as a distribution X given by degree d polynomials (as in Equation (1)) with the property that $|Supp(X)| \geq q^k$.

Definition 2.4. A function $E : \mathbb{F}_q^n \to \mathbb{F}_q^m$ is a (k, d, ϵ)*-extractor* if for all $[n, k, d]$ sources X,

$$|E(X) - U_{\mathbb{F}_q^m}| \leq \epsilon;$$

that is, the output distribution of E applied to any $[n, k, d]$-source X is within ϵ statistical distance from the uniform distribution over $\mathbb{F}_q^m$.

An important object that is used to yield extractors is known as a *disperser.* Informally, a disperser for a particular class of sources is a function D with the property that for any source X within the relevant class, the support of the output distribution $D(X)$ is sufficiently large. In fact, for intuition, you can think of a disperser as being any function that is *non-constant* on any such source X.

For simplicity, we will state the next theorem in terms of constructing a disperser.

Theorem 2.5. *There exists a deterministic function $E : \mathbb{F}_4^n \to \mathbb{F}_4$ that is non-constant on any $[n, \frac{n}{2} + 100, 2]$-source, provided the source is multi-linear homogeneous. (That is, if the source is of the type defined in Equation (1), where the f_i's are each homogeneous of degree 2 and linear in each variable).*

More explicitly, the theorem considers sources of the form $X = \{f_1(Z), ..., f_n(Z) : Z \text{ uniform on } \mathbb{F}_q^r\}$ where for each $i = 1, ..., n$,

$$f_i(z_1, ..., z_r) = \sum_{(j,\ell) \in \binom{r}{2}} a_{j,\ell}^{(i)} z_j z_\ell,$$

with coefficients $a_{j,\ell}^{(i)} \in \mathbb{F}_4$.

Consider a simpler question. Namely, suppose we want one function that takes n bits to 1 bit with the property that for any affine (deg 1) space X of dimension $> n/2$, this function is not constant. Consider the *inner product* function: i.e.,

$$E(x_1, ..., x_n) = x_1 x_2 + x_3 x_4 + \cdots + x_{n-1} x_n.$$

Why does the inner product function satisfy this property? In order for the inner product to be constant on the source X, the space spanned by the $(x_1, x_3, ..., x_{n-1})$ must be dual to the space spanned by the $(x_2, x_4, ..., x_n)$. The maximum such dimension of these spaces is $n/2$, but our space has strictly more than $n/2$ entropy.

In fact, this construction is even an *extractor*. Using Fourier analysis, one can show that $|IP(x^{(1)}, x^{(2)}) - U| < 2^{-(k\frac{n}{2})/2}$ (where IP is the inner product). You can also use this technique to extract more than one random bit. To do so, modify the standard inner product by multiplying by first transforming the two vectors via an invertible matrix M_α:

$$E_\alpha(x^{(1)}, x^{(2)}) = \begin{bmatrix} \qquad x^{(2)} \qquad \end{bmatrix} \begin{bmatrix} \\ \qquad M_\alpha \qquad \\ \\ \end{bmatrix} \begin{bmatrix} \\ x^{(1)} \\ \\ \end{bmatrix} = b.$$

Note that for $M_\alpha = I_{n/2}$ (the identity matrix), E_α corresponds to the standard inner product function. Now, pick a basis $\{\alpha_1, ..., \alpha_{n/2}\}$ for the field $\mathbb{F}_{2^{n/2}}$ over $\mathbb{F}_2$. For each element α in the basis, let M_α be the matrix corresponding to the $\mathbb{F}_2$-linear transformation of $\mathbb{F}_{2^{n/2}}$ to itself given by of multiplication by α. Note that M_α is necessarily full rank, since multiplication by α is invertible within the field. Since the α_i are independent over $\mathbb{F}_2$, we can concatenate each of the resulting output bits:

$$E(x^{(1)}, x^{(2)}) = \Big(E_{\alpha_1}(x^{(1)}, x^{(2)}), ..., E_{\alpha_{n/2}}(x^{(1)}, x^{(2)}) \Big).$$

3 Applications

Theorem 3.1 (Demenkov-Kulikov 2011). *Any circuit over the full binary basis (ie, allowing any function from 2 bits to 1 bit as single gate) computing an affine disperser for sources of min-entropy rate $o(1)$ over $\mathbb{F}_q^n$ requires $\geq 3(1 - o(1))n$ gates.*

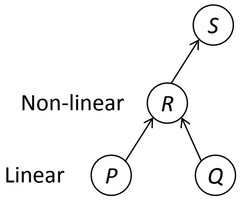
Remark 3.2. A few remarks about this result.

1. World record lower bound! This is actually the highest lower bound on circuit size that has been proved for any type of computation.

2. The lower bound is for dispersers for affine sources with $o(1)$ (low) min-entropy rate. Constructions of such dispersers are actually known [Ben-Sasson, Kopparty 2009] and [Shaltiel 2011].

Proof of Theorem 3.1. The high-level idea will be to iteratively "kill off" 3 gates of the circuit at the cost of restricting the input source by codimension 1. (By killing off a gate, we mean that this gate now evaluates to a constant for all values in the restricted input space). Repeating this process, we will eventually restrict to some subspace of the original input distribution for which the evaluation circuit of the disperser can be described by a trivial constant circuit. However, if the restricted subspace still has sufficient min-entropy, then this will contradict the fact that the circuit evaluates a *disperser*, which must be non-constant on such sources.

