Reed Muller Codes Achieve Capacity on Erasure Channels[1]

Vishvajeet N

Indian Institute of Technology Madras

October 5, 2015

[1] Summer Project with Prof. Prahladh Harsha and Prof. Vinod Prabhakaran, TIFR, Mumbai.

The basic problem

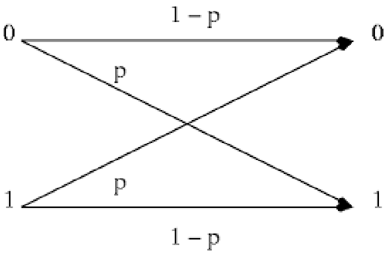
- Goal : Recover the correct codeword from a noisy received codeword
- Errors?

Communication Channel

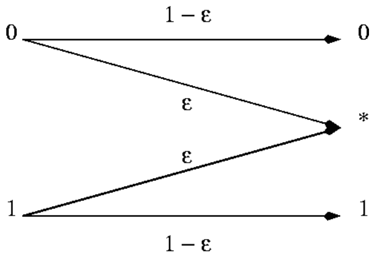
Abstraction of a physical transmission medium

- Input alphabet
- Output alphabet
- Conditional probability distribution $P(Y|X)$

Binary Symmetric Channel



Binary Erasure Channel



Symmetric, why?

A channel for which there exists a permutation π of the output alphabet Υ such that :

- $\pi^{-1} = \pi$
- $W(y|1) = W(\pi(y)|0) \quad \forall y \in \Upsilon$

We shall talk about BDMCs from here on.

Information theoretic parameters

Entropy of a discrete random variable X is defined to be

$$H(X) = -\sum_{x \in \chi} p(x) \log p(x)$$

Information theoretic parameters

Mutual information between 2 discrete RVs X and Y is defined to be

$$I(X;Y) = \sum_{x \in \chi} \sum_{y \in \Upsilon} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = H(X) - H(X|Y)$$

A Coding Scheme

- Encoding
- Decoding

Parameters of a code

Say, for every k bits of information the code generates n bits.

- Rate of a code (R) $= \frac{k}{n}$
- Distance of code (d) $= \min_{C_1 \neq C_2 \in \mathscr{C}} \triangle(C_1, C_2)$

Capacity of a channel

$$C = \text{Max}_{p_X(x)} I(X;Y)$$

Shannon's random coding approach

Given a noisy channel with channel capacity C and information transmitted at a rate R, then if R $<$ C there exist codes that allow the probability of error at the receiver to be made arbitrarily small. The converse is also true. For R $>$ C, an arbitrarily small probability of error is not achievable.

Shannon's random coding approach

Given a noisy channel with channel capacity C and information transmitted at a rate R, then if $R < C$ there exist codes that allow the probability of error at the receiver to be made arbitrarily small. The converse is also true. For $R > C$, an arbitrarily small probability of error is not achievable.

$$RS_{F,S,n,k}(m) = f(\alpha_1), f(\alpha_2), f(\alpha_3), \ldots, f(\alpha_n)),$$

where $f(X) = m_0 + m_1 X + \cdots + m_k X^k$

Vandermonde Matrix

$$\begin{bmatrix} x_1 & x_2 & x_3 & \dots & \dots & x_n \end{bmatrix} =$$

$$\begin{bmatrix} m_1 & m_2 & m_3 & \dots & \dots & m_n \end{bmatrix} * \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \dots & \alpha_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{k-1} & \alpha_2^{k-1} & \alpha_3^{k-1} & \dots & \alpha_n^{k-1} \end{bmatrix}$$

$$x_1^N = m_1^N G$$

Reed Muller Codes

Given a field size q and a number m of variables, and a total degree bound r, the RM$_{q,m,r}$ code is the code over F_q defined by the encoding map
$f(X_1, X_2, \ldots X_m) \longrightarrow < f(\alpha) > |_{\alpha \in F_q^m}$

Transitivity of Codes

1-transitivity

A code $\mathscr{C}$ is said to be 1-transitive if for any j_1 and $j_2 \in$ [N] satisfying $j_1 \neq j_2$, there exists a permutation π : [N] $\rightarrow$ [N] such that :

- $\pi(j_1) = j_2$
- $y_{\pi(1)}, y_{\pi(2)}..y_{\pi(n)} \in \mathscr{C}$ for every $y_1, y_2..y_n \in \mathscr{C}$

2-transitivity

A code $\mathscr{C}$ is said to be 2-transitive if for any j_1, j_2,j_3, $j_4 \in$ [N] satisfying $j_1 \neq j_2$ and $j_3 \neq j_4$, there exists a permutation π : [N] $\rightarrow$ [N] such that :

- $\pi(j_1) = j_3$
- $\pi(j_2) = j_4$
- $y_{\pi(1)}, y_{\pi(2)}..y_{\pi(n)} \in \mathscr{C}$ for every $y_1, y_2..y_n \in \mathscr{C}$

RS codes are 2-transitive

Proof :

RS codes are generated by the Vandermonde matrix G.

