CS225: Pseudorandomness Prof. Salil Vadhan

Lecture 11: Extractors vs. Hashing and Expanders.

March 15, 2007

Based on scribe notes by John Provine.

As mentioned in the previous lecture, we mentioned that extractors have played a unifying role in the theory of pseudorandomness, through their close connections with a variety of other pseudorandom objects. In this lecture, we will see two of these connections. Specifically, how by reinterpreting them appropriately, extractors can be viewed as providing families of hash functions, and as being a certain type of highly expanding graphs.

1 Extractors as Hash Functions

One of the results we saw last time says that for any subset $S \subseteq [N]$ of size K, if we choose a completely random hash function $h : [N] \to [M]$ for $M \ll K$, then h will map the elements of S almost-uniformly to $[M]$. Equivalently, if we let H be distributed uniformly over all functions $h : [N] \to [M]$ and X be uniform on the set S, then $(H, H(X))$ is statistically close to $(H, U_{[M]})$. Can we use a smaller family of hash functions than the set of all functions $h : [N] \to [M]$? This gives rise to the following variant of extractors.

Definition 1 (strong extractors) *Extractor* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a* strong (k, ε)-extractor *if for every k-source X on $\{0,1\}^n$, $(U_d, \mathrm{Ext}(X, U_d))$ is ε-close to (U_d, U_m). Equivalently,* $\mathrm{Ext}'(x, y) = (y, \mathrm{Ext}(x, y))$ *is a standard (k, ε)-extractor.*

The nonconstructive existence proof from last time can be extended to establish the existence of very good strong extractors.

Theorem 2 *For every $n, k \in \mathbb{N}$ and $\varepsilon > 0$ there exists a strong (k, ε)-extractor* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *with $m = k - 2\log(\frac{1}{\varepsilon}) - O(1)$ and $d = \log(n - k) + 2\log(\frac{1}{\varepsilon}) + O(1)$.*

Note that the output length $m \approx k$ instead of $m \approx k + d$; intuitively a strong extractor needs to extract randomness that is *independent* of the seed and thus can only get the k bits from the source.

We see that strong extractors can be viewed as very small families hash functions having the almost-uniform mapping property mentioned above. Indeed, our first explicit construction of extractors is obtained by using pairwise independent hash functions.

The Leftover Hash Lemma shows us how to explicitly construct an extractor from a family of pairwise independent functions $\mathcal{H}$. The extractor uses a random hash function $h \overset{\mathrm{R}}{\leftarrow} \mathcal{H}$ as its seed and keeps this seed in the output of the extractor. Thus, the extractor is strong.[1]

[1] Recall that a (k, ε)-extractor Ext is *strong* if $\mathrm{Ext}(x, y) \overset{\mathrm{def}}{=} y \circ \mathrm{Ext}'(x, y)$ for some function Ext'.

Theorem 3 (Leftover Hash Lemma) *If $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^m\}$ is a pairwise independent family where $m = k - 2\log(\frac{1}{\varepsilon})$, then $\mathrm{Ext}(x,h) \stackrel{\mathrm{def}}{=} h(x)$ is a strong (k,ε)-extractor.*

Note that the seed length is $d = O(n)$, i.e., the number of random bits required to choose $h \overset{\mathrm{R}}{\leftarrow} \mathcal{H}$. This is far from optimal; for the purposes of simulating randomized algorithms we would like $d = O(\log n)$. However, the output length of the extractor is $m = k - 2\log(\frac{1}{\varepsilon})$, which is optimal up to an additive constant.

Proof: Let X be an arbitrary k-source on $\{0,1\}^n$, $\mathcal{H}$ as above, and $H \overset{\mathrm{R}}{\leftarrow} \mathcal{H}$. Let d be the the seed length. We show that $(H, H(X))$ is ε-close to $U_d \times U_m$ in the following three steps:

1. We show that the collision probability of $(H, H(X))$ is close to that of $U_d \times U_m$.
2. We note that this is equivalent to saying that the ℓ_2 distance between $(H, H(X))$ and $U_d \times U_m$ is small.
3. Then we deduce that the statistical difference is small, by recalling that the statistical difference equals half of the ℓ_1 distance, which can be (loosely) bounded by the ℓ_2 distance.

