Lecture xx : Lossless expanders

Topics in Pseudo-randomness and Complexity Theory (Spring 2018)
Rutgers University
Swastik Kopparty
Scribe: Philip, Vishvajeet N

In this lecture we look at explicit constructions of some expanders which are not constant degree expanders, but are interesting in some other sense.

We describe what are called "Lossless Expanders" as follows:

Consider a bipartite graph such that it has N vertices on the left and M vertices on the right. Left and right sort of have special meaning in this graph. If we take any set of vertices on the left, it expands well on the right.

Definition 1. *(k, ϵ) lossless expanders These are bipartite graphs on the vertex set $[N] \times [M]$ with the edge set satisfying the following property*

$\forall S \subseteq [N]$ such that $|S| = k$
$|\Gamma(S)| \geq (1 - \epsilon)D|S|$

Where $\Gamma(S)$ denotes the neighbourhood of the set S.

It is known that these exist whenever $k \leq \epsilon \frac{M}{D}$

These objects are also called as lossless condernsors - randomness condensors.

Let's recall what a randomness extractor is. It is an object which works like this : Give it a weakly random source (variable) and a source of pure randomness, it gives out one with lots of randomness.

Lossless expanders are functions $C : [N] \times [D] \to [M]$ such that when fed with a random variable X and a purely random source, churn out more random bits, while recovering the original bits fed in. They are 'lossless' in the sense that we recover the bits which we fed in.

Formally,

$C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$

Such that $C(X, U_{\log D})$ is ϵ-close to some random variable Y with $H_\infty(Y) \geq \log k + \log D$.

These structures exist only if $k \leq \epsilon \frac{M}{D}$.

Alternatively, $\log k + \log D \leq \log M - \log(\frac{1}{\epsilon})$.

This can be proved by the probabilistic method. It should be noted that these structures are interesting only when ϵ is small, and also that we can never hope to get full min-entropy.

By now, it is easy to guess what the function C would be. $C(i, j)$ denotes the j-th neighbour of the i-th vertex on the left.

The probabilistic method proof goes this way (high level idea) :

1. It suffices to prove this for $X \subseteq [N]$ which are flat. i.e. uniform on a set of size k
2. $C(X, U_{logD}) = z$ i.e. we pick an uniform element of X, and then pick a neighbour uniformly at random.
3. Thus, z is ϵ-close to a uniform distribution over a set of size kD which has $H_\infty \geq \log k + \log D$.
4. We can adjust ϵ-fraction of the edges coming out of X, and make all kD neighbours distinct.

Let's look at examples of explicit constructions. The following example will have a non-constant degree though.

Example 2. *Projective plane graph*

Consider $[N] = \mathbb{F}_q^2$ and $[M] = \mathbb{F}_q^2$. Think of elements on the left as lines, and elements on the right as points. There is an edge between a line and a point if the point lies on the line. $\diamondsuit$

This graph has some nice properties, including that there is no $K_{2,2}$ in this graph.

The idea is that if you look at a few lines, the distribution is uniform except at the intersection points. We can pin the exact number of points where this is true because we know how many points two lines can interesect in - exactly one. Thus, if there are t lines, the distribution is uniform on at least tq - $\binom{t}{2}$ points. This has to be greater than $(1-\epsilon)tq$.

$tq - \binom{t}{2} \geq (1-\epsilon)tq$

$\implies t < O(\epsilon q)$

$k < O(\epsilon \frac{M}{D})$

We can generalize this. In this case, think of lines as degree two polynomials and points as roots. A natural way to generalize is to look at higher degree polynomials and their evaluations.

Thus, let the elements on the left be degree d polynomials i.e. points in $\mathbb{F}_q^d$ and those on the right be points in $\mathbb{F}_q^2$. We can also think of points on the right as $(x, p(x))$, joined with the curve $p(x)$ on the left. The idea is that if we look at any 2 degree d curves, they can agree in at most d evaluations.

Thus, if we take t polynomials of degree $< d$, the size of the image on the right is at least $tq - \binom{t}{2} * d$. Since we want it to be at least $(1-\epsilon) * tq$, it happens if

$t \leq \epsilon \frac{q}{d} = \epsilon \frac{M}{D}$

Suppose $T \subseteq [M]$ t $S = \{i \in [N] \text{ st. } \Gamma(i) \subseteq T\}$

We want to show that S is small. We use an algebraic argument.

Overview:

- Step 1 : Find a low-degree polynomial Q(x,y) that vanishes on T.
- Step 2 : Somehow conclude that every element of S is closely related to Q.