$x_1^N = m_1^N G$

Pick any permutation satisfying the constraints.

Given : $x_1^N = m_1^N * G$

$\pi x_1^N = \pi m_1^N * G$

$x_{\pi(1)}^{\pi N} = m_{\pi(1)}^{\pi N} * G$

Thus, $x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)} \ldots x_{\pi(n)}$ is a codeword too.

RS codes are 2-transitive

Proof :
RS codes are generated by the Vandermonde matrix G.
$x_1^N = m_1^N G$
Pick any permutation satisfying the constraints.
Given : $x_1^N = m_1^N * G$

$$\pi x_1^N = \pi m_1^N * G$$

$$x_{\pi(1)}^{\pi N} = m_{\pi(1)}^{\pi N} * G$$

Thus, $x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)} \dots x_{\pi(n)}$ is a codeword too.

The channel model

- The i-th bit is transmitted through a binary erasure channel with erasure probability p. Denote this channel by BEC($\underline{p}$).
- Uniform codeword assumption : A uniform distribution exists on space of codewords.

ML/MAP decoding

$$\hat{x}_{ML}\ (\mathrm{y}) = \arg\max_{x}\ f(y|x)$$

$$\hat{x}_{MAP}\ (\mathrm{y}) = \arg\max_{x}\ f(y|x)g(x)$$

The vector extrinsic information transfer function associated with the ith bit is defined to be

$$h_i(\underline{p}) \triangleq H(X_i | \underline{Y}_{\sim i}(\underline{p}_{\sim i}))$$

Observe that $\Pr\left(\hat{X}_i^{MAP}(\underline{Y}) \neq X_i \right) = H(X_i | \underline{Y})$

MAP EXIT functions

The vector extrinsic information transfer function associated with the ith bit is defined to be

$$h_i(\underline{p}) \triangleq H(X_i|\underline{Y}_{\sim i}(\underline{p}_{\sim i})$$

Observe that Pr ($\hat{X}_i^{MAP}(\underline{Y}) \neq X_i) = H(X_i|\underline{Y})$

The area theorem

The average EXIT function h satisfies the following :

$$\int_0^1 h(p)dp = \frac{K}{N}.$$

A characterization for the 'bad' patterns

Ω_i is the set of $w \in \{0,1\}^{N-1}$ for which $\exists c \in \mathscr{C}$ such that $c_i = 1$ and $c_{\sim i} \preceq w$

The 'bad' set Ω_i

- Ω_i is monotone
- $\mu_{\underline{p}}(\Omega_i) = h_i(\underline{p}) = \sum_{A \in \Omega_i} \prod_{l \in A} p_l \prod_{l \in A^c \setminus \{i\}} (1 - p_l)$
- Ω_i is 1-transitive if $\mathscr{C}$ is 2-transitive.

$\implies$ all EXIT functions are equal.

The 'bad' set Ω_i

- Ω_i is monotone
- $\mu_{\underline{p}}(\Omega_i) = h_i(\underline{p}) = \sum_{A \in \Omega_i} \prod_{l \in A} p_l \prod_{l \in A^c \setminus \{i\}} (1 - p_l)$
- Ω_i is 1-transitive if $\mathscr{C}$ is 2-transitive.

$\implies$ all EXIT functions are equal.

The 'bad' set Ω_i

- Ω_i is monotone
- $\mu_{\underline{p}}(\Omega_i) = h_i(\underline{p}) = \sum_{A \in \Omega_i} \prod_{l \in A} p_l \prod_{l \in A^c \setminus \{i\}} (1 - p_l)$
- Ω_i is 1-transitive if $\mathscr{C}$ is 2-transitive.

$\implies$ all EXIT functions are equal.

The 'bad' set Ω_i

- Ω_i is monotone
- $\mu_{\underline{p}}(\Omega_i) = h_i(\underline{p}) = \sum_{A \in \Omega_i} \prod_{l \in A} p_l \prod_{l \in A^c \setminus \{i\}} (1 - p_l)$
- Ω_i is 1-transitive if $\mathscr{C}$ is 2-transitive.

$\implies$ all EXIT functions are equal.

Inching towards the goal, are we?

Let's quickly look at what we know until now.

- $h_i(p)$ captures the probability of error of MAP decoder.
- Ω_i encodes $h_i(p)$.
- All EXIT functions are equal to the average exist function h.
- The area under the h vs p curve is the rate.

