Error Correcting Codes and Expanders

Lecturer: Anup B. Rao Scribe: Yan Wang

Feb. 16, 2016, Feb. 18, 2016

1 Error Correcting Codes

1.1 Introduction

Let Σ be a set (alphabet) and $q = |\Sigma|$. Often, $\Sigma = \{0, 1\}$.

Definition 1.1. A **code** C is a function from Σ^k to Σ^n for some positive integers $n \geq k$. c is called a **codeword** if c is in the range of C, and $C(\Sigma^k)$ is the set of codewords. The **distance** of a code C is $d(C) = \min_{c_1, c_2 \in C(\Sigma^k), c_1 \neq c_2} \Delta(c_1, c_2)$ where $\Delta(c_1, c_2)$ denotes the Hamming distance between c_1 and c_2. The **relative distance** of a code C is $\delta(C) = d(C)/n$. The **rate** of a code C is $\gamma(C) = \frac{\log_q |\Sigma^k|}{n} = \frac{k}{n}$.

Remark 1.2. A code C with distance $\delta(C)$ can correct $< \frac{\delta(C)}{2}$ errors.

We want both γ and δ to be high, and find an efficient way of encoding and decoding.

Example 1.3. A naive code C is repeating t times the original message, i.e.

$$\begin{aligned} C\colon \{0,1\}^k &\to \{0,1\}^{kt} \\ 0 &\mapsto (0, 0, 0, ..., 0) \\ 1 &\mapsto (1, 1, 1, ..., 1) \end{aligned}$$

Then $\delta(C) = \frac{1}{k}$ and $\gamma(C) = \frac{1}{t}$. Note that both $\delta(C)$ and $\gamma(C)$ go to 0 if t and k are large.

1.2 Reed-Solomon Codes

Let n be a positive integer and $|\Sigma| \geq n$. Let $q \geq n$ be a prime or a power of prime. We work in the finite field $\mathbb{F}_q$. For $m = (m_1, m_2, ...m_k) \in \Sigma^k$, define $P_m(x) = \sum_{i=1}^{k} m_i x^{i-1}$.

Definition 1.4. Fix distinct $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}_q$, A **Reed-Solomon code** RS is a function from Σ^k to Σ^n such that

$$\begin{aligned} RS\colon \Sigma^k &\to \Sigma^n \\ (m_1, m_2, ...m_k) &\mapsto (P_m(\alpha_1), P_m(\alpha_2), ..., P_m(\alpha_n)) \end{aligned}$$

Before proving the distance of Reed-Solomon codes, we need a fundamental theorem.

Theorem 1.5. *If P is a polynomial of degree t, then $P(x)$ has at most t distinct roots.*

Claim 1.6. $\delta(RS) = n - k + 1$.

Proof. Let $M_1, M_2 \in \Sigma^k$ be two distinct codewords of RS. $P_{M_1}(x) - P_{M_2}(x)$ has at most $k-1$ roots since it has degree at most $k-1$ and $0 \neq P_{M_1}(x) - P_{M_2}(x)$. So $RS(M_1)$ and $RS(M_2)$ can agree on at most $k-1$ places. So $\delta(RS) = n - k + 1$. □

This is optimal in the sense that the minimal distance for a code is at most $n - \log_q |\Sigma^k| + 1$.

In the following, we abuse the use of δ to be relative distance of a code, i.e. $\delta(C) =$ minimum distance of C/n. We aim to find a code C such that $\delta(C) \geq \delta_0$ and $\gamma(C) \geq \gamma_0$ for some constant $\delta_0, \gamma_0 > 0$.

Note that Reed-Solomon codes do not satisfy this. For instance, take $\delta_0 = 1/2$ and $n = 2k$. Then

$$\gamma = \frac{|\Sigma^k|}{n} = \frac{k \log q}{n} = \frac{1}{2} \log q$$

However, q also grows with n!

2 Expander codes

2.1 Expander graphs

G is a d-regular, undirected, non-bipartite graph. Let A be the adjacency matrix of G and $n = |V(G)|$. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_n$ be eigenvalues of A. Note $\lambda_1(A) = d$ and $|\lambda_i| \leq d$ for all i. G is called an ϵ-expander if $|\lambda_i| \leq \epsilon d$ for all $i \geq 2$. $\epsilon \geq \frac{2\sqrt{d-1}}{d}$. If $\epsilon = \frac{2\sqrt{d-1}}{d}$, then G is called a Ramanujan graph.

