U.C. Berkeley | Handout N20
CS294: Pseudorandomness and Combinatorial Constructions | November 8, 2005
Professor Luca Trevisan | Scribe: Alexandra Kolla

Notes for Lecture 20

In the previous two lectures, we have seen that we can use the analysis of the Nisan - Wigderson generator to argue that starting from a distribution X of min-entropy at least k over strings of length n, if our construction is not a (k, ϵ) extractor, then we obtain a short description for a non-negligible fraction of X. For this to be a contradiction, we need $k > O(m^2) + log\frac{1}{\epsilon}$. This forces k to be large $(\Omega(n))$ if we want to use only $t = O(logn)$ truly random bits. In order to achieve extractors for distributions having smaller min-entropy k, we will need to pre-process the input random source using a condenser and output a shorter string close to a distribution of the same min entropy as the original one.
To do that, we apply again the same construction used in the previous lecture based on the NW generator but now we pick the output length m to be much bigger than k. Therefore, the output cannot be close to uniform, our construction cannot be an extractor and we have a short description of a non-negligible fraction of X. Now we can proceed as follows : On input x, we give as output the string that corresponds to the short description of x. This string will be of length $m \cdot 2^a = \sqrt{n}$ and we will enable us to reconstruct x w.h.p (since we are using a suitable ECC for x), thus preserving the entropy.

Formally, in the NW generator wi use the following parameters:
m subsets of $\{1, \ldots, d\}$ $S_1, \ldots, \mathcal{S}_m$.
$|S_i| = l$
$|S_i \cap S_j| \leq a$.

For $f : \{0,1\}^l \to \{0,1\}$ we denote by $NW^f(z) = f(z|S_1) \cdots f(z|S_m)$

We use an error-correcting code $ECC : f : \{0,1\}^n \to \{0,1\}^{\bar{n}}$.

For $n = 2^l$ we view $ECC(x) \in \{0,1\}^{\bar{n}}$ as a function $f_x : \{0,1\}^l \to \{0,1\}$. We denote $NWE(x, z) = NW_x^f(z) = f_x(z|S_1) \cdots f_x(z|S_m)$

We will first consider the case where X is uniform over a set of size 2^k and therefore has min-entropy exactly k. In following lectures we will generalize for min - entropy $\geq k$.

For $m >> k$ the output cannot be close to uniform, therefore there is a statistical test that distinguishes it from uniform. By a hybrid argument, we can conclude that there is an i such that $f_x(z|S_i)$ can be predicted given $f_x(z|S_1) \cdots f_x(z|S_{i-1})$. Equivalently, for $w \in \{0,1\}^l$, $z|S_i = w, z|[d] - S_i$ random, $f_x(w)$ can be predicted given $f_x(z|S_1) \cdots f_x(z|S_{i-1})$. Each of those functions depend only on $< a$ bits of n and they need at most 2^a values to be stored in a table. This gives us a total of $m \cdot 2^a$ bits of information as promised.

Idea : input x, output $m \cdot 2^a$ bits of information. Since we are using ECC(x) (the appropriate choice is to be specified later), x can be reconstructed from $m \cdot 2^a$ bits, which will ensure that we have almost the same entropy in the output. However, there could be a catch : it could be the case that only a small fraction of the z allow us to predict x w.h.p.

In order for our condenser to succeed we want to look at the output and be able to reconstruct the input w.h.p for almost every choice of i and z. Formally :
Take $m > \frac{10}{\epsilon}k$. We want to predict $f_x(z|S_i)$ with probability $\geq 1-\epsilon/10$, given $f_x(z|S_1)\cdots f_x(z|S_{i-1})$

To achieve our goal, we would like a predictor function as follows :

Fix random source X uniform over a set of size 2^k Fix particular z Let $x \sim X$, $x \in \{0,1\}^n$, $f_x = ECC(x)$, $i \sim [1 \cdots m]$ u.a.r. (*) given $f_x(z|S_1)\cdots f_x(z|S_{i-1})$ want to compute $f_x(z|S_i)$ with probability $\geq 1 - \frac{\epsilon}{10}$ over the distribution of x and the choice of i

In order to be able to accomplish (*), let's first look at the Shannon entropy of the distribution $NW^{f_x}(z)$. By definition,

$$H(Y) = \sum_{a:Pr[a]\neq 0} Pr[Y=a] log \frac{1}{Pr[Y=a]}$$

Since NW is a deterministic procedure, it can only decrease the entropy (the probabilities of the events can only get larger). Therefore,

$$k \geq H(f_x(z|S_1)\cdots f_x(z|S_m)) = H(f_x(z|S_1)) + H(f_x(z|S_2)|f_x(z|S_1)) + \cdots + H(f_x(z|S_m)|f_x(z|S_1)\cdots f_x(z|S_{m-1}))$$

The left-hand-side of this sum has m terms, each of those measuring how much 'fresh' information there is given the previous bits. On average, this information is only $k/m = \epsilon/10$:

$$\mathop{\mathbb{E}}_{i\sim[m]} H(f_x(z|S_i)|f_x(z|S_1)\cdots f_x(z|S_{i-1})) \leq k/m \leq \epsilon/10$$

Now we are ready to define the predictor that will allow us to accomplish (*) above:

When we want to compute $f_x(z|S_i)$ output 1 if $Pr[f_x(z|S_i)=1|f_x(z|S_1)\cdots f_x(z|S_{i-1})] > Pr[f_x(z|S_i)=0|f_x(z|S_1)\cdots f_x(z|S_{i-1})]$ output 0 otherwise

In the above, the probabilities are taken over the distribution of x.