Sharp thresholds on graphs

"Every monotone graph property has a sharp threshold"

The pivotal patterns

$$\partial_j \Omega \triangleq \{\underline{x} \in \{0,1\}^N | \mathbb{1}_{\Omega}(\underline{x}) \neq \mathbb{1}_{\Omega}(\underline{x}^{(j)})\},$$

$$\text{where } \underline{x}^{(j)} = \begin{cases} x_l^{(j)} = x_l & \text{for } l \neq j \\ x_j^{(j)} = 1 - x_j & \text{otherwise} \end{cases}$$

Influences of the variables

The influence of bit j $\in$ [N] is defined by $I_j^{(p)}(\Omega) \triangleq \mu_p(\partial_j \Omega)$

Total influence $I^{(p)}(\Omega) = \sum_{l=1}^{N} I_l^{(p)}(\Omega)$

Influences of the variables

The influence of bit j $\in$ [N] is defined by $I_j^{(p)}(\Omega) \triangleq \mu_p(\partial_j \Omega)$

Total influence $I^{(p)}(\Omega) = \sum_{l=1}^{N} I_l^{(p)}(\Omega)$

Russo-Margulis lemma

If Ω be a monotone set, then $\frac{d_{\mu_p(\Omega)}}{dp} = I^{(p)}(\Omega)$

All influences are equal

There exists a bijection between $(\partial_j \Omega_i)$ and $(\partial_k \Omega_i)$ for distinct i, j ,k $\in$ [N]

Modified Russo-Margulis

Let Ω be a montone set and suppose that, for all $0 \leq p \geq 1$, the influences of all bits are equal $I_1^{(p)}(\Omega) = I_2^{(p)}(\Omega) = \cdots = I_N^{(p)}(\Omega)$. The following is true :

1. There exists an universal constant $C \geq 1$ which is independent of p, Ω and N, such that $\frac{d_{\mu_p(\Omega)}}{dp} \geq C(\log N)(\mu_p(\Omega))(1 - \mu_p(\Omega))$
2. For any $0 < \epsilon \leq \frac{1}{2}$,
   $p_{1-\epsilon} - \ p_\epsilon \leq \frac{2}{C} \frac{\log \frac{1-\epsilon}{\epsilon}}{\log M}$
   where $p_t \triangleq h^{-1} = inf\{p \in [0,1] | h(p) \geq t\}$ is the inverse function for the average EXIT function[2].

[2] h(p) is a strictly increasing continuous polynomial function and hence inverse is well-defined on [0,1]

Dictator functions

$f(x_1, x_2, x_3, ...x_n) = f(x_i)$ for some $i \in [N]$.

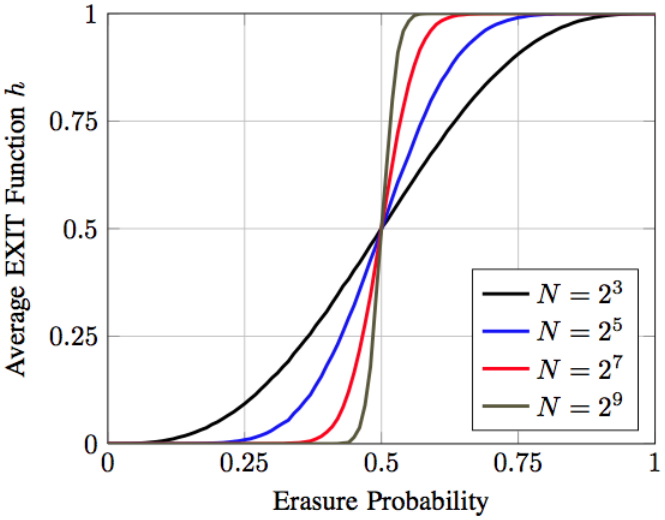
Dictators do not show the threshold property.

Dictator functions

$f(x_1, x_2, x_3, ...x_n) = f(x_i)$ for some $i \in [N]$.

Dictators do not show the threshold property.

Reed Muller and polar codes achieve capacity on BECs



Further Reading

- Shrinivas Kudekar, Marco Mondelli, Eren Sasoglu, Rudiger Urbanke *"Reed-Muller Codes Achieve Capacity on the Binary Erasure Channel under MAP Decoding."*, arXiv:1505.0583, 2015.
- Santhosh Kumar and Henry D. Pfister, *"Reed-Muller Codes Achieve Capacity on Erasure Channels"*,arXiv:1505.05123, 2015.
- Venkatesan Guruswami, Atri Rudra, *"Error-correction up to the information-theoretic limit"* Communications of the ACM, Volume 52 Issue 3, March 2009.
- C. E. Shannon, *" A mathematical theory of communication"* ACM SIGMOBILE Mobile Computing and Communications Review, Volume 5 Issue 1, January 2001.

Thank You.

Proposition: (i) A generator matrix of $\mathcal{R}(1,1)$ is

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

(ii) If G_m is a generator matrix for $\mathcal{R}(1,m)$, then a generator matrix for $\mathcal{R}(1,m+1)$ is

$$G_{m+1} = \begin{pmatrix} G_m & G_m \\ 0 \cdots 0 & 1 \cdots 1 \end{pmatrix}$$

$$\Omega_i \triangleq \{A \subseteq [N] \setminus i | \exists B \subseteq [N] \setminus i, B \cup \{i\} \in \mathscr{C}, B \subseteq A\}$$