One important property we will use is the following : whenever there is a homogeneous system of equations with more variables than constraints, it always has a non-zero solution. To this end, we define the following set of equations:

$Q(x,y) = \sum_{(i,j)\in U} a_{ij}x^i y^j$

$U = \{(i,j) \text{ st. } i,j \geq 0, i+(d)j < q\}$

$|U| = q + (q-d) + (q-2d) + \cdots + (q - \frac{q}{d} * d) \approx \frac{q}{d} * q - \frac{1}{2}(\frac{q}{d})^2 * q \approx q^2$

If $|T| < |U| \approx \frac{q^2}{2d}$ then there is a non-zero solution to the system defined above i.e. one which vanishes on all $z \in T$.

This completes step 1. Step 2 follows.

Take any $p(x) \in S$ i.e. $\forall x \in \mathbb{F}_q$, $(x, p(x)) \in T$

Consider $h(x) = Q(x,p(x)) = \sum_{(i,j)\in U} a_{ij}x^i p(x)^j$. By choice of U, $deg(h) < q$

$\implies h(x) = 0 \implies y - p(x)$divides $Q(x,y)$

It should be noted that bivariate polynomials also have a unique factorization.

Thus, the number of such $p(x) \in S$ is at most the number of $y - p(x)$ that divide $Q(x,y)$, which is $\leq y - deg(Q)$ which is $\leq \frac{q}{d}$.

This is not the best quantitative analysis, but it suffices.

Let's go over the above construction and give an overview. Here's what we did :

- We want to find the size of the set on the right side such that its preimage is one guy on the left.
- We interpolate a low degree polynomial on the set
- We say that the polynomials on the left which are connected to this point divide this interpolating polynomial
- This interpolating polynomial can't have too many divisors, and that concludes the argument.

Now, our goal for the rest of the lecture is to find an *explicit* (k, ϵ) lossless condensor for any $K = 2^k$, $N = 2^n$, $D = 2^d$- which should ideally be poly log(n)

$M = (\frac{kD}{\epsilon})^{1.1}$

The output distribution of this condensor should be ϵ-close to the min-entropy 0.9m, where $[M] = \{0,1\}^m$

Plan: Find a set $\mathbb{P} \subseteq \mathbb{F}_q[x]^m$, where each $(P_1(x), P_2(x), \ldots, P_m(x)) \in \mathbb{P}$ has degree $p_i < d$. Now think of the expander as going from $\mathbb{F}_q[x]^m$ to $\mathbb{F}_q^{(m+1)}$, where the ith neighbor of $(P_1(x), P_2(x), \ldots, P_m(x))$ is $(i, P_1(i), P_2(i), \ldots, P_m(i))$. Think of the elements of $\mathbb{F}_q[x]^m$ as curves in $m+1$ dimensions, and their neighbors as the points on the curve in $\mathbb{F}_q$.

We plan to show that for all T with size $(1-\epsilon)DK$, the set of S with $S = \{(P_1(x), P_2(x), ..., P_m(x)) \subseteq \mathbb{P}$ s.t. $\Gamma(P_1, P_2, \ldots, P_m)\}$ has $|S| < K$.

A high-level overview of the plan, which is similar to what we did previously, is as follows:

1. Find $Q(x, y_1, y_2, \ldots, y_m)$ of low degree that vanishes on T
2. Note that, if $(P_1(x), P_2(x), \ldots, P_m(x)) \in S$, then $h(x) \equiv Q(x, P_1(x), P_2(x), \ldots, P_m(x)) = 0$
3. Show that, by the choice of $\mathbb{P}$ in our construction, not too many elements of $\mathbb{P}$ are "roots" of Q.

The property we want for $\mathbb{P}$ is that it is a set of curves such that no hypersurface contains too many of them.

Toy Problem:
Suppose we have some unknown $Q(z_1, z_2, \ldots, z_M) \neq 0 \in \mathbb{Z}[z_1, z_2, \ldots, z_m]$, with $deg(Q) \leq e$ in each variable.
We want to define a set $A \subseteq \mathbb{Z}$ such that Q does not vanish on too many points of A.

Suggested Solution 1: Choose $A = (\{1, 2, 3, \ldots, 2e\})^m$. Q will be non-zero for at least $\frac{1}{2^m}$ of the points of A.

