Pseudorandomness: Lecture 21.

Agenda: [GUV Thm]: For any const $\alpha > 0$. For all n, k and $\varepsilon > 0$, there is an explicit (k,ε)-extractor $\mathrm{Ext}: [N] \times [D] \to [M]$ with $m \geq (1-\alpha) \cdot k$ and $d = O(\log(n/\varepsilon))$.

Recap:

- If n is small, then $\mathrm{Ext}_H(x, h) = (h, h(x))$ is a (k,ε)-extractor with $m = k - 2\log 1/\varepsilon$. (but seed-length n)
- If $(X_1, \ldots, X_t)$ is a $(k \times t)$-block source, then we can extract randomness by paying for just one seed block

Something that we'll use very often this class:

Lemma: (Residual entropy) Say X is a k-source and W is a correlated RV, with $\mathrm{supp}(W) \leq 2^\ell$. Then, for any $\varepsilon > 0$, with prob $\geq 1 - \varepsilon$ over $w \sim W$, $X | W = w$ is a $(k - \ell - \log \frac{1}{\varepsilon})$-source.

Pf: $\mathrm{Bad}_W = \{ w : \Pr[W = w] \leq \varepsilon / 2^\ell \} \Rightarrow \Pr[W \in \mathrm{Bad}_W] \leq \varepsilon$

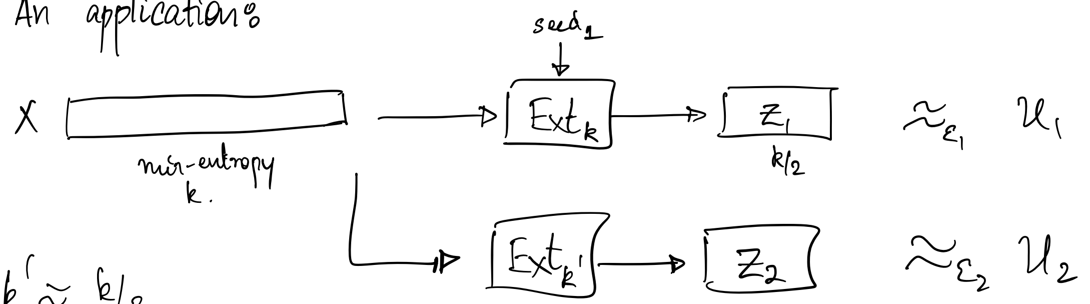
Fix any $w \notin \mathrm{Bad}_W$.

$$\frac{1}{2^k} \geq \Pr[X = x] \geq \Pr[X = x, W = x] = \Pr[W = w] \cdot \Pr[X = x | W = w]$$
$$\geq \frac{\varepsilon}{2^\ell} \cdot \Pr[X = x | W = w]$$

$$\Rightarrow \Pr[X = x | W = w] \leq \frac{1}{2^{k - \ell - \log 1/\varepsilon}}$$

$\square$.

An application:



$k' \approx k/3$

Does this work?

$$(Z_1, X) \longrightarrow (Z_1, \mathrm{Ext}_{k'}(X, \mathcal{U}_d)) = (Z_1, Z_2)$$

With prob $\geq 1-\varepsilon_3$ over $Z_1 = z_1$, we know that $X|_{Z_1=z_1}$ has entropy $\geq k - k/2 - \log 1/\varepsilon_3 \geq k/3$

With prob $\geq 1-\varepsilon_3$, $(z_1, Z_2) \approx_{\varepsilon_2} \mathcal{U}$

$\therefore$ $(Z_1, Z_2) \approx_{\varepsilon_1+\varepsilon_2+\varepsilon_3}$ of the uniform dist.

$\therefore$ If we have a way of using $O(\log n/\varepsilon)$ seed to extract $k/2$ bits, then we can also extract any $(1-\alpha)k$ bits using just $O_\alpha(\log n/\varepsilon)$ seed.

Thm [GUV weaker]: For ~~any $0<\alpha<1$~~, $n \geq k \geq 0$ $\varepsilon > 0$, there is an explicit $\mathrm{Ext}_k : [N] \times [D] \to [M]$ (k, ε)-ext. with $m = k/2$ and $d = O(\log n/\varepsilon)$.

