Lecture 28: Extracting Randomness from Trevisan's Extractors (Continued)

instructor: Jin-Yi Cai Scribe: Matthew Anderson

1 Overview

Recall that we want to show:

$$Pr_{u \leftarrow X, y \in_u \{0,1\}^d}[A(y, \bar{u}(y|_{S_1}), ..., \bar{u}(y|_{S_{i-1}})) = \bar{u}(y|_{S_i})] \leq \frac{1}{2} + \frac{\epsilon}{m}. \quad (1)$$

Let α be this probability, then there exist a 0-1 assignment on the bits $y|_{[d]-S_i}$ such that the probability is at least α, therefore it suffices to show

$$\beta = Pr_{u \leftarrow X, x \in_u \{0,1\}^l}[A(y(x), f_1(x|_{S_1 \cup S_i}), ..., f_{i-1}(x|_{S_{i-1} \cup S_i}))) = \bar{u}(x)] \geq \alpha, \quad (2)$$

where $x = y|_{S_i}$ and $f_1, f_2, ..., f_{i-1}$ are boolean functions from $|S_j \cup S_i|$ bits to $\{0, 1\}$. This shows there is a sequence of function which do not depend on u or x.

Consider the set f_i which is the sequence of $i - 2$ boolean functions each on $|S_j \cup S_i|$ bits. Every such function can be specified by $\sum_{j<i} 2^{|S_j \cup S_i|}$ bits as a concatenation of the bit strings, in particular $log_2(|F_i|) \leq \sum_{j<i} 2^{|S_j \cup S_i|}$. Let $B = \{u | \exists \vec{f} \in F_i, Pr_x[A(y(x), \vec{f}(x)) = \bar{u}(x)] > \frac{1}{2} + \frac{\epsilon}{m}\}$. Think of what each sequence in F_i, together with A as defining a boolean function on $x \in \{0, 1\}^l$. Now switch the perspective from such a boolean function to a truth table and finally to a bit string.

A little difference in the relative radius of the ball ($\frac{1}{2}$ vs. $\frac{1}{2} - \frac{\epsilon}{m}$) is a big deal in terms of number of included codes.

When viewed as a Hamming ball of relative radius less than $\frac{1}{2} - \frac{\epsilon}{2m}$ each $\vec{f} \in F_i$ defines a ball which contains less than $\left(\frac{2m}{\epsilon}\right)^2$ many $\bar{u}$'s. We can estimate the value of β from before:

$$\begin{aligned}
\beta &= Pr[u \in B] \cdot Pr[...|u \in B] + Pr[u \notin B] \cdot Pr[...|u \notin B] \\
&\leq 2^{(m-1)\rho} \left(\frac{2m}{\epsilon}\right)^2 \cdot 2^{-k} + \frac{1}{2} + \frac{\epsilon}{2m} \\
&< \frac{1}{2} + \frac{\epsilon}{m} \\
&\sim \frac{k}{m}
\end{aligned} \quad (3)$$

We can take the second step by recalling we set $\rho = \frac{k - 3log(\frac{m}{\epsilon}) - 3}{m}$. We can make the last step because $log(m) < m \leq k = H_\infty(x)$.

Theorem 1.1 *$\forall n, k, m \in N, \epsilon > 0, m \leq k \leq n$, there exists an explicit (k, ϵ)-strong extractor $(2^n \times 2^d \to 2^m)$ with:*

1. *$d = O\left(\frac{log^2(\frac{n}{\epsilon})}{log(\frac{k}{m})}\right), (m < k)$.*

2. $d = 0\left(log^2(\frac{n}{\epsilon}) \cdot log(\frac{1}{\gamma})\right), (m \geq k)$.

where $1+\gamma = \frac{k}{m-1}$ *and* $\gamma < \frac{1}{2}$.

One consequence of this theorem is that if ϵ is a small constant then we have $k \sim \rho m$ and $n = d^{\Omega(l)}$ and $\frac{1}{\rho}$ is the fraction of the entropy which is finally extracted.

2 Design Construction

We will first show existence using a randomized proof, then show construction.

