U.C. Berkeley
CS294: Pseudorandomness and Combinatorial Constructions
Professor Luca Trevisan

Handout N7
September 20, 2005
Scribe: Constantinos Daskalakis

Notes for Lecture 7

1 Increasing the Stretch of Pseudorandom Generators

In the previous lecture, we proved that if f is a permutation and B is an (S, ϵ)-hardcore predicate for f then $G : x \mapsto f(x), B(x)$, which maps n bits to $n+1$ bits, is (S, ϵ) pseudorandom. We then started investigating the performance of the iterative application of G, namely $G^{(k)}$ defined as in figure 1.

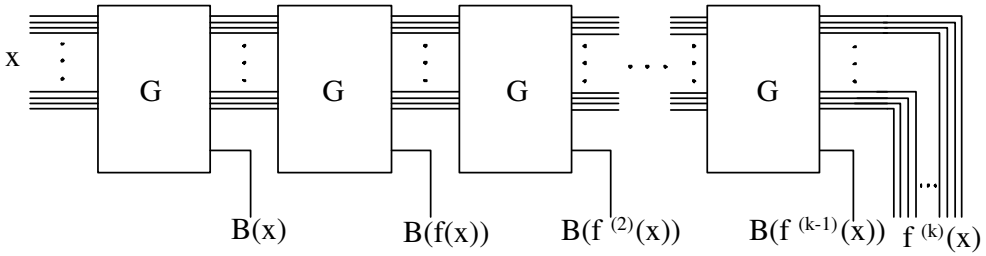

Figure 1: Definition of $G^{(k)}$

$G^{(k)}$ maps n bits to $n+k$ bits and its pseudorandom properties are determined by theorem 1 which was stated in the end of the previous lecture. Today we will finish the proof. We will go quickly through the steps of the proof that were given in the previous lecture.

Theorem 1 *If f is a permutation and B an (S, ϵ) hardcore predicate for f and, moreover, f, B are computable by circuits of size $\leq t$, then $G^{(k)}$, where $G : x \mapsto f(x), B(x)$, is $(S - O(tk), \epsilon k)$-pseudorandom.*

Proof: We will prove the contra-positive of the statement. Indeed, we will suppose that $G^{(k)}$ is not (S', ϵ')-pseudorandom and we will try to derive a distinguishing circuit for G. By definition it follows that there exists a circuit D of size $\leq S'$ such that

$$\left| \mathop{\mathbf{Pr}}_{U_n \sim \{0,1\}^n} [D(G^{(k)}(U_n)) = 1] - \mathop{\mathbf{Pr}}_{U_{n+k} \sim \{0,1\}^{n+k}} [D(U_{n+k}) = 1] \right| \geq \epsilon'$$

Without loss of generality, we can assume that circuit D satisfies the above inequality after removing the absolute value and by virtue of a hybrid argument we can argue that there exists an integer i,$0 \leq i \leq k-1$, such that:

$$\mathop{\mathbf{Pr}}_{x, r_1, \ldots, r_i} [D(r_1, \ldots, r_i, B(f^{(i)}(x)), B(f^{(i+1)}(x)), \ldots, B(f^{(k-1)}(x)), f^{(k)}(x)) = 1] -$$

$$- \mathop{\mathbf{Pr}}_{x, r_1, \ldots, r_{i+1}} [D(r_1, \ldots, r_i, r_{i+1}, B(f^{(i+1)}(x)), \ldots, B(f^{(k-1)}(x)), f^{(k)}(x)) = 1] \geq \frac{\epsilon'}{k} \qquad (1)$$

Let us denote by p_i the first term of the left hand side of the above inequality and by p_{i+1} the second term without the minus sign. Inspired by the above inequality, an algorithm that distinguishes the output of G from the uniform distribution on $n+1$ bits can look as follows.

Algorithm A
Input: An $(n+1)$-bit string b, y /*could be uniform, could be $B(z), f(z)$ where z is uniform*/
Choose uniformly at random bits $r_1, \ldots, r_i$
Output: $D(r_1, \ldots, r_i, b, B(y), B(f(y)), \ldots, B(f^{(k-i-2)}(y)), f^{(k-i-1)}(y))$

We can analyze the performance of the algorithm as follows. By a change of variables, it can be easily verified that the following equalities hold.

$$\mathop{\mathbf{Pr}}_{r_1,\ldots,r_i,b,y}[A(b,y)=1] \equiv p_{i+1} \tag{2}$$

$$\text{and} \mathop{\mathbf{Pr}}_{r_1,\ldots,r_i,z}[A(B(z),f(z))=1] \equiv p_i \tag{3}$$

Note that the only observation needed to establish these equalities is that, if z is uniformly distributed and f is a permutation, then $f(z)$ is also uniformly distributed.

