List Decoding Reed-Muller Codes over $\mathbb{F}_2$

Sahil Singla (ssingla@cmu.edu)
Manzil Zaheer (manzil@cmu.edu)

Original Paper by Gopalan, Klivans and Zuckerman

December 3, 2014
Algebraic Code

Algebraic Coding Theory

Linear Block Codes
- Partition message into blocks and encode as polynomials
- 1 codeword ↔ 1 message
  - Reed-Solomon codes: Univariate polynomials
  - Reed-Muller codes: Multivariate polynomials
- List decoding

Convolutional Codes
- Message treated as series and encoded into series
- 1 codeword is weighted sum input messages
  - Turbo codes
  - Viterbi algorithm
  - Historically used commonly as easier to implement

Both posses same error correcting power!
Background

Reed-Muller Codes

Given a field size $q$, a number $m$ of variables, and a total degree bound $r$, the $RM_q[m,r]$ code is the linear code over $\mathbb{F}_q$ defined by the encoding map:

$$f(X_1, \ldots, X_m) \rightarrow \langle f(\alpha) \rangle_{\alpha \in \mathbb{F}_q^m}$$

applies to the domain of all polynomials in $\mathbb{F}_q[X_1, \ldots, X_m]$ of total degree $\deg(f) \leq r$.

For the binary case, i.e. $q = 2$

- Block length $n = 2^m$
- Dimension $k = \sum_{i=0}^{r} \binom{m}{i}$
- Distance $d = 2^{m-r}$, $\delta = d/n = 2^{-r}$

For $r = 1$ boils down to Hadamard code.
Decoding RM Codes

- **Unique Decoding:**
  - Majority Logic Circuit Decoder [Reed, 1954, Muller, 1954]
  - Works when error rate $\eta < 2^{-r-1} - \epsilon$
Decoding RM Codes

- **Unique Decoding:**
  - Majority Logic Circuit Decoder [Reed, 1954, Muller, 1954]
  - Works when error rate $\eta < 2^{-r-1} - \epsilon$

- **List Decoding for the case $r = 1$**
  - When error rate $\eta < \frac{1}{2} - \epsilon$
  - Outputs a list of size $\leq 2m/\epsilon^2$
  - In time $\text{poly}(m, 1/\epsilon)$
Decoding RM Codes

- **Unique Decoding:**
  - Majority Logic Circuit Decoder [Reed, 1954, Muller, 1954]
  - Works when error rate $\eta < 2^{-r-1} - \epsilon$

- **List Decoding for the case $r = 1$**
  - When error rate $\eta < \frac{1}{2} - \epsilon$
  - Outputs a list of size $\leq 2m/\epsilon^2$
  - In time $\text{poly}(m, 1/\epsilon)$

- **List Decoding for the case $r \geq 2$ – This talk!**
  - Built by generalizing GL as in [Gopalan et al., 2008]
  - When error rate $\eta < 2^{-r} - \epsilon$
  - Outputs a list of size $O(\epsilon^{-8r})$
  - In time $\text{poly}_r(m, 1/\epsilon)$
Beats Johnson Bound!

- Recall Johnson Bound
  - When $\eta < J(\delta) - \epsilon$, then
  - code is list decodable with list size $O(\epsilon^2)$
  - where $J(\delta) = \frac{1}{2}(1 - \sqrt{1 - 2\delta})$

- For RM codes, we have $\delta = 2^{-r}$

<table>
<thead>
<tr>
<th></th>
<th>Johnson Bound</th>
<th>GKZ List Decoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>List Size</td>
<td>$O(\epsilon^2)$</td>
<td>$O(\epsilon^2)$</td>
</tr>
<tr>
<td>Time</td>
<td>$-$</td>
<td>$\text{poly}_r(m, 1/\epsilon)$</td>
</tr>
<tr>
<td>Max Error</td>
<td>$J(2^{-r}) - \epsilon$</td>
<td>$2^{-r} - \epsilon$</td>
</tr>
<tr>
<td>Example ($r = 2$)</td>
<td>0.146</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Can we do better?