Proof of 1: By definition, $\mathrm{CP}(H, H(X)) = \Pr[(H, H(X)) = (H', H'(X'))]$, where (H', X') is independent of and identically distributed to (H, X). Note that $(H, H(X)) = (H', H'(X))$ if and only if $H = H'$ and either $X = X'$ or $X \neq X'$ but $H(X) = H(X')$. Thus

$$\begin{aligned}\mathrm{CP}(H, H(X)) &= \mathrm{CP}(H)\left(\mathrm{CP}(X) + \Pr\left[H(X) = H(X') \mid X \neq X'\right]\right) \\ &\leq \frac{1}{D}\left(\frac{1}{K} + \frac{1}{M}\right) = \frac{1+\varepsilon^2}{DM}.\end{aligned}$$

To see the penultimate inequality, note that $\mathrm{CP}(H) = 1/D$ because there are D hash functions, $\mathrm{CP}(X) \leq 1/K$ because $\mathrm{H}_\infty(X) \geq k$, and $\Pr[H(X) = H(X') \,|X \neq X'] = 1/M$ by pairwise independence.

Proof of 2:

$$\begin{aligned}\|(H, H(X)) - U_{[D]} \times U_{[M]}\|^2 &= \mathrm{CP}(H, H(X)) - \mathrm{CP}(U_d \times U_m) \\ &\leq \frac{1+\varepsilon^2}{DM} - \frac{1}{DM} = \frac{\varepsilon^2}{DM}.\end{aligned}$$

Proof of 3: Recalling that the statistical difference between two random variables X and Y is equal to $\frac{1}{2}|X - Y|_1$, we have:

$$\begin{aligned}\Delta((H, H(X), U_d \times U_m) &= \frac{1}{2}\left|(H, H(X)) - U_d \times U_m\right|_1 \\ &\leq \frac{\sqrt{DM}}{2}\|(H, H(X)) - U_d \times U_m\| \\ &\leq \frac{\sqrt{DM}}{2} \cdot \sqrt{\frac{\varepsilon^2}{DM}} \\ &= \frac{\varepsilon}{2}.\end{aligned}$$

Thus, we have in fact obtained a strong $(k, \frac{\varepsilon}{2})$-extractor. ■

The proof above actually shows that $\mathrm{Ext}(x,h) = h(x)$ extracts with respect to CP, or equivalently, with respect to the ℓ_2-norm. This property may be expressed in terms of Renyi entropy $\mathrm{H}_2(Z) \overset{\mathrm{def}}{=} \log(1/\mathrm{CP}(Z))$. Indeed, we can define $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^m$ to be a (k,ε) *Renyi-entropy extractor* if $\mathrm{H}_2(X) \geq k$ implies $\mathrm{H}_2(\mathrm{Ext}(X,U_d)) \geq m - \varepsilon$ (or $\mathrm{H}_2(U_d, \mathrm{Ext}(X,U_d)) \geq m + d - \varepsilon$ for strong Renyi-entropy extractors). Then the above proof shows that pairwise-independent hash functions yield strong Renyi-entropy extractors.

In general, it turns out that an extractor with respect to Renyi entropy must have seed length $d \geq \Omega(\min\{m, n-k\})$ (as opposed to $d = O(\log n)$); this explains why the seed length in the above extractor is large.

2 Extractors vs. Expanders

Extractors have a natural interpretation as graphs. Specifically, we can interpret an extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^m$ as a bipartite multigraph $G = ([N],[M],E)$, where $(u,v) \in E$ if and only if $\mathrm{Ext}(u,r) = v$ for some $r \in \{0,1\}^d$. Typically $n \gg m$, so the graph is unbalanced. Note that G is D-regular on the left; indeed, this is why d is used to denote the seed length of an extractor. It turns out that the extraction property of Ext is related to various "expansion" properties of G. In this section, we explore this relationship.

2.1 Dispersers and Vertex Expansion

Let $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^m$ be a (k,ε)-extractor and $G = ([N],[M],E)$ the associated graph. Recall that it suffices to examine Ext with respect to *flat* k-sources: in this case, the extractor property says that given a subset S of size k on the left, a random neighbor of a random element of S should be close to uniform on the right. In particular, if $S \subseteq [N]$ is a subset on the left of size k, then $|N(S)| \geq (1-\varepsilon)M$. This property is just like vertex expansion, except that it ensures expansion only for sets of size exactly K, not any size $\leq K$. Indeed, this gives rise to the following weaker variant of extractors.

Definition 4 (dispersers) *A function* $\mathrm{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a* (k,ϵ)-disperser *if for every k-source X on $\{0,1\}^n$, $\mathrm{Disp}(X,U_d)$ has a support of size at least $(1-\varepsilon) \cdot 2^m$.*

While extractors can be used to simulate **BPP** algorithms with a weak random source, dispersers can be used simulate **RP** algorithms with a weak random source.

