Expander Graphs in Computer Science **WS 2010/2011**

Lecture 11: List-Decodable Codes

Lecturer: He Sun

In this course we discuss list-decodable codes. In contrast to Lecture 4 in which we use expander graphs to construct codes, list-decodable codes present nice combinatorial properties and we will see how to use these codes to construct condensers and unbalanced expanders.

1 List-Decodable Codes

Definition 11.1 *For two strings $x, y \in \Sigma^n$, their (relative) Hamming distance $d_H(x, y)$ equals $\Pr_i[x_i \neq y_i]$. The agreement is defined by $\operatorname{agr}(x, y) = 1 - d_H(x, y)$.*

Throughout this lecture, we use the relative Hamming distance to evaluate the difference between two strings.

Definition 11.2 *A q-ary code is a set $\mathcal{C} \subseteq \Sigma^n$, where Σ is an alphabet of size q. Elements of $\mathcal{C}$ are called codewords. Some key parameters:*

- *n is the block length.*
- *$k = \log_2 |\mathcal{C}|$ is the message length.*

Definition 11.3 *Let $\mathrm{Enc} : \{0,1\}^k \to \Sigma^n$ be an encoding algorithm for a code $\mathcal{C}$. A δ-decoding algorithm for Enc is a function $\mathrm{Dec} : \Sigma^n \to \{0,1\}^k$ such that for every $m \in \{0,1\}^k$ and $r \in \Sigma^n$ satisfying $d_H(\mathrm{Enc}(m), r) < \delta$, we have $\mathrm{Dec}(r) = m$. If such a function Dec exists, we call the code δ-decodable.*

A (δ, L)-list-decoding algorithm for Enc is a function $\mathrm{Dec} : \Sigma^n \to \left(\{0,1\}^k\right)^L$ such that for every $m \in \{0,1\}^k$ and $r \in \Sigma^n$ satisfying $d_H(\mathrm{Enc}(m), r) < \delta$, we have $m \in \mathrm{Dec}(r)$. If such a function Dec exists, we call the code (δ, L)-list-decodable.

The main goals in constructing codes are to have infinite families of codes in which we

- Maximize the faction δ of errors correctible.
- Maximize the rate $\rho = k/n$.
- Minimize the alphabet size $q = |\Sigma|$.
- Keep the list size L relatively small.
- Have computationally efficient encoding and decoding algorithms.

Proposition 11.4 (Johnson Bound) *(1) If $\mathcal{C}$ has minimum distance $1 - \varepsilon$, then it is a $(1 - O(\sqrt{\varepsilon}), O(1/\sqrt{\varepsilon}))$-list-decodable. (2) If a binary code $\mathcal{C}$ has minimum distance $1/2 - \varepsilon$, then it is $(1/2 - O(\sqrt{\varepsilon}), O(1/\varepsilon))$-list-decodable.*

Definition 11.5 *Let $\mathcal{C}$ be a code with encoding function* $\mathrm{Enc} : \{0,1\}^k \to \Sigma^n$. *For $r \in \Sigma^n$, define* $\mathrm{LIST}(r, \varepsilon) = \{w : \mathrm{agr}(w, r) > \varepsilon\}$.

Let us look at two examples of list-decodable codes.

Definition 11.6 (Hadamard Code) *For $k \in \mathbb{N}$, the (binary) Hadamard code of message length k is the binary code of blocklength $n = 2^k$ consisting of all functions $c : \mathbb{Z}_2^k \to \mathbb{Z}_2$ that are linear (modulo 2).*

Proposition 11.7 *The Hadamard code:*

- *is explicit with respect to the encoding function that takes a message $m \in \mathbb{Z}_2^m$ to the linear function c_m defined by $c_m = \sum_i m_i x_i \bmod 2$.*
- *has minimum distance $1/2$.*
- *is $O\left(1/2 - \varepsilon, O\left(1/\varepsilon^2\right)\right)$ list decodable for every $\varepsilon > 0$.*

Proof: It suffices to prove Item (2) and Item (3). Since for any two distinct functions $c_1, c_2 : \mathbb{Z}_2^k \to \mathbb{Z}_2$, $\Pr[c_1(x) = c_2(x)] = \Pr[(c_1 - c_2)(x) = 0] = 1/2$, therefore the minimum distance is $1/2$. Item (3) follows from the Johnsen Bound. ■

