Pseudorandomness: Lecture 20.

Instructor: Ramprasad Saptharishi
Date: 2021-11-09
Lecture: #20.

Agenda:
- Extractors from expanders and hash families
- Block sources
- Extractors for block sources.

Recap:
- Min-entropy: $X \subseteq \{0,1\}^n$

$$\text{min entropy}(X) = \min_{p_x \neq 0} \log\left(\frac{1}{p_x}\right)$$

$$\text{min-entropy}(X) \geq k \Rightarrow \Pr[X = x] \leq 1/2^k \quad \forall x$$

(convex comb of flat k-sources)

- Seeded Extractors: $\text{Ext}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε)-extractor if, for every k-source X,

$$\text{Ext}(X, \mathcal{U}_d) \overset{TV}{\approx}_{\varepsilon} \mathcal{U}_m.$$

- Ideal parameters: $d = O(\log(n/\varepsilon))$, $m = k + d - O(\log\frac{1}{\varepsilon})$

Final thm we will prove:

Thm: [... Guruswami-Umans-Vadhan] Let $\alpha > 0$ be a constant. For all $0 \leq k \leq n$, and $\varepsilon > 0$, there is an explicit (k,ε)-ext

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

with $m = (1-\alpha)k$ and $d = O(\log(n/\varepsilon))$.

(Another with $m = k - O(\log 1/\varepsilon)$ and $d = O(\log k \cdot \log(n/\varepsilon))$.

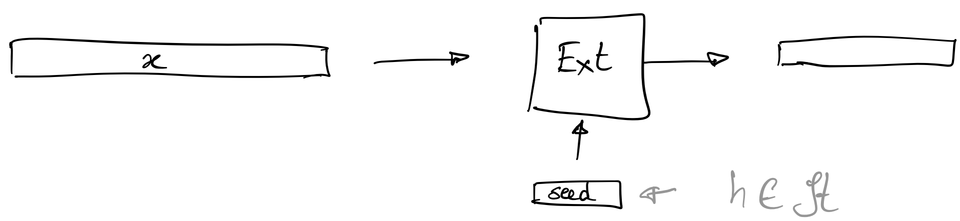
$\text{Ext}'(x,y) = y, \text{Ext}(x,y)$

Strong extractors: $\{(y, \mathrm{Ext}(X,y))\}_{y\sim\mathcal{U}_d} \overset{TV}{\approx_\varepsilon} \mathcal{U}_{d+m}$

"remains an extractor even if we 'reveal' the seed"

These also exist with $d = O(\log n/\varepsilon)$ $\quad m = k - \log 1/\varepsilon$.

Towards building extractors:



Qn: Are hash functions extractors?!

Lemma: [Leftover Hash Lemma] Let $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^m\}$ be a p.i.h.f. with $m = k - 2\log 1/\varepsilon$. Then
$$\mathrm{Ext}(x,h) = (h, h(x)) \text{ is a } (k,\varepsilon)\text{-extractor.}$$

Pf: hash functions $\approx$ collision resistance $\approx$ ℓ^2-norms

$$CP(Y) = \Pr_{y,y'\sim Y^2}[y = y'] = \sum p_y^2 = \mathop{\mathbb{E}}_{y\sim Y}\left[\Pr[Y=y]\right]$$

$$CP((H, H(X))) = \Pr[H = H'] \cdot \Pr_{\substack{h\sim\mathcal{H}\\ x,x'}}[H(x) = H(x')]$$

$$\boxed{\begin{matrix}(H, H(X))\\ (H', H'(X'))\end{matrix}} \quad \le \frac{1}{2^d}\cdot\left(\Pr[X = X'] + \Pr[H(X) = H(X') \mid X \neq X']\right)$$

$$\le \frac{1}{2^d}\left(\frac{1}{2^k} + \frac{1}{2^m}\right) \le \frac{1}{2^{d+m}}\cdot(1+\varepsilon^2)$$

$$CP(\mathcal{U}_{d+m}) = \frac{1}{2^{d+m}}$$

$$\|\underset{p_i}{(H, H(x))} - \mathcal{U}_{d+m}\|_2^2 = \sum p_i^2 - 2\frac{1}{DM}\sum p_i + \sum\left(\frac{1}{DM}\right)^2$$

