Class 19

Extractors and Expanders
Warm-up

• Say that $X$ is a $k$-source on $\{0,1\}^n$. Let $N = 2^n$.
• Let $\sigma \in \mathbb{R}^N$ correspond to the pmf for $X$.

1. Why is $\|\sigma\|_\infty \leq 2^{-k}$?
2. Argue that $\|\sigma\|_2 \leq 2^{-k/2}$.
   • Hint: $\|x\|_2^2 \leq \|x\|_\infty \|x\|_1$
Announcements

• Welcome to week 10!!!
• HW8 due Friday.
• Practice exam is out now. (With solutions).
  • We hope it’s about the same difficulty as the real final, although TBH I think it’s not as “good” an exam as the real final... (good exams are hard to write).
• Today:
  • Pseudorandomness! Not on the exam.
• Thursday:
  • Research talks! Also not on the exam.
• EXAM: Thursday 12/15, 12:15-3:15pm, Room 420-040.
Pseudorandomness

• Deterministic (or not-so-random) objects that behave like random ones.
• Useful for derandomization.
Extractors

- $n$-bit “weak” source of randomness
- $d$-bit “seed” of uniform randomness
- $m$-bit output that is “close” to uniformly random.
Expanders

• Let $G = (V, E)$ be an unweighted, undirected, regular graph with degree $D$ and with $N$ vertices.
• Let $A$ be the normalized adjacency matrix of $G$.
• Say that the eigenvalues of $A$ are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$

• The expansion of $A$ is $\lambda(G) = \max \{ \lambda_2, |\lambda_n| \}$

Theorem:

• Let $\{X_t\}$ be a random walk on $G = (V, E)$.
• The stationary distribution of $\{X_t\}$ is $\pi = \text{uniform on } V$.
• If $\lambda(G) < 0.99$, then $\tau_{\text{mix}} = O(\log n)$
Questions?
Minilectures, Warm-up?

Warm-up:
• Say that $X$ is a $k$-source on $\{0,1\}^n$. Let $N = 2^n$.
• Let $\sigma \in \mathbb{R}^N$ correspond to the pmf for $X$.
1. Why is $\|\sigma\|_\infty \leq 2^{-k}$?
2. Argue that $\|\sigma\|_2 \leq 2^{-k/2}$.
   • Hint: $\|x\|_2^2 \leq \|x\|_\infty \|x\|_1$
Warm-Up

• Say that $X$ is a $k$-source on $\{0,1\}^n$
• Let $N = 2^n$, and let $\sigma$ be the “vectorized” version of the distribution of $X$

1. Why is $\|\sigma\|_\infty \leq 2^{-k}$?
Warm-Up

• Say that $X$ is a $k$-source on $\{0,1\}^n$
• Let $N = 2^n$, and let $\sigma$ be the “vectorized” version of the distribution of $X$

2. Argue that $\|\sigma\|_2 \leq 2^{-k/2}$.
Warm-Up

• Say that $X$ is a $k$-source on $\{0,1\}^n$
• Let $N = 2^n$, and let $\sigma$ be the “vectorized” version of the distribution of $X$

Then $\|\sigma\|_{\infty} \leq 2^{-k}$, by def of $k$-source:

$H_\infty(X) \leq 2^{-k}$
Warm-Up

• Say that $X$ is a $k$-source on $\{0,1\}^n$

• Let $N = 2^n$, and let $\sigma$ be the “vectorized” version of the distribution of $X$

$$\left\| \sigma \right\|_2 = \left( \sum_{i \in [N]} \sigma_i^2 \right)^{1/2} \leq \left\| \sigma \right\|_\infty \left( \sum_{i \in [N]} \sigma_i \right)^{1/2} = \left\| \sigma \right\|_\infty \leq 2^{-k/2}$$

$$\sigma_i = \Pr[X = i], \quad \forall i \in \{0, \ldots, N-1\}$$

For, the binary expansion of $i$. 
Today

• We will consider a way to make an extractor out of an expander graph.
Today

• We will consider a way to make an extractor out of an expander graph.

• Recall: An extractor looks like this:
Today

• We will consider a way to make an extractor out of an expander graph.

• Recall: An expander graph looks like this:

Degree D graph with N vertices

Normalized adjacency matrix
$A \in \mathbb{R}^{N \times N}$

• The eigenvalues of $A$ are
$1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$

• The expansion is
$\lambda(G) = \max\{\lambda_2, |\lambda_N|\}$

• For an expander, $\lambda(G)$ is decently less than 1.

