Derandomization (V)

Yijia Chen

Shanghai Jiao Tong University

December 23, 2012

Error Correcting Codes

Definition

For $x, y \in \{0,1\}^m$, the *fractional Hamming distance* of x and y is

$$\Delta(x,y) := \frac{\big|\{i \in [m] \mid x_i \neq y_i\big|}{m}.$$

Let $0 \leq \delta \leq 1$. A function $E : \{0,1\}^n \to \{0,1\}^m$ is an *error-correcting code with distance* δ, if for every $x \neq y \in \{0,1\}^n$ we have

$$\Delta(E(x), E(y)) \geq \delta.$$

The set

$$\mathrm{Im}(E) := \big\{E(x) \;\big|\; x \in \{0,1\}^n\big\}$$

is the set of *codewords* of E.

Lemma

Let $\delta < 1/2$ *and* $n \in \mathbb{N}$*. Then there exists a function*

$$E : \{0,1\}^n \to \{0,1\}^{n/(1-H(\delta))}$$

that is an error-correcting code with distance δ*, where*

$$\begin{aligned} H(\delta) &:= \delta \cdot \log \frac{1}{\delta} + (1-\delta) \cdot \log \frac{1}{1-\delta} \\ &= -\delta \cdot \log \delta - (1-\delta) \cdot \log(1-\delta) \end{aligned}$$

is Shannon's binary entropy function.

A useful inequality

Lemma

Let $\delta < 1/2$ *and* $n \in \mathbb{N}$. *Then*

$$\sum_{i=0}^{\lfloor \delta \cdot n \rfloor} \binom{n}{i} \leq 2^{H(\delta) \cdot n}.$$

Proof.

$$\begin{aligned} 1 = \left(\delta + (1-\delta)\right)^n &\geq \sum_{i=0}^{\lfloor \delta \cdot n \rfloor} \binom{n}{i} \delta^i \cdot (1-\delta)^{n-i} \\ &= \sum_{i=0}^{\lfloor \delta \cdot n \rfloor} \binom{n}{i} 2^{i \cdot \log \delta + (n-i) \cdot \log(1-\delta)} \\ &\geq \sum_{i=0}^{\lfloor \delta \cdot n \rfloor} \binom{n}{i} 2^{-H(\delta) \cdot n} \end{aligned}$$

□

Proof of the Gilbert-Varshamov Bound

Let $m \in \mathbb{N}$, and a string $x \in \{0,1\}^m$. Then

$$\left|\{y \in \{0,1\}^m \mid \Delta(y,x) < \delta\}\right| \leq \sum_{i=0}^{\lfloor \delta \cdot m \rfloor} \binom{m}{i} \leq 2^{H(\delta)\cdot m}$$

Therefore, by a simple greedy algorithm, we can choose a set of codewords with pairwise distance δ of size

$$2^{(1-H(\delta))\cdot m}.$$

□

Explicit codes

We need to show an explicit function $E : \{0,1\}^n \rightarrow \{0,1\}^m$ with the following properties:

Efficient encoding: There is an $m^{O(1)}$-time algorithm to compute $E(x)$ from x.

Efficient decoding: Let $\rho < \delta/2$. Then there is a polynomial time algorithm to compute the unique x from every y with $\Delta(y, E(x)) < \rho$.

Walsh-Hadamard code

Let $x, y \in \{0,1\}^n$. We define

$$x \odot y := \sum_{i=1}^{n} x_i \cdot y_i \pmod 2.$$

Definition

The *Walsh-Hadamard code* is the function $\mathrm{WH} : \{0,1\}^n \to \{0,1\}^{2^n}$ that maps every string $x \in \{0,1\}^n$ into the string $z \in \{0,1\}^{2^n}$ satisfying

$$z_y = x \odot y$$

for every $y \in \{0,1\}^n$, where z_y denotes the yth coordinate of z, identifying $\{0,1\}^n$ with $[2^n]$ in some canonical way.

Walsh-Hadamard code (cont'd)

Lemma

The function WH is an error-correcting code of distance $1/2$.

Proof.

Let $x \neq y \in \{0,1\}^n$. We can show that

$$\left|\{w \in \{0,1\}^n \mid x \odot w \neq y \odot w\}\right| = 2^{n-1}.$$

Assume $x \odot w \neq y \odot w$, i.e., $\sum_{i=1}^{n} x_i \cdot w_i \neq \sum_{i=1}^{n} y_i \cdot w_i \pmod 2$, which is equivalent to

$$\sum_{i=1}^{n} (x_i - y_i) \cdot w_i = 1 \pmod 2.$$

As $x \neq y$, there exists a $j \in [m]$ with $x_j - y_j = 1 \pmod 2$. Therefore, for every $w \in \{0,1\}^m$, let w' be the string which only differs from w on the jth bit, then

$$\sum_{i=1}^{n} (x_i - y_i) \cdot w_i \neq \sum_{i=1}^{n} (x_i - y_i) \cdot w'_i \pmod 2.$$

□

Nonbinary alphabets

Definition

Let Σ be a finite alphabet and $x, y \in \Sigma^m$. Again we define $\Delta(x, y) := \big|\{i \in [m] \mid x_i \neq y_i\}\big|/m$.