Another application: (any high entropy source is close to a block source)

Lemma: Suppose X is an $(n-\Delta)$-source. Then for any $\varepsilon > 0$, $X = (X_1, X_2)$ is ε-close to a ~~an~~ $(n_1 - \Delta,\ n_2 - \Delta - \log 1/\varepsilon)$-source

Pf: X_1 an $n_1 - \Delta$ source:

$$\Pr[X_1 = x_1] \le \sum_{x_2} \Pr[X = x_1 x_2] \le \frac{1}{2^{n-\Delta}} \cdot 2^{n_2} = \frac{1}{2^{n_1 - \Delta}}.$$

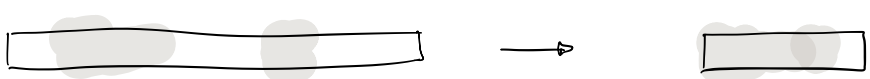
To show that $X_2 \mid X_1 = x_1$ has high-min-entropy (w.h.p. over x_1)

just use the REL.

$\Rightarrow X_2 \mid X_1 = x_1$ is an $n_2 - \Delta - \log 1/\varepsilon$ src w.p $\ge 1 - \varepsilon$ over x_1 $\square$.

$\therefore$ If min-entropy(X) is <u>really</u> high, then we can just break X into blocks and get a block source.

What if it was not this high?



Defn: (Condensers) $\mathrm{Cond}: [N] \times [D] \to [M]$ is a $(k \to k', \varepsilon)$ condenser if for any k-source X, we have that $\mathrm{Cond}(X, U_d) \overset{TV}{\approx}_{\varepsilon} Y$ where Y is a k'-source.

The condenser is loss-less if $k' = k + d$.

We would want $k'/m \gg k/n$ so that "entropy density" inc.

Thm: (Guruswami-Umans-Vadhan) For any $\alpha > 0$, and $n \geq k$ and $\varepsilon > 0$, there is an explicit $k \rightarrow_{\varepsilon} k+d$ lossless condenser

$$\mathrm{Cond}: [N] \times [D] \longrightarrow [M]$$

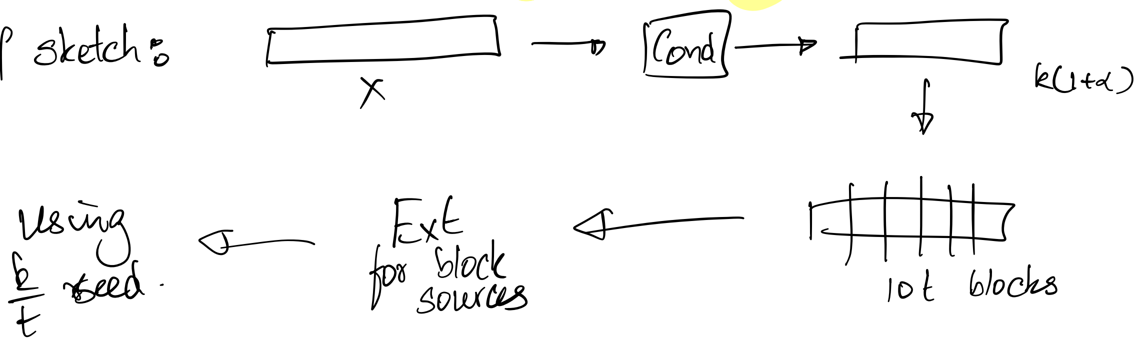
with $m = (1+\alpha)k + O(\log n/\varepsilon)$ and $d = O(\log n/\varepsilon)$.

Putting this all together.

Lemma: For any $t > 1$ and $n \geq k$ and $\varepsilon > 0$, there is an explicit (k, ε)-extractor $\mathrm{BBExt}^{(t)}: [N] \times [D] \rightarrow [M]$ with $m \geq k/2$ and $d = \frac{k}{t} + O(\log n/\varepsilon)$.

$\alpha \ll \frac{1}{t}$

Pf sketch:



$\square$.

Ex: Prove this formally (in PS4)

Thm: For all n, k, and $\varepsilon > 0$, there is an exp (k, ε)-ext
(main). $\mathrm{Ext}: [N] \times [D] \longrightarrow [M]$ with $m \geq k/2$, $d = O(\log n/\varepsilon)$.

Our lego pieces:

- $\mathrm{WLOG}_{\varepsilon}$, $n \approx k$. (Condensers)
- $\mathrm{WLOG}_{\varepsilon}$, we can assume that X is a block source.
- For block sources, on seed-block is sufficient.
- We have explicit extractors with seed length $\approx k/10000$

- If source has more entropy after 1 extraction, we can extract more using REL.