This is $\frac{1}{D}$ times the standard adjacency matrix.
Our extractor

• Associate each vertex of $G$ with a string in $\{0,1\}^n$

$N = 2^n$, and choose $k \leq n$ and $\epsilon > 0$

Let $d = \ell \cdot \log(D)$, where $\ell = \frac{n-k}{2} + \log\left(\frac{1}{\epsilon}\right)$
Our extractor

• Associate each vertex of $G$ with a string in $\{0,1\}^n$
• Take a random walk on $G$, starting from $x \sim X$, and following a random walk given by the seed $s \sim U_d$. 

$N = 2^n$, and choose $k \leq n$ and $\epsilon > 0$

Let $d = \ell \cdot \log(D)$, where $\ell = \frac{n-k}{2} + \log\left(\frac{1}{\epsilon}\right)$
Our extractor

- Associate each vertex of $G$ with a string in $\{0,1\}^n$
- Take a random walk on $G$, starting from $x \sim X$, and following a random walk given by the seed $s \sim U_d$.

$G = \text{Degree D graph with N vertices}$

$N = 2^n$, and choose $k \leq n$ and $\epsilon > 0$

Let $d = \ell \cdot \log(D)$, where $\ell = \frac{n-k}{2} + \log \left( \frac{1}{\epsilon} \right)$
Our extractor

- Associate each vertex of $G$ with a string in $\{0,1\}^n$
- Take a random walk on $G$, starting from $x \sim X$, and following a random walk given by the seed $s \sim U_d$.

\[ N = 2^n, \text{ and choose } k \leq n \text{ and } \epsilon > 0 \]
\[ \text{Let } d = \ell \cdot \log(D), \text{ where } \ell = \frac{n-k}{2} + \log\left(\frac{1}{\epsilon}\right) \]

- The source $x \sim X$ tells us a vertex to start at.
- For each step $1, 2, \ldots, \ell$, that chunk of the seed tells us what our next step should be.
- Output the label on the vertex where we are after $\ell$ steps.
Claim

• If we choose $\ell = \frac{n-k}{2} + \log \left(\frac{1}{\epsilon}\right)$, then this is a $(k, \epsilon)$-extractor.
  • Seed length: $d = \ell \cdot \log(D) = O(\ell)$
  • Output length: $n$

This is not as good as our existential result, since the seed length is really long unless $k$ is quite large, but it’s still non-trivial!

This is pretty good when $k = n - \log n$, for example.
Comparison to optimal

- $\ell = \frac{n-k}{2} + \log\left(\frac{1}{\epsilon}\right)$
- Seed length $d = \ell \cdot \log(D) = O\left(\frac{n-k}{2} + \log\left(\frac{1}{\epsilon}\right)\right)$
- Output length: $n$

---

- Seed length $d = k + d - 2 \log\left(\frac{1}{\epsilon}\right) - O(1)$
- Output length: $\log(n - k) + 2 \log(1/\epsilon) + O(1)$

The seed length is much longer than we’d like unless $k$ is big. However, the output length in that case is pretty good: ideally it wouldn’t be much smaller than $k + d$ (the total amount of randomness going in), so it’s the right order of magnitude.
Group Work: prove the claim!

1. Let $\sigma \in \mathbb{R}^n$ represent the probability mass function of our input $X$. Explain why $Ext(X, U_d) \sim A^\ell \cdot \sigma$, where $A$ is the normalized adjacency matrix for $G$.

2. Let $\pi = \frac{1}{N} \mathbf{1}$ correspond to the uniform distribution. Explain why

$$\|U_n - Ext(X, U_d)\|_{TV} = \|\pi - A^\ell \cdot \sigma\|_{TV} \leq \frac{\sqrt{N}}{2} \lambda(G)^\ell \|\pi - \sigma\|_2$$

3. Argue that $\|\pi - \sigma\|_2 \leq 2 \cdot 2^{-k/2}$

4. Conclude that $\|U_n - Ext(X, U_d)\|_{TV} \leq \epsilon$, which means that $Ext$ is a $(k, \epsilon)$-extractor.

• If we choose $\ell = \frac{n-k}{2} + \log \left(\frac{1}{\epsilon}\right)$, then this is a $(k, \epsilon)$-extractor.
  - Seed length: $d = \ell \cdot \log(D) = O(\ell)$
  - Output length: $n$
Solutions
1. Why is $\text{Ext}(X, U_d) \sim A^\ell \cdot \sigma$
1. Why is $\text{Ext}(X, U_d) \sim A^\ell \cdot \sigma$

• By definition, $\sigma$ is the distribution of $X$, the starting distribution for our random walk.

