U.C. Berkeley
CS294: Pseudorandomness and Combinatorial Constructions
Professor Luca Trevisan

Handout N21
November 10, 2005
Scribe: Gatis Midrijanis

Notes for Lecture 21

1 Condensers

In previous lectures we saw how to construct expanders. But to apply expanders for inputs of large size we need condensers that first reduce inputs size. In the last lecture we saw such construction which we will sketch now.

Let $S_1, .., S_m$ be sets such that $S_i \in \{1, ..., d\}$, $|S_i| = l$ and $|S_i \bigcap S_j| \leq a$ for each i, j. Let $ECC : \{0,1\}^n \rightarrow \{0,1\}^{\overline{n}}$ be an error correcting code with min-distance at least $\overline{n}/5$, where $\overline{n} = 2^l$. We defined condenser like $Cond(x, z, i)$ as $n' = m2^a$ bit string from $ECC(x)$, where x is n bit string, z is d bit string and $i \in \{1, ..., m\}$.

Now we will state the main result of the last lecture. Informally, it says that the condenser doesn't loose much information. We formalize this by giving deterministic procedure Dec_x that can reconstruct the input of the condenser by it's output.

Lemma 1 *Suppose X is a distribution such that $H(X) \leq \frac{\varepsilon m}{10}$. Then there is a decoding procedure Dec_x such that*

$$\mathop{\mathbf{Pr}}_{x \sim X, z \sim \{0,1\}^d, i \sim \{1,...,m\}} [Dec_x(z, i, Cond(x, z, i)) = x] \geq 1 - \varepsilon.$$

In this lecture we will finish the proof of the correctness of composition of condensers and an extractor. First we state two lemmas without a proof.

Lemma 2 *Let X is uniform distribution over a set of size 2^k, where $k \leq \frac{\varepsilon m}{10}$. Then $Cond(X, U_d, U_{[m]})$ is ε-close to a distribution Y of min-entropy at least k.*

Lemma 3 *If X has min-entropy at least k, where $k \leq \frac{\varepsilon m}{10}$. Then $Cond(X, U_d, U_{[m]})$ is ε-close to a distribution Y of min-entropy at least k.*

At the very end we are using condensers before applying extractors, because for extractors the ratio between min-entropy and the length of a message should be high. It is where condensers help us - they reduce the length of a message until we can start to use extractors. One can see it in Figure 1.

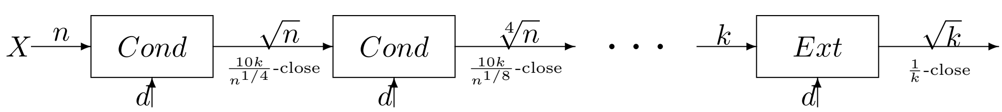

Figure 1: The very end construction by composing condensers and an expander.

Figure 1 shows the very end construction that is composition of sufficiently many condensers reducing the size of the input to k and an extractor at the end. We chose the parameters to be

$a = \log m$, $n' = m^2$ and $m = n^{1/4}$ thus having a condenser $Cond : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{\sqrt{n}}$ with length of the random input $t = d + \log m$. If input X has min-entropy at least k then output $Cond(X, U_t)$ is $\frac{10k}{n^{1/4}}$-close to a distribution of min-entropy at least k.
In the rest of the lecture we check that this composition works. Before that let precisely define finite version of condensers.

Definition 4 *$Cond : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^{n'}$ is a (k, ε)-condenser if for every X of min-entropy at least k $Cond(X, U_t)$ is ε-close to a distribution Y of min-entropy at least k.*

Lemma 5 *If $Cond_1 : \{0,1\}^n \times \{0,1\}^{t_1} \to \{0,1\}^{n_1}$ is a (k, ε_1)-condenser and $Cond_2 : \{0,1\}^{n_1} \times \{0,1\}^{t_2} \to \{0,1\}^{n_2}$ is a (k, ε_2)-condenser then $Cond(x, z_1, z_2) = Cond_2(Cond_1(x, z_1), z_2)$ is $(k, \varepsilon_1 + \varepsilon_2)$-condenser.*

PROOF: Figure 2 shows the relation between outputs of condenses. $Cond_1$ just outputs a distribution that is ε_1-close to Y_1 with min-entropy at least k. Take Y_1 as an input of condenser $Cond_2$. Then its output is Y_2 with min-entropy at least k and such that $||Y_1 - Y_2||_{SD} \leq \varepsilon_2$. By triangle inequality, a statistical distance sums up to at most $\varepsilon_1 + \varepsilon_2$.

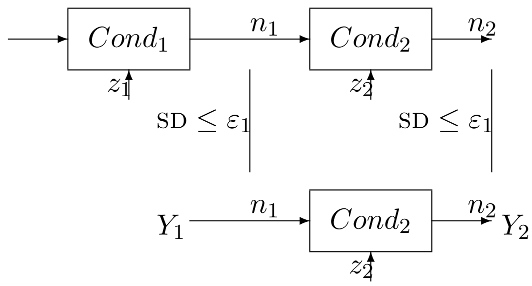

Figure 2: Composing condensers.

□

Lemma 6 *If $Cond : \{0,1\}^n \times \{0,1\}^{t_1} \to \{0,1\}^{n_1}$ is a (k, ε_1)-condenser and $Ext_1 : \{0,1\}^{n_1} \times \{0,1\}^{t_2} \to \{0,1\}^{n_2}$ is a (k, ε_2)-extractor then $Ext(x, z_1, z_2) = Ext_1(Cond(x, z_1), z_2)$ is a $(k, \varepsilon_1 + \varepsilon_2)$-extractor.*

PROOF: The same reason as in Lemma 5. □

Next lemma we showed in the previous lectures.