- No! as exponentially many codewords at distance of $2^{-r}$
- An example:
  - Let $\mathbf{V}_1, \ldots, \mathbf{V}_t \subset \mathbb{F}_2^m$ such that $\forall i : \dim(\mathbf{V}_i) = m - r$.
  - Each $\mathbf{V}_i$ has a parity check matrix $[H^{(i)}]_{r \times m}$
  - Consider the polynomials
    \[
    P_i(x) = \prod_{j=1}^{r} (1 + \langle H^{(i)}_j, x \rangle) = \begin{cases} 
      1 & \text{if } x \in \mathbf{V}_i \\
      0 & \text{else}
    \end{cases}
    \]
  - All $P_i$'s are unique
  - They are valid codewords in $RM(m, r)$ code!
  - If we receive $R = 0$, then all these are at distance $2^{-r}$
  - Note $t = \text{Number of subspace of dimension } m - r > 2^{r(m-r)}$
GL: Hadamard List Decoding

- Let the message be $s \in \mathbb{F}_2^m$ and define $P(x) = \langle s, x \rangle$
- Then $\text{Had}(s) = \langle P(\alpha) \rangle_{\alpha \in \mathbb{F}_2^m}$
- We receive a noisy function $R : \mathbb{F}_2^m \to \mathbb{F}_2$ such that $\Delta(P, R) \leq \eta < \frac{1}{2} - \epsilon$
- Goal: Recover the message $s$ (or equivalently $P$) from $R$

- Enumerated $R$
- Error $R(x) \neq P(x)$
- Correct $R(x) = P(x)$
GL: Hadamard List Decoding

Set $k := O(\log(m/\epsilon))$

Begin by selecting a random subspace $A$ of $\dim(A) = k$

Assume $\forall x \in A : R(x) = P(x)$

Call them “hints”
GL: Hadamard List Decoding

- Set $k := O(\log(m/\epsilon))$
- Begin by selecting a random subspace $A$ of $\dim(A) = k$
- Assume $\forall x \in A : R(x) = P(x)$
- Call them “hints”
GL: Hadamard List Decoding

- Given the hints
- For any \( b \in \mathbb{F}_2^m \)
- Consider the space \( b + A \)
- Error in \( A = 0 \) (assumed)
- Error in \( b + A < \eta + \epsilon \) (with constant probability)
GL: Hadamard List Decoding

▶ Error in $A = 0$
▶ Error in $b + A < \eta + \epsilon$
▶ Error in combined subspace $< \frac{\eta + \epsilon}{2} < \frac{1}{4}$
▶ Unique Decode!
GL: Hadamard List Decoding

- Error in $A = 0$
- Error in $b + A < \eta + \epsilon$
- Error in combined subspace $< \frac{\eta + \epsilon}{2} < \frac{1}{4}$
- Unique Decode!
GL: Hadamard List Decoding

- Error in $A = 0$
- Error in $b + A < \eta + \epsilon$
- Error in combined subspace $< \frac{\eta + \epsilon}{2} < \frac{1}{4}$
- Unique Decode!
Interpolating Sets

- Q: For how many \( b \)'s do we need to run this?
- A: As many times as it needs to uniquely determine the polynomial \( P \)
  - In case of Hadamard codes, \( P \) is linear in \( m \) variables
  - It suffices to run for \( b = e_1, ..., e_m \)

- In general, for a degree \( r \) polynomial in \( m \) variables
  - The set sufficient to efficiently determine the polynomial uniquely is called the interpolating set
  - Any Hamming ball of radius \( r \) is an interpolating set having \( O(m^r) \) points.
Summary

But we don't know the hints!
Iterate over all possible hints
\[ \# \text{ hints} = 2^k = \text{poly}(m, 1/\epsilon) \]
\[ \therefore \text{ still polynomial in list size and time} \]
Problems porting to RM

- Most of the steps for GL can be directly ported for general RM$[r, m]$ codes

- Brute forcing over guess doesn’t work any more
  - Too many choices for $r \geq 2$
  - For being able to evaluate $Q(a + b)$, we need to make $2^{O(k^r)}$ guess
Finding restriction $P_A$

- Note with high probability $\Delta(P_A, R_A) \leq \eta + \epsilon$
- Thus, find list $L$ of every degree $r$ polynomial $Q$ on $k$ dimensions s.t. $\Delta(Q, R_A) \leq \eta + \epsilon$
- Moreover, since $k = O(\log \frac{m}{\epsilon})$, we can use a global list decoding algorithm
Finding restriction $P_A$

- Note with high probability $\Delta(P_A, R_A) \leq \eta + \epsilon$
- Thus, find list $\mathcal{L}$ of every degree $r$ polynomial $Q$ on $k$ dimensions s.t. $\Delta(Q, R_A) \leq \eta + \epsilon$
- Moreover, since $k = O(\log \frac{m}{\epsilon})$, we can use a global list decoding algorithm