Solution 2: Choose $A = \{(i, i^e, i^{e^2}, \ldots, i^{e^{(m-1)}}) : i \leq 2e^m\}$. Then Q is non-zero on at least $\frac{1}{2}$ of the points of A.
To see why that is, note that $R(T) \equiv Q(T, T^e, T^{e^2}, \ldots, T^{e^{(m-1)}})$ is a non-zero univariate polynomial of degree $< e^m$. If Q vanishes on $(i, i^e, i^{e^2}, \ldots, i^{e^{m-1}})$, then R vanishes on i. Since there are fewer than e^m real roots of R, there must be fewer than e^m points that vanish on Q.

We want these results for fields in order to apply them to our lossless condenser construction, but we need to factor out common divisors.

Some observations:
Consider $A* = \{(i, i^e, i^{e^2}, \ldots i^{e^{m+1}}) \mod p\} \subseteq \{0, 1, \ldots, p-1\}^m \subseteq \mathbb{Z}^m$
Analysis:
If all coefficients of Q are divisible by p, then divide out p. So then Q is non-zero $\mod p$.
Let $\bar{Q}(z_1, z_2, \ldots, z_m) \in \mathbb{F}_p[z_1, z_2, ..., z_m]$ be the reduction. Since Q is non-zero $\mod p$, $\bar{Q}$ is non-zero.
Consider $\bar{R}(T) = \bar{Q}(T, T^e.T^{e^2}, \ldots, T^{e^{m-1}})$. Again, since $\bar{Q}$ is non-zero, $\bar{R}(T)$ must also be non-zero.

As before, if $Q((i, i^e, i^{e^2}, \ldots, i^{e^{m-1}}) \mod p) = 0$, then $Q(i, i^e, i^{e^2}, \ldots, i^{e^{m-1}}) = 0 \mod p$, and so $\bar{Q}(i, i^e, i^{e^2}, \ldots, i^{e^{m-1}}) = 0$ in $\mathbb{F}_p$, and therefore $\bar{R}(i) = 0$. So then $Q((i, i^e, i^{e^2}, \ldots, i^{e^{m-1}}) \mod p) = 0$ can be true for at most e^m values of $i \mod p$.

Take $p > 2e^m$. Then we get that Q vanishes on $\leq \frac{1}{2}$ of the points of $A*$.

Final Construction:

Choose $E(x)$ to be an irreducible polynomial in $\mathbb{F}_q[x]$ of degree d.

Let $\mathbb{P} = \{(A(x), A(x)^e, A(x)^{e^2}, \ldots, A(x)^{e^{m-1}}) mod E(x) : deg(A(x)) < d\}$

Claim 3. *For any $Q(x, y_1, y_2, \ldots, y_m)$ with individual degrees $< e$, Q vanishes on at most e^m elements of $\mathbb{P}$*

First, divide Q by any power of $E(x)$ that divides all coefficients of $Q(x, y_1, y_2, \ldots, y_m)$, i.e. divides

$$\sum_{i_0, i_1, \ldots, i_m} a_{1,2,\ldots,m} x^{i_0} y_1^{i_1} y_2^{i_2} \ldots y_m^{i_m}.$$

Now we may assume that $Q(x, y_1, y_2, \ldots, y_m) \neq 0 \mod E(x)$. Let $\bar{Q}(y_1, y_2, \ldots, y_m) \in \mathbb{F}_{q^d}[y_1, y_2, \ldots, y_m]$, noting that $\mathbb{F}_q[x]/E(x) \cong \mathbb{F}_{q_d}$.

Let $\bar{R}(T) \in \mathbb{F}_{q^d}[T]$ be

$$\bar{R}(T) = \bar{Q}(T, T^e, T^{e^2}, \ldots, T^{e^{m-1}})$$

Then $\bar{R}(T)$ is non-zero and of degree $\leq e^m$.

If $Q(x, (A(x), A(x)^e, A(x)^{e^2}, \ldots, A(x)^{e^{m-1}}) \mod E(x)) = 0$, then $Q(x, A(x), A(x)^e, A(x)^{e^2}, \ldots, A(x)^{e^{m-1}}) = 0(\mod E(x))$ which means that $\bar{Q}(A(x), A(x)^e, A(x)^{e^2}, \ldots, A(x)^{e^{m-1}}) = 0$ in $\mathbb{F}_{q^d}$, and so

$$\bar{R}(A(x)) = 0.$$

But $\bar{R}(T)$ is non-zero, so fewer than $deg(\bar{R}) \leq e^m$ polynomials $A(x) \mod E(x)$ exist, and so it immediately follows that fewer than e^m such polynomials $A(x)$ exist of degree no greater than d.