A Laplacian of G is $L_G = D - A = dI - A$. The eigenvalues of L_G are $\mu_i(L_G) = d - \lambda_i(A)$. So $\mu_1 \leq \mu_2 \leq ... \leq \mu_n$. If G is ϵ-expander, then $(1-\epsilon)d \leq \mu_i \leq (1+\epsilon)d$ for $i > 1$. In particular, $\mu_1 = 0$ and the eigenvector corresponds to it is $\mathbf{1}$.

Let K_n be the complete graph on n vertices. Let G be an ϵ-expander and $\vec{x} \in \mathbb{R}^n, \vec{x} \perp \mathbf{1}$. So

$$(1-\epsilon)d||\vec{x}||^2 \leq \vec{x}^T L_G \vec{x} \leq (1+\epsilon)d||\vec{x}||^2$$

Note $L_{K_n} = nI - \mathbf{1}\cdot\mathbf{1}^T$ and L_{K_n} has eigenvalues $0, n, n, ..., n$. In addition, $d||\vec{x}||^2 = \frac{d}{n}\vec{x}^T L_{K_n}\vec{x}$. In some sense, an expander is an approximate of complete graph. We have:

$$(1-\epsilon)\frac{d}{n}L_{K_n} \leq L_G \leq (1+\epsilon)\frac{d}{n}L_{K_n}$$

Lemma 2.1. *For any $S, T \subset V, S \cap T = \emptyset, |S| = \alpha n, |T| = \beta n$. $||E(S,T)| - d\alpha\beta n| \leq \epsilon dn\sqrt{\alpha\beta}$.*

Proof. Let x_S, x_T be indicative vectors of S, T. Let $L_G = dI - A$. We have

$$x_S^T L_G x_T = d x_S^T I x_T - x_S^T A x_T = d|S \cap T| - |E(S,T)| = -|E(S,T)|$$

$$\frac{d}{n}x_S^T L_{K_n} x_T = \frac{d}{n}x_S^T I x_T - \frac{d}{n}x_S^T(\mathbf{1}\mathbf{1}^T - I)x_T = d|S\cap T| - \frac{d}{n}|S||T| + \frac{d}{n}|S\cap T| = -\frac{d}{n}|S||T|$$

Hence,

$$||E(S,T)| - d\alpha\beta n| = |x_S^T(\frac{d}{n}L_{K_n} - L_G)x_T| \leq ||x_S||||\frac{d}{n}L_{K_n} - L_G||||x_T|| \leq \epsilon dn\sqrt{\alpha\beta}$$

□

By using $x'_S = x_S - \alpha\mathbf{1}$ and $x'_T = x_T - \beta\mathbf{1}$, we can show an improved bound.

2.2 Expander Codes [1]

Expander codes are of particular interest since they have a constant positive rate, a constant positive relative distance, and a constant alphabet size. In fact, the alphabet contains only two elements, so expander codes belong to the class of binary codes. Furthermore, expander codes can be both encoded and decoded in time proportional to the block length of the code. Expander codes are the only known asymptotically good codes which can be both encoded and decoded from a constant fraction of errors in polynomial time.

We remark that both Reed-Solomon codes and expander codes are linear codes, i.e. the set of codewords is a subspace of Σ^n.

Given some positive integer d, let G be a d-regular bipartite expander graph with partitions U and V where $|U| = |V| = n$. Fix a linear code C_0 with codeword in $\{0,1\}^d$ with distance $\delta_0 := \delta_0(C_0)$ and rate $\gamma_0 := \gamma_0(C_0) > 1/2$. We construct an expander code C based on G and C_0. (In some reference, it's called Zémor code)

For each $u \in U$ we give a labeling $e_1, e_2, ..., e_d$ of the edges incident to u, and we associate a codeword $(x_{i_1}, x_{i_2}, ..., x_{i_d})$ of C_0 to the edges incident to u, namely we put weight x_{i_j} on the edge e_j for $j = 1, ..., d$.

A bit string $x \in \{0,1\}^{nd}$ is a codeword in C if x restricted to the edges of each vertex is a codeword in C_0. Note that each edge label must satisfy constraints imposed by both its left and right endpoint.

Let us first calculate the rate of C.

Lemma 2.2. $\gamma(C) \geq 2\gamma_0 - 1$.

Proof. The number of constraints in C_0 is at most $d - \gamma_0 d$. Since we have such many constraints on each vertex of G, the number of constraints in C is at most $2n(d-\gamma_0 d)$. Hence, the number of codewords in $C \geq nd - 2n(d-\gamma_0 d) = nd(2\gamma_0 - 1)$. Therefore, $\gamma(C) \geq 2\gamma_0 - 1$. □

Theorem 2.3. *[2] In $O(n \log_{4/3} n)$ time, the expander code can correct up to $\frac{\delta_0^2 dn}{18}$ errors.*

Now, we shall prove the distance of the code C. We need a lemma for expander graph.