**Challenges:**

1. Design a global RM list decoding algorithm.
2. Argue $|\mathcal{L}|$ is $O(\epsilon^{-8r})$
Global RM list decoding

\[ \eta = \frac{1}{2} (\eta_0 + \eta_1) \]

\[ \text{Assume } \eta_0 \leq \eta_1 \]

\[ \text{Thus, } \eta_0 \leq \eta \text{ and } \eta_1 \leq 2\eta \]
Global RM list decoding

\[ \eta = \frac{1}{2} (\eta_0 + \eta_1) \]

- Assume \( \eta_0 \leq \eta_1 \)
- Thus, \( \eta_0 \leq \eta \) and \( \eta_1 \leq 2\eta \)

- Note \( Q = Q_0(X_1, \ldots, X_{k-1}) + X_k Q'(X_1, \ldots, X_{k-1}) \)
- Recurse over \( Q_0 \): \( \eta_0 \leq \eta \) and degree at most \( k \)
- Recurse over \( Q' \): \( \eta_1 \leq 2\eta \) and degree at most \( k - 1 \)

Since we don’t know if \( \eta_0 \leq \eta_1 \), try every possible \( 2^k \) orders
Reduction of $A'$s dimension

- The original algorithm has $k \geq O(\log \frac{m}{\epsilon})$.
- Instead, $k \geq O(\log \frac{1}{\epsilon})$ suffices
- First showed using clever interpolating sets, Dvir-Shpilka [Dvir and Shpilka, 2008]
- Later showed by implementing Reed’s Majority Logic Decoder locally
- Hence, $l(r, m, 2^{-r} - \epsilon) = O(l(r, k, 2^{-r}))$
- We bound $l(r, k, 2^{-r})$ by $O(\epsilon^{-8r})$
Deletion lemma

Johnson Bound

For any code $C$ with distance $\delta n$ and any $R \in \{0, 1\}^n$

- Number of $C$ such that $\Delta(R, C) < J(\delta) - \gamma$ is at most $O(\gamma^{-2})$
- Number of $C$ such that $\Delta(R, C) < J(\delta)$ is at most $2n$
Deletion lemma

**Johnson Bound**

For any code $C$ with distance $\delta n$ and any $R \in \{0, 1\}^n$

- Number of $C$ such that $\Delta(R, C) < J(\delta) - \gamma$ is at most $O(\gamma^{-2})$
- Number of $C$ such that $\Delta(R, C) < J(\delta)$ is at most $2n$

Let $A(\alpha)$ be number of codewords of weight less than $\alpha$

**Deletion lemma**

For any linear code $C$ and $\alpha \in [0, 1]$ and $R \in \{0, 1\}^n$

- Number of $C$ such that $\Delta(R, C) < J(\alpha) - \gamma$ is at most $A(\alpha)O(\gamma^{-2})$
- Number of $C$ such that $\Delta(R, C) < J(\alpha)$ is at most $2A(\alpha)n$

- Generalization of Johnson Bound for $\alpha = \delta$ and $A(\delta) = 1$
Bounding list size $|\mathcal{L}|$

Let $\alpha = 2(2^{-r} - 2^{-2r})$

**Corollary of Kasami-Tokura lemma**

$A(\alpha) \leq 2.2^{(4r-2)(k+1)}$

Recollect $l(r, m, 2^{-r} - \epsilon) = O(l(r, k, 2^{-r}))$

$l(r, k, 2^{-r}) \leq 2A(\alpha)n$, by Deletion lemma

$= 2A(\alpha)2^k$

$= O(\epsilon^{-8r})$, using above corollary
Open Problem

Conjecture
For field $\mathbb{F}_q$ and $\epsilon > 0$, $\exists c(q, \epsilon, r)$ independent of $n$ s.t. for all $m$ and $r$

$$l_q(r, m, \delta_q(r) - \epsilon) \leq c(q, \epsilon, r)$$

- GKZ also proves for small $q$ when $q - 1$ divides $r$
- Proven for quadratic polynomials $r = 2$ [Gopalan, 2010]
- List decoding over $\mathbb{F}_p$ for prime $p$
  shown [Bhowmick and Lovett, 2014]
Many of the images were adopted from David Zuckerman’s presentation!

List decoding reed-muller codes over small fields.


Noisy interpolating sets for low degree polynomials.


A hard-core predicate for all one-way functions.

In Proceedings of the twenty-first annual ACM symposium on Theory of computing, pages 25–32. ACM.
Reference II

A fourier-analytic approach to reed-muller decoding.

List-decoding reed-muller codes over small fields.
In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 265–274. ACM.

Application of boolean algebra to switching circuit design and to error detection.
Reference III

A class of multiple-error-correcting codes and the decoding scheme.