From (1), (2) and (3) it follows that:

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{r_1,\ldots,r_i,z}[A(B(z),f(z))=1] - \mathop{\mathbf{Pr}}_{r_1,\ldots,r_i,b,y}[A(b,y)=1] &\geq \frac{\epsilon'}{k} \\
\Rightarrow \mathop{\mathbb{E}}_{r_1,\ldots,r_i,z,b,y}[A(B(z),f(z)) - A(b,y)] &\geq \frac{\epsilon'}{k} \\
\Rightarrow \exists r_1^*, r_2^*, \ldots, r_i^* \text{ s.t. } \mathop{\mathbb{E}}_{z,b,y}[A_{r_1^*,r_2^*,\ldots,r_i^*}(B(z),f(z)) - A_{r_1^*,r_2^*,\ldots,r_i^*}(b,y)] &\geq \frac{\epsilon'}{k} \\
\Rightarrow \exists r_1^*, r_2^*, \ldots, r_i^* \text{ s.t. } \mathop{\mathbf{Pr}}_{z,b,y}[A_{r_1^*,r_2^*,\ldots,r_i^*}(B(z),f(z))=1] - \mathop{\mathbf{Pr}}_{z,b,y}[A_{r_1^*,r_2^*,\ldots,r_i^*}(b,y)=1] &\geq \frac{\epsilon'}{k}
\end{aligned}$$

Thus algorithm $A_{r_1^*,r_2^*,\ldots,r_i^*}$ distinguishes the output of G from the uniform distribution with probability at least $\frac{\epsilon'}{k}$. Note that $A_{r_1^*,r_2^*,\ldots,r_i^*}$ can be realized by a circuit of size $S' + O(kt)$ since it just requires at most k copies of the circuits implementing f and B and one copy of circuit D. Thus, if we set $S' \leq S - O(kt)$ (the constant in the O notation same as above) and $\epsilon' \geq \epsilon k$ we get a contradiction to the fact that $G : x \mapsto f(x), B(x)$ is (S, ϵ)-pseudorandom. This concludes the proof. □

Now we will generalize theorem 1 to the following scenario: suppose $G : \{0,1\}^n \to \{0,1\}^{n+1}$ is an (S, ϵ)-pseudorandom generator of stretch $l(n) = n+1$ and suppose that we consider $G^{(k)} : \{0,1\}^n \to \{0,1\}^{n+k}$ defined as in figure 1. For convenience let's denote the output of the generator G as $G(x) =: F(x), B(x)$, where $F : \{0,1\}^n \to \{0,1\}^n$ and $B : \{0,1\}^n \to \{0,1\}$; we will prove the following theorem which is a generalization of theorem 1.

Theorem 2 *If $G : \{0,1\}^n \to \{0,1\}^{n+1}$ is an (S, ϵ) pseudorandom generator which is computable by a circuit of size $\leq t$, then $G^{(k)}$ is $(S - O(tk), \epsilon k)$-pseudorandom.*

PROOF: As in the proof of theorem 1 we will prove the contrapositive by virtue of a hybrid argument. Indeed, by way of contradiction let us suppose that $G^{(k)}$ is not (S', ϵ')-pseudorandom. We will try to derive a distinguishing circuit for G. By definition it follows that there exists a circuit D of size $\leq S'$ such that

$$\left| \mathop{\mathbf{Pr}}_{U_n \sim \{0,1\}^n} [D(G^{(k)}(U_n)) = 1] - \mathop{\mathbf{Pr}}_{U_{n+k} \sim \{0,1\}^{n+k}} [D(U_{n+k}) = 1] \right| \geq \epsilon'$$

Without loss of generality, we can assume that circuit D satisfies the above inequality after removing the absolute value namely that

$$\mathop{\mathbf{Pr}}_{U_n \sim \{0,1\}^n} [D(G^{(k)}(U_n)) = 1] - \mathop{\mathbf{Pr}}_{U_{n+k} \sim \{0,1\}^{n+k}} [D(U_{n+k}) = 1] \geq \epsilon' \tag{4}$$

Now let us consider the following hybrid distributions:

H_0, pick x :	$B(x)$,	$B(F(x))$,	$B(F^{(2)}(x))$,	$B(F^{(3)}(x))$	$\ldots$,	$B(F^{(k-1)})$,	$F^{(k)}(x)$
H_1, pick x, r_1 :	r_1,	$B(x)$,	$B(F(x))$,	$B(F^{(2)}(x))$,	$\ldots$,	$B(F^{(k-2)})$,	$F^{(k-1)}(x)$
H_2, pick x, r_1, r_2 :	r_1,	r_2,	$B(x)$,	$B(F(x))$,	$\ldots$,	$B(F^{(k-3)})$,	$F^{(k-2)}(x)$
.							
.							
.							
H_k, pick $x, r_1, r_2, \ldots, r_k$:	r_1,	r_2,	r_3,	r_4,	$\ldots$,	r_k,	x

Note that the distribution H_0 is exactly the distribution $G^{(k)}(U_n)$, observed at the output of $G^{(k)}$ if the uniform distribution U_n is given in the input, and distribution H_k is exactly the uniform distribution U_{n+k}. Thus, we can write (4) as follows:

$$\epsilon' \leq \mathbf{Pr}[D(H_0) = 1] - \mathbf{Pr}[D(H_k) = 1] = \sum_{i=0}^{k-1} [\mathbf{Pr}[D(H_i) = 1] - \mathbf{Pr}[D(H_{i+1}) = 1]]$$

from which it follows that there exists an integer $i, 0 \leq i \leq k-1$ such that:

$$\mathbf{Pr}[D(H_i) = 1] - \mathbf{Pr}[D(H_{i+1}) = 1] \geq \frac{\epsilon'}{k}$$

or

$$\mathop{\mathbf{Pr}}_{x, r_1, \ldots, r_i} [D(r_1, \ldots, r_i, B(x), B(F(x)), \ldots, B(F^{(k-i-1)}(x)), F^{(k-i)}(x)) = 1] -$$

$$- \mathop{\mathbf{Pr}}_{x, r_1, \ldots, r_{i+1}} [D(r_1, \ldots, r_i, r_{i+1}, B(x), \ldots, B(F^{(k-i-2)}(x)), F^{(k-i-1)}(x)) = 1] \geq \frac{\epsilon'}{k} \tag{5}$$

Let us denote by p_i the first term of the left hand side of the above inequality and by p_{i+1} the second term without the minus sign. Inspired by the above inequality, an algorithm that distinguishes the output of G from the uniform distribution on $n+1$ bits can look as follows.

Algorithm A
Input: An $(n+1)$-bit string b, y /*could be uniform, could be $B(z), F(z)$ where z is uniform*/
Choose uniformly at random bits $r_1, \ldots, r_i$
Output: $D(r_1, \ldots, r_i, b, B(y), B(F(y)), \ldots, B(F^{(k-i-2)}(y)), F^{(k-i-1)}(y))$

We can analyze the performance of the algorithm as follows. It can be easily verified that the following equalities hold.

$$\mathop{\mathbf{Pr}}_{r_1,\ldots,r_i,b,y}[A(b,y)=1] \equiv p_{i+1} \tag{6}$$

$$\text{and } \mathop{\mathbf{Pr}}_{r_1,\ldots,r_i,z}[A(B(z),F(z))=1] \equiv p_i \tag{7}$$

From (5), (6) and (7) it follows that:

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{r_1,\ldots,r_i,z}[A(B(z),F(z))=1] - \mathop{\mathbf{Pr}}_{r_1,\ldots,r_i,b,y}[A(b,y)=1] &\geq \frac{\epsilon'}{k} \\
\Rightarrow \mathop{\mathbb{E}}_{r_1,\ldots,r_i,z,b,y}[A(B(z),F(z)) - A(b,y)] &\geq \frac{\epsilon'}{k} \\
\Rightarrow \exists r_1^*, r_2^*, \ldots, r_i^* \text{ s.t. } \mathop{\mathbb{E}}_{z,b,y}[A_{r_1^*,r_2^*,\ldots,r_i^*}(B(z),F(z)) - A_{r_1^*,r_2^*,\ldots,r_i^*}(b,y)] &\geq \frac{\epsilon'}{k} \\
\Rightarrow \exists r_1^*, r_2^*, \ldots, r_i^* \text{ s.t. } \mathop{\mathbf{Pr}}_{z,b,y}[A_{r_1^*,r_2^*,\ldots,r_i^*}(B(z),F(z))=1] - \mathop{\mathbf{Pr}}_{z,b,y}[A_{r_1^*,r_2^*,\ldots,r_i^*}(b,y)=1] &\geq \frac{\epsilon'}{k}
\end{aligned}$$

Thus algorithm $A_{r_1^*,r_2^*,\ldots,r_i^*}$ distinguishes the output of G from the uniform distribution with probability at least $\frac{\epsilon'}{k}$. Note that $A_{r_1^*,r_2^*,\ldots,r_i^*}$ can be realized by a circuit of size $S'+O(kt)$ since it requires at most $k-1$ copies of the circuit implementing G and one copy of circuit D. Thus, if we set $S' \leq S - O(kt)$ (the constant in the O notation same as above) and $\epsilon' \geq \epsilon k$ we get a contradiction to the fact that $G : x \mapsto F(x), B(x)$ is (S,ϵ)-pseudorandom. This concludes the proof. □

Now, that we have proved theorems 1 and 2, figure 2 presents the state of our knowledge. Note that we haven't yet seen the implications denoted by broken arrows.