Definition 11.8 (Reed-Solomon Code) *For a prime power q and $d \in \mathbb{N}$, the q-ary Reed-Solomon code of degree d is the code of blocklength $n = q$ and message length $k = (d+1)\log q$ consisting of all polynomials $p : \mathbb{F}_q \to \mathbb{F}_q$ of degree at most d.*

Proposition 11.9 *The q-ary Reed-Solomon Code of degree d:*

- *is explicit with respect to the encoding function that takes a vector of coefficients $m \in \mathbb{F}_q^{d+1}$ to the polynomial p_m defined by $p_m(x) = \sum_{i=1}^{d} m_i x^i$.*
- *has minimum distance $\delta = 1 - d/q$ and*
- *is $\left(1/2 - O(\sqrt{d/q}), O\left(\sqrt{q/d}\right)\right)$-list-decodable.*

2 List-Decoding Views of Expanders and Extractors

Given a code $\mathrm{Enc} : [N] \to [M]^D$, we define the corresponding extractor $\mathrm{Ext} : [N] \times [D] \to [D] \times [M]$ and the neighbor function of the corresponding expander $\Gamma : [N] \times [D] \to [D] \times [M]$ via the correspondence:

$$\mathrm{Ext}(x, y) = \Gamma(x, y) = (y, \mathrm{Enc}(x)_y). \tag{1}$$

Definition 11.10 *Let $\mathrm{Enc}, \mathrm{Ext}$ and Γ be the corresponding code, extractor and expander defined by Eq. (1). For a subset $T \subseteq [D] \times [M]$ and $\varepsilon \in [0, 1)$, we define*

$$\begin{aligned}
\mathrm{LIST}(T, \varepsilon) &:= \{x : \Pr[(y, \mathrm{Enc}(x)_y) \in T] > \varepsilon\} \\
&= \left\{x : \Pr_y[\mathrm{Ext}(x, y) \in T] > \varepsilon\right\} \\
&= \left\{x : \Pr_y[\Gamma(x, y) \in T] > \varepsilon\right\}.
\end{aligned}$$

We define $\mathrm{LIST}(T, 1)$ *analogously, except that replace "$> \varepsilon$" with "$= 1$".*

According to this definition, we have the following proposition.

Proposition 11.11 Enc : $[N] \to [M]^D$ *is a* $(1 - 1/M - \varepsilon, K)$*-list-decodable iff for every* $r \in [M]^D$*, we have*

$$|\mathrm{LIST}(T_r, 1/M + \varepsilon)| \leq K, \tag{2}$$

where $T_r = \{(y, r_y) | y \in [D]\}$.

Proposition 11.12 *If* Ext : $[N] \times [D] \to [M]$ *is a* (k, ε)*-extractor then for every* $T \subseteq [D] \times [M]$*, we have*

$$|\mathrm{LIST}(T, \mu(T) + \varepsilon)| < K, \tag{3}$$

where $K = 2^k$ *and* $\mu(T) = |T|/M$*. Conversely, if Eq. (3) holds for every* $T \subseteq [D] \times [M]$*, then* Ext *is a* $(k + \log(1/\varepsilon), 2\varepsilon)$*-extractor.*

Proof: ($\Rightarrow$): The proof is by contradiction. Suppose that there is a set $T \subseteq [D] \times [M]$ with the property that $|\mathrm{LIST}(T, \mu(T) + \varepsilon)| \geq K$. Let X be a random variable distributed uniformly over $\mathrm{LIST}(T, \mu(T) + \varepsilon)$. Then $\mathbf{H}_\infty(X) \geq k$. However, we have

$$\begin{aligned}
\Pr[\mathrm{Ext}(X, U_{[D]}) \in T] &= \mathop{\mathbb{E}}_{x \in_R X} \left[\Pr[\mathrm{Ext}(x, U_{[D]}) \in T]\right] \\
&> \mu(T) + \varepsilon \\
&= \Pr[U_{[D]\times[M]} \in T] + \varepsilon.
\end{aligned}$$

So $\mathrm{Ext}(X, U_{[D]})$ is ε-far from $U_{[D]\times[M]}$, which contradicts that Ext is a (k, ε)-extractor.

