 Handout N12 (Draft)
February 26, 2009

Notes for Lecture 12 (Draft)

Summary

Today we prove the Goldreich-Levin theorem.

1 Goldreich-Levin Theorem

We use the notation

$$\langle x, r\rangle := \sum_i x_i r_i \bmod 2 \tag{1}$$

Theorem 1 (Goldreich and Levin) *Let $f : \{0,1\}^n \to \{0,1\}^n$ be a permutation computable in time r. Suppose that A is an algorithm of complexity t such that*

$$\mathbb{P}_{x,r}[A(f(x), r) = \langle x, r\rangle] \geq \frac{1}{2} + \epsilon \tag{2}$$

Then there is an algorithm A' of complexity at most $O((t + r)\epsilon^{-2} n^{O(1)})$ such that

$$\mathbb{P}_{x}[A'(f(x)) = x] \geq \frac{\epsilon}{3}$$

Last time we proved the following partial result.

Lemma 2 *Suppose we have access to a function $H : \{0,1\}^n \to \{0,1\}$ such that, for some unknown x, we have*

$$\mathbb{P}_{r\in\{0,1\}^n}[H(r) = \langle x, r\rangle] \geq \frac{7}{8} \tag{3}$$

where $x \in \{0,1\}^n$ is an unknown string.

Then there is an algorithm that runs in time $O(n^2 \log n)$ and makes $O(n \log n)$ oracle queries into H and, with probability at least $1 - \frac{1}{n}$, outputs x.

This gave us a proof of a variant of the Goldreich-Levin Theorem in which the right-hand-side in (2) was $\frac{15}{16}$. We could tweak the proof Lemma 2 so that the right-hand-side of (4) is $\frac{3}{4}+\epsilon$, leading to proving a variant of the Goldreich-Levin Theorem in which the right-hand-side in (2) is also $\frac{3}{4}+\epsilon$.

We need, however, the full Goldreich-Levin Theorem in order to construct a pseudorandom generator, and so it seems that we have to prove a strengthening of Lemma 2 in which the right-hand-side in (4) is $\frac{1}{2}+\epsilon$.

Unfortunately such a stronger version of Lemma 2 is just false: for any two different $x, x' \in \{0,1\}^n$ we can construct an H such that

$$\mathop{\mathbb{P}}_{r \sim \{0,1\}^n}[H(r) = \langle x, r\rangle] = \frac{3}{4}$$

and

$$\mathop{\mathbb{P}}_{r \sim \{0,1\}^n}[H(r) = \langle x', r\rangle] = \frac{3}{4}$$

so no algorithm can be guaranteed to find x given an arbitrary function H such that $\mathbb{P}[H(r) = \langle x, r\rangle] = \frac{3}{4}$, because x need not be uniquely defined by H.

We can, however, prove the following:

Lemma 3 (Goldreich-Levin Algorithm) *Suppose we have access to a function $H : \{0,1\}^n \to \{0,1\}$ such that, for some unknown x, we have*

$$\mathop{\mathbb{P}}_{r \in \{0,1\}^n}[H(r) = \langle x, r\rangle] \geq \frac{1}{2} + \epsilon \tag{4}$$

where $x \in \{0,1\}^n$ is an unknown string, and $\epsilon > 0$ is given.

Then there is an algorithm GL that runs in time $O(n^2\epsilon^{-2}\log n)$ and makes $O(n \cdot \epsilon^{-2}\log n)$ oracle queries into H and, with probability at least $1 - \frac{1}{n}$, outputs a set $L \subseteq \{0,1\}^n$ such that $|L| = O(\epsilon^{-2})$ and $x \in L$.

The Goldreich-Levin Theorem is an easy consequence of Lemma 3. The Goldreich-Levin algorithm GL has other interpretations (an algorithm that learns the Fourier coefficients of H, an algorithm that decodes the Hadamard code is sub-linear time) and various applications outside cryptography.