Lemma 2.4. *If G is an ϵ-expander, then $\forall S \subset U, \forall T \subset V$, $|E(S,T) - \frac{d}{n}|S||T|| \leq \epsilon d\sqrt{|S||T|}$.*

Corollary 2.5. *If $|S| = \sigma n, |T| = \tau n$, the average degree of $G[S \cup T] \leq \frac{2d\sigma\tau}{\sigma+\tau} + \epsilon d$.*

2.2.1 Decoding algorithm

We decode the vertices in U one by one and then decode the vertices in V one by one. Repeat this process until there are no errors on the code.

2.2.2 Minimum distance

Lemma 2.6. $\delta(C) \geq \phi \geq 2/3\delta_0^2$.

Proof. Recall that C_0 and C are both linear code. We have:

$$\delta(C) = \min_{c_1, c_2 \in C, c_1 \neq c_2} \Delta(c_1, c_2) = \min_{c_1, c_2 \in C, c_1 \neq c_2} weight(c_1 - c_2) = \min_{c \in C} weight(c)$$

Let $c \in \{0,1\}^{nd}$ and $F = \{i : c_{(i)} = 1\}$ so that $|F| = \phi dn$. Every vertex $u \in U$ that is incident on F should have at least $\delta_0 d$ edges incident on F. (by definition of minimum distance of C_0 and the fact that C_0 is a linear code)

Let $S \subset U$ be the vertices that are incident on F. Let $T \subset V$ be the vertices that are incident on F. So, the average degree of $G[S \cup T] \geq \delta_0 d$. Notice that $|S| \leq \frac{\phi dn}{\delta_0 d} = \frac{\phi n}{\delta_0}$ and $|T| \leq \frac{\phi dn}{\delta_0 d} = \frac{\phi n}{\delta_0}$. We apply Corollary 2.5 and obtain:

$$d(G[S \cup T]) \leq \frac{\phi}{\delta_0} d + \epsilon d$$

Hence, $\delta_0 d \leq d(G[S \cup T]) \leq \frac{\phi}{\delta_0} d + \epsilon d$ and $\phi \geq \delta_0(\delta_0 - \epsilon) \geq 2/3\delta_0^2$ if $\epsilon \leq \delta_0/3$. So $\delta(C) \geq \phi \geq 2/3\delta_0^2$. □

2.2.3 Efficient decoding

We show that if there are at most $\frac{\delta_0^2 nd}{18}$ the errors, then the algorithm outputs the correct codewords in $\log_{4/3}(n)$ steps.

Lemma 2.7. *Let $F \subset E$ be a subset of edges and $S \subset U$ be the vertices incident to F, $|S| \leq \frac{\delta_0 n}{9}$. Suppose $T \subset V$ is the set of vertices which are incident to at least $\frac{\delta_0 d}{2}$ edges. Then $|T| \leq \frac{3}{4}|S|$.*

Proof. By definition, $|F| \geq \frac{\delta_0 d}{2}|T|$. Let $|S| = \sigma n$ and $|T| = \tau n$. Consider the average degree d' of $G[S \cup T]$. On the one hand,

$$d' = \frac{|F|}{|S| + |T|} \geq \frac{\delta_0 d \tau}{\sigma + \tau}$$

On the other hand, by Corollary 2.5,

$$d' \leq \frac{2d\sigma\tau}{\sigma + \tau} + \epsilon d$$

So $\tau \leq \frac{\epsilon\sigma}{\delta_0 - 2\sigma - \epsilon}$. If $\epsilon \leq \delta_0/3, \sigma \leq \delta_0/9$, then $\tau \leq \frac{3}{4}\sigma$. □

First, let F be the edges that have errors initially. So $|F| \geq \frac{\delta_0^2 nd}{18}$. After the first decoding step on U, the number of error vertices in U, say S, $|S| \leq \frac{\delta_0 d}{9}$; and we can apply Lemma 2.7 to obtain a bound on the number of error vertices in V. Note that if $v \in V$ has less than $\frac{\delta_0 d}{2}$ edges in F, then one decoding step will correct them.

References

[1] Dan Spielman, Expander Codes, Available at `http://www.cs.yale.edu/homes/spielman/561/2009/lect12-09.pdf`.

[2] Gilles Zemor, On expander codes, IEEE Transactions on Information Theory, volume 47, number 2, 835–837 (2001)