We now describe this process more formally. Begin with the original circuit evaluating the affine disperser, and sort the gates of the circuit topologically. Consider the first gate R within this ordering that is non-linear on its two inputs. Since this is the first such gate, the two gates P, Q preceding R are necessarily linear (affine) functions.



We will restrict our input space X to either X_0 or X_1, where X_b is defined as

$$X_b := \{x \in X : \text{gate } P \text{ evaluates to } b \text{ on circuit input } x\}.$$

Note that, since P is linear, if X is an affine source, then so will each X_b. Now, since gate R is non-linear, then $R(z_1, \cdot)$ must be constant for either $z_1 = 0$ or $z_1 = 1$. We will restrict to the space X_b for this choice of $z_1 = b$.

Consider the effects on the circuit of restricting to X_b. By choice of the space X_b, we know that gates P and R always evaluate to a constant, and so can be removed. Indeed, the constant either corresponds to the identity gate for one of the incoming wires, which can be removed altogether, or a not gate, which can be absorbed into each of the following gates (this is where we use that our circuit is over the full binary basis). Further, there must be at least one gate S following gate R. Since R evaluates to a constant, S is either the identity or negation of its second input, which can also be absorbed into the following gates. Hence, by restricting our input source by codimension 1 (from X to X_b), we have removed 3 gates from the circuit needed to compute the disperser on the restricted input space. Hence, by repeating this process as described above, the theorem holds.

□

Theorem 3.3 (Hou, Leung, Xiang 2002)**.** *Let n be prime. Then for any two sets $A, B \subseteq \mathbb{F}_{q^n}$ with $A, B \neq \{0\}$, the product set $A \cdot B = \{a \cdot b : a \in A, b \in B\}$ satisfies*

$$\dim(span(A \cdot B)) \geq \min\{n, \dim(span(A)) + \dim(span(B)) - 1\}.$$

Here, dimensions are taken over $\mathbb{F}_q$. (Recall that $\mathbb{F}_{q^n} \cong \mathbb{F}_q^n$ as $\mathbb{F}_q$-vector spaces).

This result says that if you know something about the *additive* structure of A and B (ie, $span(A), span(B)$), then you can conclude something also about the multiplicative structure.

Proof. Assume $A = span(A)$ and $B = span(B)$. (The proof easily extends to general A, B). We will prove by induction on $\dim(A)$.

If $\dim(A) = 1$, then the claim holds directly, since $A \cdot B$ is isomorphic to B as $\mathbb{F}_q$-vector spaces, for example by the map $\phi_a : B \to A \cdot B$ defined by multiplication by any fixed element $a \in A \setminus \{0\}$. Thus, $\dim(span(A \cdot B)) = \dim(span(B))$, and we are done.

Now, suppose $1 < \dim(A) < n$. Without loss of generality, we can assume $1 \in A \cap B$. Otherwise, we can move to any $A \to g \cdot A$, $B \to g' \cdot B$ for $g, g' \in \mathbb{F}_{q^n} \setminus \{0\}$ without changing dimension, since this is an invertible linear transformation. Pick $a \in A \setminus \mathbb{F}_q$. Such an a exists since $\dim(A) > 1$. Take ℓ to be the smallest power of a that is not contained in B. We know that $0 < \ell \leq n-1$: $\ell > 0$ because $1 \in B$, and we can assume $\ell < n$ since otherwise B is necessarily the entire space $\mathbb{F}_{q^n}$ (in which case the claim holds immediately). Now, multiply B by $a^{-(\ell-1)}$, (which exists, since we are within a field). The result is $1 \in (A \cap B)$ and $a \in A \setminus B$. Thus, $1 < \dim(A \cap B) < \dim(A)$. So, by induction, it holds that

$$\dim((A \cap B) \cdot (A + B)) \geq \min\{n, \dim(A \cap B) + \dim(A + B) - 1\},$$

where $A+B$ is the span of $\{a+b : a \in A, b \in B\}$. But $\dim(A \cap B) + \dim(A+B) = \dim(A) + \dim(B)$. Notice that $span(A \cdot B) \supseteq (A \cap B) \cdot (A+B)$, and so $\dim(A \cdot B) \geq \dim(A \cap B) \cdot (A+B)$. Indeed, for general $a \in A,\ b \in B,\ c \in A \cap B$, $c(a+b) = ca + cb = b'a + a'b \in span(A \cdot B)$. □

This theorem is very powerful, as it tells us a way to expand the dimension of a source so it fills up the whole space. It is something to keep in the back of your mind when attempting to deranomdomize algorithms.

Ending question: Show that the following transformation expands the linear dimension of its input. Take $A \subseteq \mathbb{F}_2^n$ a space of dimension d. Prove that

$$\dim\left(span\{a^3 : a \in A\}\right) > d + 3.$$

What is strange is that, if $d > n/2$, then you can show that the space expands to the full dimension n. But if the original dimension is, say, $n/2 - 1$, then we don't even know if it expands a little bit. If you can show this, it would be an important step for constructing better pseudorandom objects.

References

[BG11] Eli Ben-Sasson and Ariel Gabizon. Extractors for Polynomial Sources over Constant-Size Fields of Small Characteristic. ECCC Report 2011/129, 2011.