2 The Goldreich-Levin Theorem

The computational problem addressed by the Goldreich-Levin theorem is along the following lines:

- Suppose:
  - $l_a : \{0,1\}^n \to \{0,1\}$ is the mapping $x \mapsto \sum_i a_i x_i$, where parameter $a \in \{0,1\}^n$ is unknown
  - $f : \{0,1\}^n \to \{0,1\}$ is a boolean function such that $f(x) \equiv \sum_i a_i x_i$ for at least a fraction $\frac{1}{2} + \epsilon$ of $\{0,1\}^n$
- Given oracle access to f find a.

It turns out that the problem as stated is not well defined. Indeed, we can very well find a pair $a, b \in \{0,1\}^n$ such that the mappings:

$$x \mapsto \sum_i a_i x_i \text{ and } x \mapsto \sum_i b_i x_i$$

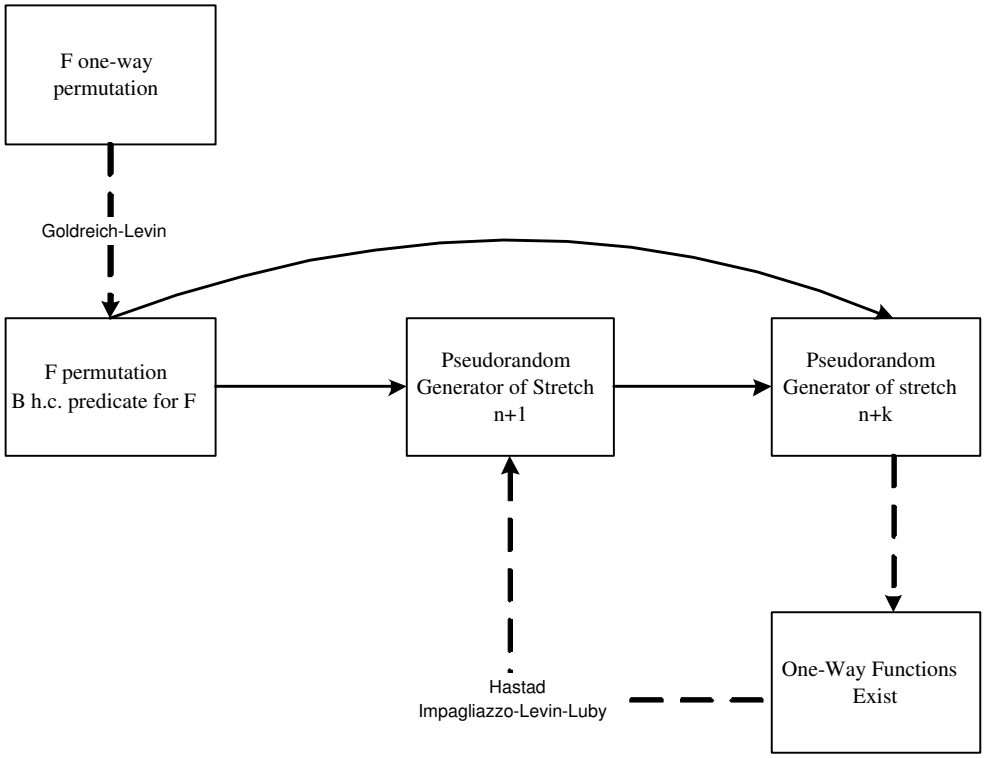

Figure 2: The picture of our knowledge.

agree on exactly half of the inputs. In this case, we can define a boolean function f which agrees with each mapping on 3/4 of the inputs and, thus, the computational problem stated above does not have a unique answer. So the computational problem is better stated as follows:

- Given oracle access to a boolean function $f : \{0,1\}^n \to \{0,1\}$ and a parameter ϵ
- Find, in time polynomial in n, $\frac{1}{\epsilon}$ and with high probability, all $a \in \{0,1\}^n$ such that f and l_a agree on at least a $\frac{1}{2} + \epsilon$ fraction of the inputs.

The elegant algorithm of Goldreich and Levin that solves this problem will be described in the next lecture.