Then, we have:

Proposition 5 *A function* $\mathrm{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a (k,ϵ)-disperser iff the corresponding bipartite graph $G = ([N],[M],E)$ with left-degree D is a (K,A) vertex expander for $A = (1-\varepsilon) \cdot M/K$.*

Note that extractors and dispersers are interesting even when $M \ll K$, so the expansion parameter A may be less than 1. Indeed, $A < 1$ is interesting for vertex 'expanders' when the graph is highly imbalanced. Still, for an *optimal* extractor, we have $M = \Theta(\varepsilon^2 KD)$ (because $m = k + d - 2\log(1/\varepsilon) - \Theta(1)$), which corresponds to expansion factor $A = \Theta(\varepsilon^2 D)$. (An optimal disperser actually gives $A = \Theta(D/\log(1/\varepsilon))$.) Note this is smaller than the expansion factor of $D/2$ in Ramanujan graphs and $D - O(1)$ in random graphs; the reason is that those expansion factors are for 'small' sets, whereas here we are asking for sets to expand to almost the entire right-hand side.

Now let's look for a graph-theoretic property that is *equivalent* to the extraction property. Ext is a (k, ε)-extractor iff for every set $S \subseteq [N]$ of size K,

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

where U_S denotes the uniform distribution on S. This inequality may be expressed in graph-theoretic terms as follows. For every set $T \subseteq [M]$,

$$\begin{aligned} & \left| \Pr\left[\mathrm{Ext}(U_S, U_{[D]}) \in T\right] - \Pr\left[U_{[M]} \in T\right] \right| \leq \varepsilon \\ \Leftrightarrow \quad & \left| \frac{e(S,T)}{|S|D} - \frac{|T|}{M} \right| \leq \varepsilon \\ \Leftrightarrow \quad & \left| \frac{e(S,T)}{ND} - \mu(S)\mu(T) \right| \leq \varepsilon\mu(S) \end{aligned}$$

Thus, we have:

Proposition 6 Ext *is a (k, ε)-extractor iff the corresponding bipartite graph $G = ([N], [M], E)$ with left-degree D has the property that $\left|\frac{e(S,T)}{ND} - \mu(S)\mu(T)\right| \leq \varepsilon\mu(S)$ for every $S \subseteq [N]$ of size K and every $T \subseteq [M]$.*

Note that this is very similar to the Expander Mixing Lemma, which states that if a graph G has spectral expansion λ, then for *all* sets $S, T \subseteq [N]$ we have

$$\left| \frac{e(S,T)}{ND} - \mu(T) \right| \leq \lambda\sqrt{\mu(S)\mu(T)}.$$

It follows that if $\lambda\sqrt{\mu(S)\mu(T)} \leq \varepsilon\mu(S)$ for all $S \subseteq [N]$ of size K and all $T \subseteq [N]$, then G gives rise to a (k, ε)-extractor (by turning G into a D-regular bipartite graph with N vertices on each side in the natural way). It suffices for $\lambda \leq \varepsilon \cdot \sqrt{K/N}$ for this to work.

We can use this connection to turn our explicit construction of spectral expanders into an explicit construction of extractors. To achieve $\lambda \leq \varepsilon \cdot \sqrt{K/N}$, we can take an appropriate power of a constant-degree expander. Specifically, if G_0 is a D_0-regular expander on N vertices with bounded second eigenvalue, we can consider the tth power of G_0, $G = G_0^t$, where $t = O(\log((1/\varepsilon)\sqrt{N/K})) = O(n - k + \log(1/\varepsilon))$. The degree of G is $D = D_0^t = \mathrm{poly}(1/\lambda) = \mathrm{poly}(1/\varepsilon, N/K)$. This yields the following result:

Theorem 7 *For every $n, k \in \mathbb{N}$ and $\varepsilon > 0$, there is an explicit (k, ε)-extractor* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^n$ *with $d = O(n - k + \log(\frac{1}{\varepsilon}))$.*

Note that the seed length is significantly better than in the construction from pairwise-independent hashing when k is close to n, say $k \geq n - O(\log n)$ (i.e. $K = \Omega(N/\log N)$). The output length is just n, which is much larger than the typical output length for extractors (usually $m \ll n$). Using a Ramanujan graph (rather than an arbitrary constant-degree expander), the seed length can be improved to $d = n - k + 2\log(1/\varepsilon) + O(1)$, which yields an optimal output length $n = k + d - 2\log(1/\varepsilon) - O(1)$.

Another way of proving Theorem 7 is to use the fact that a random step on an expanders decreases the ℓ_2 distance to uniform, like in the proof of the Leftover Hash Lemma. This analysis shows that we actually get a Renyi-entropy extractor; and thus explains the large seed length $d \approx n - k$.

The following table summarizes the main differences between "classic" expanders and extractors.

Expanders	**Extractors**
Measured by vertex or spectral expansion	Measured by min-entropy/statistical difference
Typically constant degree	Typically logarithmic or poly-logarithmic degree
All sets of size *at most* K expand	All sets of size *exactly* (or at least) K expand
Typically balanced	Typically unbalanced, bipartite graphs

Figure 1: Differences between "classic" expanders and extractors