Define $S_1, ..., S_m \subseteq [d]$ with each $|S_i| = l$ and for all $i \sum_{j<i} 2^{|S_i \cup S_j|} < \rho(m-1)$.

Lemma 2.1 *$\forall l, m \in N, \rho > 1$, there exists a weak (l,ρ)-design such that $d = \lceil \frac{l}{log(\rho)} \rceil \cdot l$ and it can be found in polytime(m,d).*

Proof: Construct l sets $B_1, ..., B_l$, with each $|B_i| = \lceil \frac{l}{log(\rho)} \rceil \cdot l$. Each S_i will have one point in each B_j.

Base case, assign the first element of each B_j to be in S_1. Inductively assume S_1 to S_{i-1} have been found $(i \geq 2)$. Probabilistically pick the k^{th} point of S_i uniformly on B_k. Let $|S_j \cup S_i \cup B_k| = Y_{j,k}$; $Y_{j,i}$ is a 0-1 random variable. Notice $|S_i \cup S_j| = \sum_{k=1}^{l} |S_i \cup S_j \cup B_k|$. Then,

$$\begin{aligned} E[\sum_{j<i} 2^{|S_i \cup S_j|}] &= \sum_{j<i} E[2^{\sum_k Y_{j,k}}] \\ &= \sum_{j<i} \prod_{k=1}^{l} E[2^{Y_{j,k}}] \\ &= \sum_{j<i} \left(\frac{1}{\lceil \frac{l}{log(\rho)}}\right)^l \\ &\leq (i-1)\rho \leq (m-1)\rho \end{aligned} \tag{4}$$

The second step in the above equation requires an independence assumption on $Y_{j,k}$. This proves existence of such a design $(E[X] < y \Rightarrow \exists x \in X, x < y)$. ∎

To construct this design (or derandomize the above construction), we use a method of conditional expectations. We will evaluate the expectation of $\sum_{j<i} 2^{|S_i \cup S_j|}$ conditional to the first $(k-1)$ points in $B_1, ..., B_{k-1}$ of S_i, which are $\alpha_1, ..., \alpha_{k-1}$ respectively. Then:

$$\begin{aligned} E[\sum_{j<i} 2^{|S_i \cup S_j|} | S_i \cup B_s = \alpha_s, 1 \leq s \leq k-1] &= \sum_{j<i} E[2^{|S_i \cup S_j|} | S_i \cup B_s = \alpha_s, 1 \leq s \leq k-1] \\ &= \sum_{j<i} 2^{\left(\sum_{s=1}^{k-1} X_{\{\alpha_s \in S_i\}}\right)} \cdot \left(1 + \frac{1}{\lceil \frac{l}{log(\rho)} \rceil}\right)^{p-(k-1)} \end{aligned} \tag{5}$$

This value is easy to compute in polynomial time because the exponent of the two is a constant, so we can compute this conditional expected value and use to derandomize the construction, always choosing the branch that will give us the smallest expectation.

The weak design does, in fact, buy you something.

LEMMA 2.2 *$S_1, ..., S_m \subset [d]$, and an $(l, /rho)$-design then $m \leq \frac{\binom{d}{log_2(\rho)}}{\binom{l}{log_2(\rho)}}$, in particular, $d \geq m^{\frac{1}{log_2(\rho)}}(l-log(\rho))$.*

Proof: Let $I = max_{i \neq j}|S_i \cup S_j| \leq \lfloor log_2(\rho) \rfloor$. For $j = 1, ..., m$, let $\Gamma_j = \{T \subseteq S_j || T| = I + 1\}$. Notice, $|\Gamma_j| = \binom{l}{I+1}$ and all the Γ_j's are pairwise disjoint. Then $\binom{d}{I+1} \geq |\bigcup_j \Gamma_j| = m\binom{l}{I+1}$.

So we have:

$$\begin{aligned} m &\leq \frac{\binom{d}{I+1}}{\binom{l}{I+1}} \\ &\leq \frac{d(d-1)...}{l(l-1)...} \\ &\leq \frac{d}{l - log(\rho)}. \end{aligned} \tag{6}$$

■