Together, we ought to have what we want... just a matter of putting them together.

Pf of GUV Weaker:

Fix $\varepsilon_0 > 0$ $\left(\varepsilon_0 = \varepsilon/\mathrm{poly}(n) \right)$ and $d = c \log n/\varepsilon_0$ for a fixed large constant c

We'll have a seq $0 < \varepsilon_0 \leq \varepsilon_1 \leq \varepsilon_2 \leq \cdots \leq \varepsilon_{\log n} = \varepsilon$

Define $i(k)$ = smallest i : $k \leq 2^i \cdot 8d$.

We'll prove: For all k, there is an exp $(k, \varepsilon_{i(k)})$ extractor $\mathrm{Ext}_k : [N] \times [D] \to [M_k]$ with $m_k = k/2$.

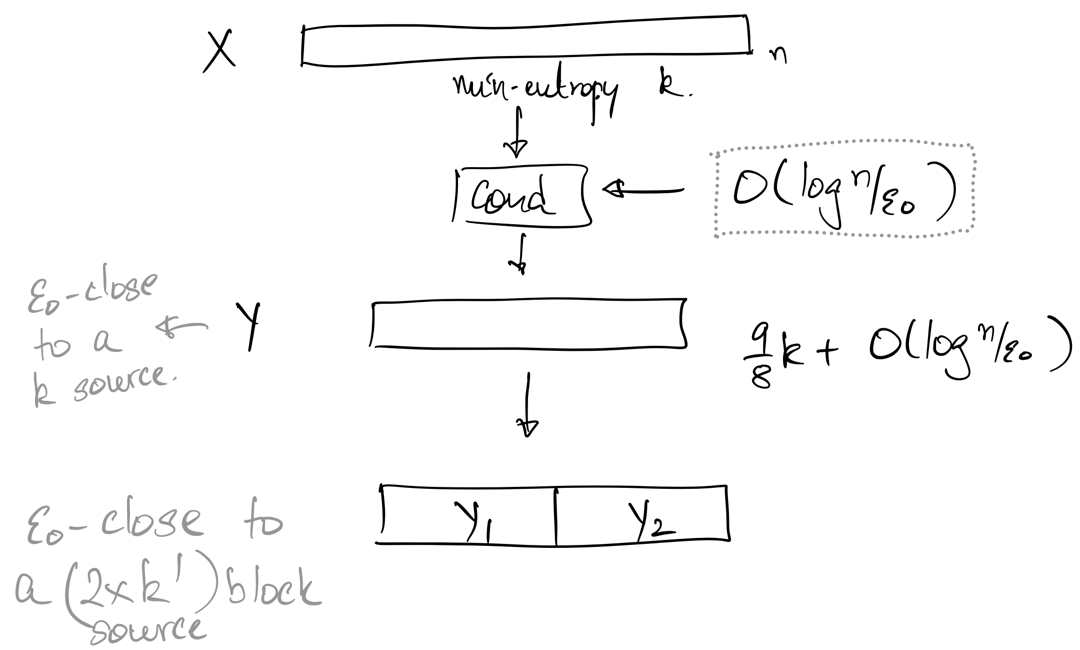
Imp: The seed length is d, nomatter what k is., but error might slightly inc.

Base case: $i(k) = 0 \Rightarrow k \leq 8d$.

$\Rightarrow$ We have such an ext. by the prev. lemma.