Now suppose that $f_x(z|S_1) = b_1 \cdots f_x(z|S_{i-1}) = b_{i-1}$. There are only some of the original x that can lead to these values. Over the distribution of those x's, let

$$Pr[f_x(z|S_i) = 1 | f_x(z|S_1) = b_1 \cdots f_x(z|S_{i-1}) = b_{i-1}] = p_{b_1,\cdots,b_{i-1}} = p$$

It follows that conditioning on those values, the predictor will be wrong with probability $\min\{p, 1-p\} \le plog\frac{1}{p} + (1-p)log\frac{1}{1-p} = H(p)$.
Let $\digamma(i,z)$ be the event that the predictor is wrong for the specific i and z. Taking probability over the distribution X :

$$Pr_{x \sim X}[\digamma(i,z)] \le \sum_{b_1,\cdots,b_{i-1}} Pr[f_x(z|S_i) = b_1, \cdots, f_x(z|S_{i-1}) = b_{i-1}] \cdot H(p_{b_1,\cdots,b_{i-1}}) =$$

$$= H(f_x(z|S_i)|f_x(z|S_1) \cdots f_x(z|S_{i-1}))$$

If we want to choose i as well :

$$Pr_{x \sim X, i \sim [m]}[\digamma(i,z)] \le \mathop{\mathbb{E}}_i \sum_{b_1,\cdots,b_{i-1}} Pr[f_x(z|S_i) = b_1, \cdots, f_x(z|S_{i-1}) = b_{i-1}] \cdot H(p_{b_1,\cdots,b_{i-1}}) =$$

$$= \mathop{\mathbb{E}}_i (H(f_x(z|S_i)|f_x(z|S_1) \cdots f_x(z|S_{i-1}))) \le k/m = \epsilon/10$$

Therefore the algorithm we specified above is correct with probability $\ge 1 - \epsilon/10$ over the choice of i and x. We can now conclude that for every z there is a function p_z (the predictor defined above) such that:

$$= Pr[p_z(f_x(z|S_1) \cdots f_x(z|S_{i-1}) = f_x(z|S_i))] \ge 1 - \epsilon/10$$

We are now ready to define our condenser:

$Cond(x,z,i)$ with $z \in \{0,1\}^d, i \in [m]$ Compute $f_x = ECC(x)$, view f_x as function $f_x : \{0,1\}^l \to \{0,1\}$ for $j = 1, \cdots, i-1$ { for every z' that differs from z only in $S_i \cap S_j$ output $f_x(z'|S_j)$} output z, i

In the rest of the lecture, we will present the main lemma which will allow us later to prove that indeed the output of the condenser is ϵ-close to a distribution with the same min-entropy as the original one. Intuitively, we want to prove that the output of the condenser doesn't loose much entropy, and for this to be proved we will need a deterministic reconstruction procedure that can reconstruct the input x of the condenser with high probability. More precisely :

Lemma 1 ***Main Lemma*** *Assuming that the ECC has min-distance $> \bar{n}/5$, there is a deterministic function Dec such that*

$$Pr_{x \sim X, Z \sim U_d, i \sim [m]}[Dec(Cond(x, z, i)) = x] \geq 1 - \epsilon$$

Proof: Let us first describe what Dec should do :

$Dec(C)$ with $C = Cond(x, z, i), z \in \{0,1\}^d, i \in [m]$ For every $w \in \{0,1\}^l$ define z' such that $z'\|S_i = w$ $z'\|[d] - S_i = z$ Compute $p_{z'}(f_x(z'\|S_1) \cdots f_x(z'\|S_{i-1}) = g(w)$ output the unique x such that f_x, g are $\frac{1}{10}$-close if such an x exists otherwise output ERROR

In order to prove our lemma, it is enough to prove the following claim :

Claim 2 *With probability $\geq 1 - \epsilon$ over z, i, x g and f_x agree on more than 0.9 fraction of the inputs.*

The lemma will follow from the properties of our error-correcting code with the correct choice of min-distance as stated. Proof:(Claim)

$$Pr_{w \sim \{0,1\}^l, x \sim X, Z \sim U_d, i \sim [m]}[g(w) = f_x(w)] \geq 1 - \epsilon/10$$

This follows from the fact that for each specific z' the probability is $\geq 1 - \epsilon/10$ therefore the same should hold for the average. By a Markov argument, we get :

$$Pr_{x \sim X, z \sim U_d, i \sim [m]}[Pr[g(w) = f_x(w)] \geq 0.9] \geq 1 - \epsilon$$

With the suitable ECC for x, the above is just the probability that we retrieve x, therefore :

$$Pr_{x \sim X, z \sim U_d, i \sim [m]}[Dec(Cond(x, z, i)) = x] \geq 1 - \epsilon$$

which concludes the proof. □

□