• The normalized adjacency matrix $A$ is the transition matrix for the random walk on $G$.

• Since $U_d$ is uniformly random, we just take an $\ell$-step random walk on $G$ starting from the distribution $\sigma$ to get the output of Ext.

• The distribution of that is $A^\ell \cdot \sigma$, as we saw before when we studied Markov chains.
2. Bounding $\|U_n - Ext(X, U_d)\|

$\|U_n - Y\|_{TV} = \frac{1}{2} \|\pi - A^l\sigma\|_1$

- Let $Y = Ext(X, U_d)$
- Let $\pi$ be the uniform distribution.
2. Bounding $\|U_n - Ext(X, U_d)\|$

$$\|U_n - Y\|_{TV} = \frac{1}{2} \|\Pi - A^l \sigma\|_1$$

$$\|\Pi - A^l \sigma\|_1 = \|A^l (\Pi - \sigma)\|_1$$

Since $A\pi = \pi$

$$\leq \sqrt{N} \|A^l (\Pi - \sigma)\|_2$$

Cauchy-Schwarz

$$\leq \sqrt{N} \lambda(G)^l \|\Pi - \sigma\|_2$$

Since $\pi - \sigma \perp \pi$, and $\pi$ is the top eigenvector.
3. Bounding $||\pi - \sigma||_2$

$$||\pi - \sigma||_2 \leq ||\pi||_2 + ||\sigma||_2$$
3. Bounding $\|\pi - \sigma\|_2$

\[
\|\pi - \sigma\|_2 \leq \|\pi\|_2 + \|\sigma\|_2 \leq 2^{-\frac{n}{2}} + 2^{-\frac{k}{2}} \leq 2^{-\frac{k}{2} + 1}
\]

\[
\|\pi\|_2 = \left( \sum_{x \in \{0,1\}^n} \frac{1}{2^n} \right)^{1/2} = 2^{-\frac{n}{2}}
\]

\[
\|\sigma\|_2 \leq 2^{-\frac{k}{2}} \quad \text{Warmup!}
\]
4. Ext is a $(k, \epsilon)$-extractor

\[ \|U_n - Y\|_{TV} \leq \epsilon? \]

We know:

• \( \|U_n - Y\|_{TV} \leq \frac{\sqrt{N}}{2} \cdot \lambda(G)^\ell \cdot \|\pi - \sigma\|_2 \)

• \( \|\pi - \sigma\|_2 \leq 2 \cdot 2^{-\frac{k}{2}} \)

• \( \ell = \frac{n-k}{2} + \log\left(\frac{1}{\epsilon}\right) \)

• \( \lambda(G) \leq \frac{1}{2} \)
4. Ext is a \((k, \epsilon)\)-extractor

\[
\|U_n - Y\|_{TV} \leq \sqrt{N} \cdot \chi(G)^l \cdot 2^{-k/2} \leq 2^{\frac{n-k}{2}} \cdot \left(\frac{1}{2}\right)^l
\]

\[
N = 2^n \quad \chi(G) \leq \nu_2
\]

\[
l = \frac{n-k}{2} + \log(\nu_2)
\]

\[
= 2^{\frac{n-k}{2}} \cdot 2^{-\left(\frac{n-k}{2} + \log(\nu_2)\right)} = 2^{-\log(\nu_2)} = \epsilon
\]

- Let \(Y = Ext(X, U_d)\)
- Let \(\pi\) be the uniform distribution.
Hooray!

• So Ext is a \((k, \varepsilon)\) extractor.
• It’s a pretty good one when \(k = n - O(\log n)\), say.
  • In that case the seed length is \(O(\log\left(\frac{n}{\varepsilon}\right))\)
• Why do we care? If \(k\) is large (as above), then we can actually just exhaust over the seeds! We don’t need true randomness!
Recap

• We can use a good spectral expander to get an okay extractor.
• This extractor is pretty good when $k$ is large!