Lecture 28: List Decoding Hadamard Code and Goldreich-Levin Hardcore Predicate

Recall

- Let $H\colon \{0,1\}^n \to \{+1,-1\}$
- Let:
  $L_\varepsilon = \{S\colon \chi_S$ agrees with H at $(1/2+\varepsilon)$ fraction of points$\}$
- Given oracle access to H output a list $L \in 2^{[n]}$ such that: For all $S \in L_\varepsilon$, we have: $\Pr[S \in L] \geqslant 1/2$. The probability here is over the internal randomness of the algorithm generating L
- This procedure is identical to *list decoding* of Hadamard Code (Hadamard code is a linear code that maps the message $S \subseteq [n]$ to $\chi_S \in \{+1,-1\}^{2^n}$)

Basic Example

- Let H be a oracle that agrees with χ_S at every $x \in \{0,1\}^n$

```
function Basic-Decode(H)
    for i from 1 to n do
        a_i = H(e_i)
    end for
    Output (a_1, ..., a_n)
end function
```

- Reconstruction of S: If $a_i = -1$ then $i \in S$; otherwise $i \notin S$

Unique Decoder

- Let H be an oracle that agrees with χ_S (for some $S \subseteq [n]$) at some $3/4 + \varepsilon$ fraction of inputs

function Unique-Decode(H)
  for i from 1 to n **do**
    for j from 1 to t **do**
      Choose $r_{i,j} \xleftarrow{\$} \{0,1\}^n$
      Let $a_{i,j} = H(r_{i,j} + e_i) \cdot H(r_{i,j})$
    end for
    Let $a_i = \mathrm{Maj}\{a_{i,1}, \ldots, a_{i,t}\}$
  end for
  Output $(a_1, \ldots, a_n)$
end function

Analysis of Unique Decoder

- Let $E_{i,j} = \mathbf{1}_{(a_{i,j}=\chi_S(e_i))}$
- Note that $\Pr[E_{i,j} = 0] \leqslant (1/4 - \varepsilon) + (1/4 - \varepsilon) = 1/2 - 2\varepsilon$ and, for each i, the $E_{i,j}$s are i.i.d. variables
- So, $a_i = \chi_S(e_i)$ except with probability $\exp(-\Theta(t/\varepsilon^2))$. Using $t = \Theta\left(\frac{1}{\varepsilon}^2 \log n\right)$ we can make this failure probability $1/n^2$
- So, $a_i = \chi_S(e_i)$ for all $i \in [n]$, except with probability $1/n$

List Decoder

- Consider any $S \in L_\varepsilon$
- Given H that agrees with χ_S at $1/2 + \varepsilon$ fraction of inputs, we want to *mimic* another *more precise* oracle $\widetilde{H}$ that agrees with $7/8$ fraction of inputs
- And we will successfully mimic $\widetilde{H}$ with probability at least $3/4$
- So, given access to $\widetilde{H}$, the unique decoder can recover S, except with probability $1/n$
- So, we recover S with probability $3/4 - 1/n \geqslant 1/2$

- Consider $S \in L_\varepsilon$

function Mimic-Hypothetical(H)
  for $i \in [\alpha]$ **do**
    Sample $x_i \xleftarrow{\$} \{0,1\}^n$
    Assume that we have *magically obtained* $b_i = \chi_S(x_i)$
  end for
  Define the following oracle $\widetilde{H}$:
  function $\widetilde{H}(H, z)$
    for $j \in [\alpha]$ **do**
      $a_i = H(z + x_i) \cdot b_i$
    end for
    Return $a = \text{Maj}\{a_1, \ldots, a_\alpha\}$
  end function
end function

Analysis of Mimicking

- Note that $a_i = \chi_S(a)$ with probability $1/2 + \varepsilon$
- Since a_is are i.i.d. we have that $a = \chi_S(z)$, except with probability $\exp(-\Theta(\alpha/\varepsilon^2))$
- So, choosing $\alpha = O(1/\varepsilon^2)$ we can achieve the correctness probability to be $31/32$
- Formally:

$$\Pr_{z, x_1, \dots, x_\alpha}[a = \chi_S(z)] \geqslant 31/32$$

- Using averaging-argument:

$$\Pr_{x_1, \dots, x_\alpha}\left[\Pr_z[a = \chi_S(z)] \geqslant 7/8\right] \geqslant 3/4$$

- Summary: Over the random choices of $x_1, \dots, x_\alpha$ we succeed with probability at least $3/4$ in implementing an oracle $\widetilde{H}$ that agrees with χ_S at at least $7/8$ fraction of inputs