Inductive step: Say $i(k) \geq 1$. and we have

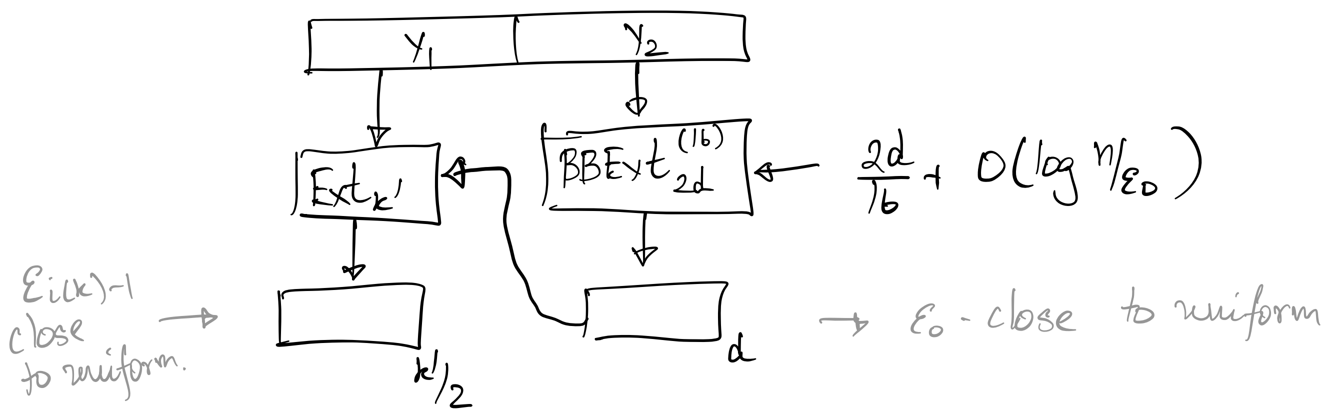
$\mathrm{Ext}_{k'}$ for all k' with $i(k') < i(k)$.



where $k' \geq \frac{k}{2} - \frac{k}{8} - O(\log n/\varepsilon_0)$

Note that $\frac{k}{3} \leq k' \leq k/2 \Rightarrow i(k') < i(k)$

$k \geq 8d$ $\qquad k' \geq 2d$



But we wanted $k/2$ random bits... we only have $k/6$ bits

$\Rightarrow$ there are $5k/6$ bits still in the system. [REL]

$\Rightarrow$ We can extract another $1/6$th of that.

$\Rightarrow$ There are $(5/6)^2$ bits still in the system.

$\left(\frac{5}{6}\right)^4 < \frac{1}{2} \Rightarrow$ 4 applications of REL gives us what we want.

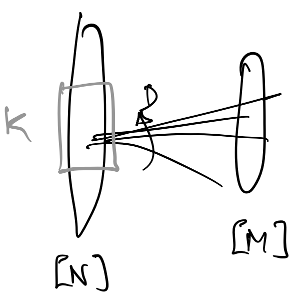
$\therefore$ Total seed : $4\left(\underbrace{O(\log\frac{n}{\varepsilon_0})}_{\text{cond.}} + \frac{2d}{16} + \underbrace{O(\log\frac{n}{\varepsilon_0})}_{\text{BB-Ext}}\right)$

$\le d$

Total error : $\varepsilon_{i(k)} \le 4\left(3\varepsilon_0 + \varepsilon_{i(k)-1}\right) = 16 \cdot \varepsilon_{i(k)-1}$

And $i(n) = \log n \Rightarrow \varepsilon_{i(n)} = \varepsilon_0 \cdot \mathrm{poly}(n) \le \varepsilon$. $\square$

Condensers as graphs:



If X is a k-source on $[N]$, then, $\Gamma(X)$ must be ε-close to a k'-source on $[M]$.

set of size K

$\Rightarrow |\Gamma(X)| \ge kD\cdot(1-\varepsilon)$ if lossless.

Fact: If Cond: $[N]\times[D] \to [M]$ is a $k \to_\varepsilon k+d$ lossless condenser, then G is a $(=K, D(1-\varepsilon))$-vertex expander.

($\Gamma(S)$ is almost distinct for any S of size K).

"unbalanced expanders" (as we want $m \ll n$.)
with almost optimal expansion on size K sets

GUV Graph: $G : \mathbb{F}_q^n \times \mathbb{F}_q \to \mathbb{F}_q^{m+1}$. Fix $E(x) \in \mathbb{F}_q[x]$, irred of degree $n-1$

$$\mathrm{Rot}_G(f, y) = [f^{(0)}(y), f^{(1)}(y), \ldots, f^{(m)}(y)]$$

where $f^{(i)} = f^{h^i} \bmod E(x)$. ($h$ is a parameter).

(Based on Parvaresh - Vardy codes).

Lemma: The above graph is a $(\leq K, A)$-vertex expander for $K = h^m$ and $A = q - (n-1)(h-1).m$.

Can suitably set parameters to get

$$d = \left(1 + \frac{1}{\alpha}\right) \cdot \log\left(\frac{2nk}{\varepsilon}\right), \quad m \leq 2d + (1+\alpha).k$$

Next: Trevisan's extractor (Interpreting the NW PRG in this framework).