$$= \frac{1+\varepsilon^2}{DM} - \frac{2}{DM} + \frac{1}{DM}$$

$$= \frac{\varepsilon^2}{DM}$$

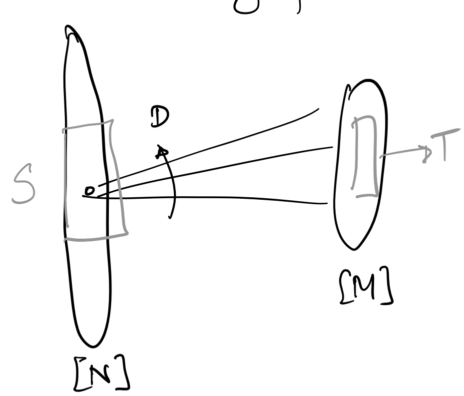
$$|(H, H(x)) - \mathcal{U}_{d+m}|_1 \leq \sqrt{DM}\cdot\frac{\varepsilon}{\sqrt{DM}} = \varepsilon. \quad \square.$$

∴ This shows that PWI HF are extractors

Cool! How good are the parameters?

- How small is $|\mathcal{H}| = 2^d$?
  - Seed length $= n$.
    this is bad!

Extractors as graphs:



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

For any subset $S \subseteq [N]$ with $|S| \geq k$, we want $\Gamma(S)$ to hit $[M]$ almost evenly.

In particular, $|\Gamma(S)| \geq (1-\varepsilon)M$.

Defn: (Dispersers) $E: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε)-disperser if for every k-source X, $E(X, U_d)$ has support at least $(1-\varepsilon)\cdot 2^m$.

... perfectly suited for RP / coRP algos.

What do we want for extractors?

$$\Delta\left(\mathrm{Ext}(U_S, U_{[D]}),\ U_{[M]}\right) = \max_{T \subseteq [M]} \left| \Pr[\mathrm{Ext}(\,) \in T] - \frac{|T|}{M} \right| \le \varepsilon \quad (*)$$

$$\Pr\left[\mathrm{Ext}(U_S, U_{[D]}) \in T\right] = \frac{|E(S,T)|}{D \cdot |S|}$$

Want $\left| \dfrac{E(S,T)}{D\cdot|S|} - \dfrac{|T|}{M} \right| \le \varepsilon$ for all T $\Leftrightarrow (*)$

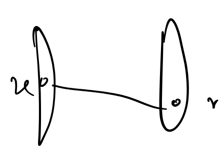
$$\Updownarrow$$

$$\left| \frac{E(S,T)}{ND} - \mu(T)\cdot\mu(S) \right| \le \varepsilon \cdot \mu(S).$$

EML: G is a λ-spectral expander

$$\Rightarrow \left| \frac{E(S,T)}{ND} - \mu(S)\mu(T) \right| \le \lambda \cdot \sqrt{\mu(S)\cdot\mu(T)}$$

Can convert a usual expander G to a bipartite "double cover".



Wait $\lambda \cdot \sqrt{\mu(S)\,\mu(T)} \leq \varepsilon \cdot \mu(S)$

$\Rightarrow \lambda \leq \varepsilon \sqrt{K/N}$ is good enough.

We know how to build $G = (N, D_0, 1/2)$-expanders for a constant D_0.

$G^t = (N, D_0^t, 1/2^t)$ - expander.

$\frac{1}{2^t} \leq \varepsilon \sqrt{K/N}$ $\quad t = \Theta(n - k - \log\frac{1}{\varepsilon})$

$\Rightarrow D = D_0^t = \exp(t) \Rightarrow d = \Theta(n - k - \log 1/\varepsilon)$

Prop [Expanders as extractors] For all n, k and $\varepsilon \geq 0$. there is an explicit extractor $Ext: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m = n$ and $d = O(n - k + \log\frac{1}{\varepsilon})$.

If $k = n - O(\log n)$, then this is quite good.
But if $k \ll n$, this sucks!

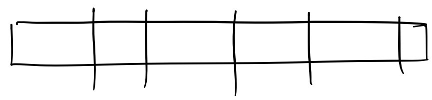
Revisiting our nemesis:

Unpredictabe Bit Sources (δ):

$$\delta \leq \Pr[X_i = 1 \mid X_1 = a_1, \dots, X_{i-1} = a_{i-1}] \leq 1 - \delta$$

How do we build extractors for these sources?