We say that $E : \Sigma^n \to \Sigma^m$ is an *error-correcting code with distance* δ *over the alphabet* Σ if for every $x \neq y \in \Sigma^n$ we have $\Delta(x, y) \geq \delta$.

Definition

Let $\mathbb{F}$ be a (finite) field. And let $n, m \in \mathbb{N}$ with $n \leq m \leq |\mathbb{F}|$. Then the *Reed-Solomon code* from $\mathbb{F}^n \to \mathbb{F}^m$ is the function $\mathrm{RS} : \mathbb{F}^n \to \mathbb{F}^m$ that on input $a_0, \ldots, a_{n-1} \in \mathbb{F}$ outputs the string $z_0, \ldots, z_{m-1}$ where

$$z_j = \sum_{i=0}^{n-1} a_i \cdot f_j^i.$$

One natural way to understand the Reed-Solomon code, is to view the input as a polynomial of degree $n-1$ over the field $\mathbb{F}$:

$$F(x) := \sum_{i=0}^{n-1} a_i x_i^i,$$

while the output is the evaluation of $F(x)$ on the points $f_0, \ldots, f_{m-1} \in \mathbb{F}$.

Reed-Solomon code (cont'd)

Lemma

The Reed-Solomon code $RS : \mathbb{F}^n \to \mathbb{F}^m$ *has distance* $1 - (n-1)/m$.

Definition

Let $\mathbb{F}$ be a (finite) field. And let $\ell, d \in \mathbb{N}$ with $d < |\mathbb{F}|$. Then the *Reed-Muller code* with parameter $\mathbb{F}, \ell, d$ is the function

$$\mathsf{RM} : \mathbb{F}^{\binom{\ell+d}{d}} \to \mathbb{F}^{|\mathbb{F}|^{\ell}}$$

that maps every ℓ-variable polynomial P over $\mathbb{F}$ of total degree d to the values of P on all the inputs in $\mathbb{F}^{\ell}$.

That is, the input is a multivariate polynomial of the form

$$P(x_1, \ldots, x_\ell) = \sum_{i_1 + \cdots + i_\ell \leq d} c_{i_1, \ldots, i_\ell} x_1^{i_1} \cdots x_\ell^{i_\ell},$$

and the output is the evaluation of P on the every $e_1, \ldots, e_\ell \in \mathbb{F}$.

Reed-Muller code (cont'd)

Lemma

The Reed-Muller code $RM : \mathbb{F}^n \to \mathbb{F}^m$ *has distance* $1 - d/|\mathbb{F}|$.

Assume that $|\mathbb{F}|$ is a power of 2.

Definition

If RS is the Reed-Solomon code mapping $\mathbb{F}^n$ to $\mathbb{F}^m$ and WH is the Walsh-Hadamard code mapping $\{0,1\}^{\log|\mathbb{F}|}$ to $\{0,1\}^{2^{\log|\mathbb{F}|}} = \{0,1\}^{|\mathbb{F}|}$, then the code WH $\circ$ RS maps $\{0,1\}^{n\log|\mathbb{F}|}$ to $\{0,1\}^{m|\mathbb{F}|}$ in the following way:

1. View RS as a code from $\{0,1\}^{n\log|\mathbb{F}|}$ to $\mathbb{F}^m$ and WH as a code from $\mathbb{F}$ to $\{0,1\}^{|\mathbb{F}|}$ using *the canonical representation of elements in $\mathbb{F}$ as strings in $\{0,1\}^{\log|\mathbb{F}|}$*.
2. For every input $x \in \{0,1\}^{n\log|\mathbb{F}|}$

$$\mathrm{WH}\cdot\mathrm{RS}(x) := \mathrm{WH}(RS(x)_1),\ldots,\mathrm{WH}(\mathrm{RS}(x)_m),$$

   where $\mathrm{RS}(x)_i$ denotes the i-th symbol of $\mathrm{RS}(x)$ (which is an element of $\mathbb{F}$ identified by a string in $\{0,1\}^{\log|\mathbb{F}|}$.

Concatenated codes – WH $\circ$ RS (cont'd)

Theorem

Let $\delta_1 := 1 - (n-1)/m$ be the distance of RS and $\delta_2 = 1/2$ the distance of WH. Then WH $\circ$ RS is an error correcting code of distance $\delta_1 \cdot \delta_2$.

Concatenated codes – the general case

Theorem

Let $E_1 : \{0,1\}^n \to \Sigma^m$ *and* $E_2 : \Sigma \to \{0,1\}^k$ *be two error correcting codes with distance* δ_1 *and* δ_2*, respectively. Then* $E_2 \circ E_1 : \{0,1\}^n \to \{0,1\}^{m \cdot k}$ *is an error correcting code of distance* $\delta_1 \cdot \delta_2$*.*

Thank you