Lemma 7 *There is a universal constant c and $Cond : \{0,1\}^n \times \{0,1\}^{c \log n} \to \{0,1\}^{\sqrt{n}}$ that is a $(k, \frac{10k}{n^{1/4}})$-condenser. There is $Ext : \{0,1\}^n \times \{0,1\}^{O(\log n)} \to \{0,1\}^m$ that is a $(O(m^2), O(1/m))$-extractor for $m = n^{\Omega(1)}$.*

Combining lemmas above we see that the construction of Figure 1 works. Now let start to proof lemmas.

Proof:[Lemma 3] If X has min-entropy at least k then it is a convex combination $\sum_i p_i X_i$ of distributions X_i such that each X_i is a uniform over a set of size 2^k. To state it equivalently: there exist sets $S_1, ..., S_M$, such that $|S_i| = 2^k$ and distribution X is given by picking S_i with probability p_i and outputting a random element of S_i. p_i is a probability because $p_i \geq 0$ and $\sum_i p_i = 1$.

To see it let define a polygon from $X \in \mathbb{R}^{2^n}$ and let $X(a) = Pr[X = a]$. Now by

$$\begin{cases} X(a) \geq 0 & \forall a; \\ \sum_a X(a) = 1; & \\ X(a) \leq \frac{1}{2^k} & \forall a. \end{cases}$$

we give a polygon in $\mathbb{R}^{2^n}$ such that X is into it. But it is known that any point inside a polygon can be described as a convex combination of vertices of polygon. Every vertex v is described by

$$\begin{cases} X(a) = 0 & \forall a \notin v; \\ \sum_a X(a) = 1; & \\ X(a) = \frac{1}{2^k} & \forall a \in v. \end{cases}$$

So X_i is just a set of all non-zero entries in vertex v.

Now, since $Cond(X_i, U_d, U_{[m]})$ is ε-close to a distribution Y_i of min-entropy at least k, X is ε-close to a distribution $\sum_i p_i Y_i$ that has a min-entropy at least k. $\square$

Proof:[Lemma 2] Let $t = d + \log m$ as before. For each $z \in \{0,1\}^t$ consider

$$supp(z) = \{y : \Pr_{x \sim X}[Cond(x, z) = y] \geq 0\}.$$

In other words, $supp(z) = Cond(X, z) = \{y : \exists_{x \in X} Cond(x, z) = y\}$.

Claim 1: $\mathbb{E}_{z \sim U_t} |supp(z)| \geq (1 - \varepsilon) 2^k$. Proof is by considering decoding function $dec(z) = \{x \in X : Dec_x(Cond(x, z), z) = x\}$. By analyzing decoding process in Figure 3 we see

$$1 - \varepsilon \leq \frac{1}{2^k} \mathop{\mathbb{E}}_{z \sim U_t} |dec(z)|.$$

Since for every z $|supp(z)| \geq |dec(z)|$ then Claim 1 holds.

Define $A_z \in \{0,1\}^{n'}$ any set of size 2^k that contains $supp(z)$. Let Y be a distribution defined by sampling $z \sim \{0,1\}^t$ at random and outputting a random element from A_z. By definition, Y has a min-entropy at least k because it is a convex combination of distributions of min-entropy k.

Let define the analogous of distributions X and Y that include outputting : Y' by sampling $z \sim \{0,1\}^t$, $y \sim A_z$ and outputting (y, z); X' by sampling $z \sim \{0,1\}^t$, $x \sim X$ and outputting $(Cond(x, z), z)$.

It is easy to see that

$$||Y - Cond(X, U_t)||_{SD} \leq ||Y' - (Cond(X, U_t), U_t)||_{SD}$$

because one can always ignore first t bits to test the statistical distance. Now, to finish the whole proof it is enough to show the following claim.

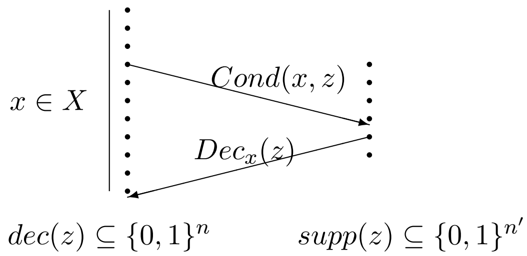

Figure 3: Decoding condenser for fixed z. For each $x_1 \neq x_2$ such that $Cond(x_1, z) = Cond(x_2, z)$ we loose an element in $dec(z)$ because obviously $Dec_x(Cond(x_1, z), z) = Dec_x(Cond(x_2, z), z)$. But this happens with probability at most ε.

Claim 2: $||Y' - X'|| \leq \varepsilon$. By straightforward calculation,

$$||Y' - X'|| = \frac{1}{2} \sum_{(z,y)} |Pr[Y' = (z, y)] - Pr[X' = (z, y)]| =$$

$$\frac{1}{2} \sum_{z \in \{0,1\}^t, y \in A_z} |Pr[Y' = (z, y)] - Pr[X' = (z, y)]| =$$

$$\frac{1}{2} \left(\sum_{z \in \{0,1\}^t, y \in supp(z)} |Pr[Y' = (z, y)] - Pr[X' = (z, y)]| + \sum_{z \in \{0,1\}^t, y \in A_z - supp(z)} |Pr[Y' = (z, y)] - 0| \right) =$$

$$\frac{1}{2} \left(1 - \frac{1}{2^t} \sum_{z \in \{0,1\}^t} \frac{supp(z)}{2^k} + \frac{1}{2^t} \sum_{z \in \{0,1\}^t} \frac{2^k - supp(z)}{2^k} \right) = \frac{1}{2} \left(2 - \frac{2\,\mathbb{E}\,|supp(z)|}{2^k} \right) \leq \varepsilon.$$

□