Partial Solution

- We enumerate all $2^{O(1/\varepsilon^2)}$ possible $b_1, \ldots, b_\alpha$ bits
- And we execute unique decoding algorithm with the corresponding $\widetilde{H}$ oracle
- Add the output of the unique decoding algorithm to the list L

- The list size is at most $2^{O(1/\varepsilon^2)}$ and for all $S \in L_\varepsilon$, with probability $\geqslant 1/2$ we have $S \in L$

- This procedure is inefficient if $1/\varepsilon$ is super-logarithmic in n

Changing the Analysis of the Mimicking Algorithm

- We do not need $\{a_1, \dots, a_\alpha\}$ to be i.i.d.
- We just need them to be *pairwise-independent*
- In this case, we can apply Chebyshev's inequality
- The probability of $a \neq \chi_S(z)$ is defined as follows: Let $X_i = a_i \cdot \chi_S(z)$

$$\Pr\left[\sum_{i\in[\alpha]} X_i \leqslant \frac{\alpha}{2}\right] \leqslant \Pr\left[\left|\sum_{i\in[\alpha]} X_i - \left(\frac{1}{2}+\varepsilon\right)\alpha\right| \leqslant \varepsilon\alpha\right]$$

$$\leqslant \frac{\mathrm{Var}\left[\sum_{i\in[\alpha]} X_i\right]}{\varepsilon^2\alpha^2} = \Theta\left(\frac{1}{\varepsilon^2\alpha}\right)$$

- Choose $\alpha = O(1/\varepsilon^2)$ we can make the success probability $\geqslant 31/32$ as earlier

Pairwise-Independent Distributions

- Let $u_1, \dots, u_\beta$ be uniform random strings from $\{0,1\}^n$
- Let $v_1, \dots, v_\beta$ be particular values of elements in $\{+1, -1\}$
- Let $\alpha = 2^\beta - 1$
- Interpret every $i \in [\alpha]$ as the characteristic vector of the subset of $[\beta]$
- Define $x_i := \bigoplus_{k \in i} u_k$, for $i \in [\alpha]$
- Define $b_i := \prod_{k \in i} v_k$, for $i \in [\alpha]$
- Note that x_i and $x_{i'}$ are pairwise independent for $i \neq i'$
- Similarly, note that a_i and $a_{i'}$ are pairwise independent for $i \neq i'$

- To perform the "mimicking algorithm" choose random $u_1, \dots, u_\beta$ and enumerate all possible $v_1, \dots, v_\beta$
- The number of possible enumerations is $2^\beta = \alpha + 1 = O(1/\varepsilon^2)$

Time Complexity

- We have $O(\frac{1}{\varepsilon^2})$ iterations for each setting of $v_1, \ldots, v_\beta$
- Each iteration of unique decoding takes $O(\frac{1}{\varepsilon^2} n \log n)$ time
- Overall time-complexity: $O(\frac{1}{\varepsilon^4} n \log n)$
- The list size is $\leqslant 2^\beta = O(\frac{1}{\varepsilon^2})$

Goldreich-Levin Hardcore Predicate

Lemma (Hardcore Lemma)

Let $f\colon \{0,1\}^n \to \{0,1\}^m$ be a one-way function. Let X and R be a uniform random strings from $\{0,1\}^n$. Then, given $(f(X), R)$ no polynomial time algorithm cannot predict $B := R \cdot X$ with $\varepsilon \geqslant 1/\mathrm{poly}(n)$ advantage.

- $B = R \cdot X$ is known as the hardcore predicate
- Proof Idea: Proof by Contradiction. Given an adversary that predicts B, we use the adversary as an oracle to recover x using the list-decoding algorithm described previously

Proof

- Suppose there exists an adversary A that, given $(f(X), R)$, can predict the random variable B with $\varepsilon = 1/\mathrm{poly}(n)$ advantage

$$\Pr_{x,r}[A(f(x), r) = r \cdot x] \geqslant (1/2 + \varepsilon)$$

- Using an averaging argument:

$$\Pr_x \left[\Pr_r [A(f(x), r) = r \cdot x] \geqslant (1/2 + \varepsilon/2) \right] \geqslant \varepsilon/2$$

  Call such an input x as a *good input*
- Conditioned on a good input x, the adversary A is an oracle that agrees with the function χ_x at $(1/2 + \varepsilon/2)$ fraction of inputs
- Using this oracle, recover x from the list L with probability $1/2$ in $\mathrm{poly}(m + n + 1/\varepsilon)$ time using Goldreich-Levin List-Decoding Algorithm
- With probability $(\varepsilon/2) \cdot (1/2)$ we successfully recover x in polynomial time and violate the one-way-ness of f