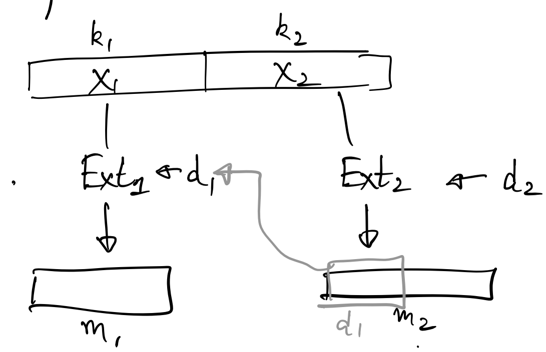
Defn: (Block sources) $X = (X_1, \dots, X_t)$ is a $(k_1, \dots, k_t)$- block source if $\forall i$: $X_i |_{X_1 = x_1, \dots, X_{i-1} = x_{i-1}}$ has min-entropy k_i for all $x_1, \dots, x_{i-1}$



Each block guarantees some min-entropy.

Obs: For any $t \le n$, UnpredBits_δ is a $t \times \alpha \frac{n}{t}$ - block source for $\alpha \approx \log \frac{1}{(1-\delta)}$

Extractors for block sources:



$m_2 \geq d_1$

Lemma: Suppose $\text{Ext}_i : \{0,1\}^{n_i} \times \{0,1\}^{d_i} \to \{0,1\}^{m_i}$ is a (k_i, ε_i) extractor for $i = 1, 2, \ldots, t$. with $m_{i+1} > d_i$.
Then, there is an explicit ε-extractor
$$\text{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$$
$d = d_t$

for $(k_1, \ldots, k_t)$- block sources with $d = d_t$, $\varepsilon = \sum \varepsilon_i$,
$$m = m_1 + (m_2 - d_1) + \cdots + (m_t - d_{t-1})$$

Pf: We'll prove it for $t = 2$.

$$(X_1, X_2) \leftarrow Z_2 \in \{0,1\}^{d_2}$$
$$\downarrow \text{Ext}_2$$
$$(X_1, Z_1, Y_2) \qquad \text{where } Z_1 Y_2 = \text{Ext}_2(X_2, Z_2)$$

$$\approx_{\varepsilon_2} \left(X_1, U_{d_1}, U_{m_2 - d_1} \right)$$

This is because for any $X_1 = x_1$, $X_2 \mid X_1 = x_1$ still has k_2 bits of entropy.

$$\xrightarrow{Ext_1} \quad (Y_1, Y_2) \approx_{\varepsilon_1} \left(U_{m_1}, U_{m_2 - d_1} \right)$$

(Check this)

$\square$.

That is, for block sources, you can build an extractor where seed length depends only on Ext_2.

... does this sound familiar? (See Sec 6.3.5 in Vadhan's text)

What we have so far:

- If min-entropy really high, then expanders work.
- PWIHFs work, but seed length too high.
- If we have a block source, then we can use a zig-zag like construction to pay for just the seed of one block.

Unpred $Bits_s$

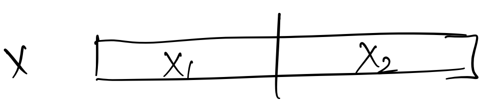
Do we know of such an ext?

Yes! PWIHFs work when length is small!

Prop: We can extract $k - \log 1/\varepsilon$ bits of randomness from UnpredBits_δ sources with $O(\log n/\varepsilon)$ seed length.

Next class: Extractors for general sources

"Reduce any source to a block source"



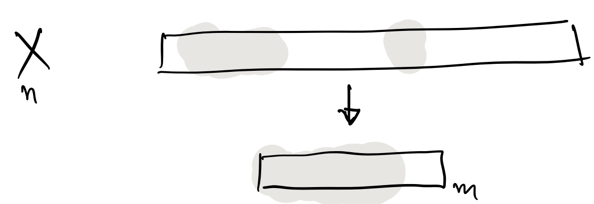
Fail. if say X_1 determines X_2

Suppose min entropy $(X) \geq 99\%$ of n.

Then, this actually works!

Lemma: If X is an $(n-\Delta)$-source and $X = (X_1, X_2)$
Then (X_1, X_2) is ε-close to a (k_1, k_2)-block src
with $k_1 = n_1 - \Delta$, $k_2 = n_2 - \Delta - \log 1/\varepsilon$.

ie if $\Delta = 0.01n$, then $(X_1, X_2) - \left(\frac{n}{2} - 0.1n, \frac{n}{2} - 0.1n - \log\frac{1}{\varepsilon}\right)$



Defn (Condenser) $\mathrm{Cond}: [N] \times [D] \to [M]$ is a $(k \to_{\varepsilon} k')$-condenser if for any k-source X,

$$\mathrm{Cond}(X, U_d) \approx_{\varepsilon} Y \qquad \text{where } Y \text{ is a } k'\text{-source.}$$

There are exp. condensers with $k' = k + d$ and $m = (1+\alpha) \cdot k$. $\quad d = O(\log n/\varepsilon)$