($\Leftarrow$): Suppose Eq. (3) holds, we show that Ext is a $(k + \log(1/\varepsilon), 2\varepsilon)$-extractor. Let X be any $(k + \log(1/\varepsilon))$-source taking values in $[N]$. We need to show that $\mathrm{Ext}(X, U_{[D]})$ is 2ε-close to $U_{[M]}$, i. e. for every $T \subseteq [M]$, it holds that $\Pr[\mathrm{Ext}(X, U_{[D]}) \in T] < \mu(T) + 2\varepsilon$. Let T be any subset of $[M]$. Then

$$\begin{aligned}
&\Pr[\mathrm{Ext}(X, U_{[D]}) \in T] \\
\leq &\Pr\left[X \in \mathrm{LIST}(T, \mu(T) + \varepsilon)\right] + \Pr[\mathrm{Ext}(X, U_{[D]}) \in T | X \notin \mathrm{LIST}(T, \mu(T) + \varepsilon)] \\
\leq &|\mathrm{LIST}(T, \mu(T) + \varepsilon)| \cdot 2^{-(k + \log(1/\varepsilon))} + (\mu(T) + \varepsilon) \\
\leq &K \cdot 2^{-(k + \log(1/\varepsilon))} + \mu(T) + \varepsilon \\
= &\mu(T) + 2\varepsilon.
\end{aligned}$$

■

Corollary 11.13 *If* Ext : $[N] \times [D] \to [D] \times [M]$ *is a* (k, ε)*-extractor, then corresponding code* Enc *is* $(1 - 1/M - \varepsilon, K)$*-list-decodable.*

Proposition 11.14 *If* Enc : $[N] \to [M]^D$ *is* $(1 - 1/M - \varepsilon, K)$*-list-decodable, then the corresponding function* Ext : $[N] \times [D] \to [D] \times [M]$ *given by* $\mathrm{Ext}(x, y) = (y, \mathrm{Enc}(x)_y)$ *is a* $(k + \log(1/\varepsilon), M \cdot \varepsilon)$*-extractor.*

Proof: Let X be a $(k + \log(1/\varepsilon))$-source and $Y = U_{[D]}$. Then the statistical difference between $\mathrm{Ext}(X, Y)$ and $Y \times U_{[M]}$ equals

$$\begin{aligned}
\Delta(\mathrm{Ext}(X, Y), Y \times U_{[M]}) &= \mathop{\mathbb{E}}_{y \in_R Y} \left[\Delta(\mathrm{Enc}(X)_y, U_{[M]})\right] \\
&\leq \frac{M}{2} \cdot \mathop{\mathbb{E}}_{y \in_R Y} \left[\max_z \Pr[\mathrm{Enc}(X)_y = z] - 1/M\right]
\end{aligned}$$

So if we define $r \in [M]^D$ by setting r_y to be the value z maximizing $\Pr[\mathrm{Enc}(X)_y = z] - 1/M$, we have

$$\begin{aligned}
\Delta(\mathrm{Ext}(X,Y), Y \times U_{[M]}) &\le \frac{M}{2} \cdot (\Pr[(Y, \mathrm{Enc}(X)_Y) \in T_r] - 1/M) \\
&\le \frac{M}{2} \cdot (\Pr[X \in \mathrm{LIST}(T_r, 1/M + \varepsilon)] + \varepsilon) \\
&\le \frac{M}{2} \cdot \left(2^{-(k+\log(1/\varepsilon))} \cdot K + \varepsilon\right) \\
&\le M \cdot \varepsilon.
\end{aligned}$$

■

Lemma 11.15 *For $k \in \mathbb{N}$, $\Gamma : [N] \times [D] \to [D] \times [M]$ is an $(= K, A)$ expander iff for every set $T \subseteq [D] \times [M]$, such that $|T| < KA$, we have*

$$|\mathrm{LIST}(T, 1)| < K.$$

Proof:

$$\begin{aligned}
\Gamma \text{ is not an } (= K, A) \text{ expander} &\Leftrightarrow \exists S \subseteq [N] \text{ s. t. } |S| = k \text{ and } |N(S)| \le KA \\
&\Leftrightarrow \exists S \subseteq [N] \text{ s. t. } |S| \ge k \text{ and } |N(S)| \le KA \\
&\Leftrightarrow \exists T \subseteq [D] \times [M] \text{ s. t. } |\mathrm{LIST}(T, 1)| \ge k \text{ and } |T| < KA.
\end{aligned}$$

■