Shooting for the Stars! The May-Ozerov Algorithm for Syndrome Decoding is “Galactic”

M. Hamdad

October, 2023
Shooting for the Stars! The May-Ozerov Algorithm for Syndrome Decoding is “Galactic”

M. Hamdad

October, 2023
Algorithms for Nearest Neighbor Problem and application to cryptanalysis of McEliece cryptosystem

M. Hamdad

October, 2023
The Nearest Neighbor Problem and application to cryptanalysis

- **Boolean context**
- **Decoding random linear codes (DRLC)**: 
  Find $x$ such that $Hx = s$ with $|x| \leq w$ (NP-hard)
- **Improve DRLC**
  $\implies$ improve McEliece cryptosystem’s cryptanalysis
- The best-known algorithms use in a crucial way a subroutine that solve **NNP**
Nearest Neighbor Problem in $\mathbb{F}_2^m$

Goal: Find $C = (x, y) \in L \times R$ such that $|x + y| \leq \gamma m$
Nearest Neighbor Problem in $\mathbb{F}_2^m$

Goal: Find $C = (x, y) \in L \times R$ such that $|x + y| \leq \gamma m$

Here $\gamma m = 1 \implies \gamma = \frac{1}{m}$

<table>
<thead>
<tr>
<th>Bit n°</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Line 2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Line 5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bit n°</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Line 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Line 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Line 6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$N = 2^{\lambda m}$
The projection method

Probability that 2 bit strings we search coincide on $k$ columns: $\frac{\binom{m}{k}}{\binom{m(1-\gamma)}{k}}$

The algorithm

- Pick $k$ columns randomly
- Sort the 2 lists in lexicographic order according to the selected columns
- Compare all pairs of bit strings that coincide on the $k$ columns
- Repeat $\sim \frac{\binom{m}{k}}{\binom{m(1-\gamma)}{k}}$ times
$k = 2$, drawn column numbers = \{0, 2\}

<table>
<thead>
<tr>
<th>Bit n°</th>
<th>2</th>
<th>0</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Line 3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Line 7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$\textbf{L sorted}$

C

<table>
<thead>
<tr>
<th>Bit n°</th>
<th>2</th>
<th>0</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Line 4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Line 6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Line 7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$\textbf{R sorted}$
\[ k = 2, \text{ drawn column numbers} = \{1, 4\} \]

<table>
<thead>
<tr>
<th>Bit n°</th>
<th>4</th>
<th>1</th>
<th>0</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Line 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Line 7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bit n°</th>
<th>4</th>
<th>1</th>
<th>0</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Line 1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Line 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Line 3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line 7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Complexity

\[ C_{Proj} = O \left( \left( N + \frac{N^2}{2^k} \right) \frac{\binom{m}{k}}{\binom{m-l}{k}} \right) \]

A well known complexity tradeoff is \( k = \lambda m \) then

\[ C_{Proj} = O \left( 2^m \left( \lambda + h(\lambda, \gamma) \right) \right) \]
A well know complexity tradeoff is $k = \lambda m$ then

$$\mathcal{C}_{Proj} = O \left( 2^m (\lambda + h(\lambda, \gamma)) \right)$$

where $h(\lambda, \gamma) = H(\lambda) - (1 - \gamma) H\left( \frac{\lambda}{1 - \gamma} \right)$.
In practice

Parameters

\( n = 200, \ \gamma m = 60, \ N \in [250, 2300], \ step = 10 \)
\[ C_{Proj} = O \left( 2^{m(\lambda + h(\lambda, \gamma))} \right) \]
\[ C_{Proj} = O \left( 2^m \left( \lambda + h(\lambda, \gamma) \right) \right) \]

\[ C_{MO} = \tilde{O}(2^{y(\lambda, \gamma)m}) \]
$C_{Proj} = O \left( 2^m (\lambda + h(\lambda, \gamma)) \right)$

$C_{MO} = \tilde{O} (2^{y(\lambda, \gamma)m})$ where $y(\lambda, \gamma) = (1 - \gamma) \left( 1 - \frac{H(H^{-1}(1-\lambda)-\gamma/2)}{1-\gamma} \right)$
\[ C_{Proj} = O\left(2^m(\lambda + h(\lambda, \gamma))\right) \]

\[ C_{MO} = \tilde{O}(2^{y(\lambda, \gamma)m}) \text{ where } y(\lambda, \gamma) = (1 - \gamma) \left(1 - \frac{H(H^{-1}(1-\lambda)-\gamma/2)}{1-\gamma}\right) \]

With \( \lambda = 0.025 \) and \( \gamma = 0.1 \) \( C_{MO} \geq C_{Proj} \implies m \geq 256000 \implies |L| = |R| \geq 2^{8000} \)
At a fixed list size $N$ and a fixed $\gamma$, what happens if the vectors are twice as long?

$$2^\lambda m = 2^{\frac{\lambda}{2}} 2^m$$

Goal: Find $C = (x', y') \in L_2 \times R_2$ such that $|x' + y'| \leq \gamma 2m$
Complexity

\[
C_{\text{Proj}2} = O \left( \left( N + \frac{N^2}{2k} \right) \frac{\binom{2m}{k}}{(2m(1-\gamma))^k} \right)
\]

If we choose \( k = \lambda m \) then

\[
C_{\text{Proj}2} = O \left( 2^m(\lambda + 2h(\frac{\lambda}{2}, \gamma)) \right) \text{ with } 2h(\lambda/2, \gamma) \leq h(\lambda, \gamma)
\]
Complexity

\[ C_{Proj2} = O \left( \left( N + \frac{N^2}{2^k} \right) \left( \frac{\binom{2m}{k}}{2^m(1-\gamma)} \right) \right) \]

If we choose \( k = \lambda m \) then

\[ C_{Proj2} = O \left( 2^m(\lambda+2h(\frac{\lambda}{2}, \gamma)) \right) \text{ with } 2h(\frac{\lambda}{2}, \gamma) \leq h(\lambda, \gamma) \]

And if we concatenate \( L \) with itself, the same for \( R \)?
Complexity

\[ C_{Proj2} = O \left( \left( N + \frac{N^2}{2^k} \right) \frac{\binom{2m}{k}}{2m(1-\gamma)} \right) \]

If we choose \( k = \lambda m \) then

\[ C_{Proj2} = O \left( 2^m(\lambda+2h(\frac{\lambda}{2},\gamma)) \right) \text{ with } 2h(\lambda/2, \gamma) \leq h(\lambda, \gamma) \]

And if we concatenate \( L \) with itself, the same for \( R \)?

That won't work: some of the \( k \) columns drawn can be identical
If we choose $k = \lambda m$ then

$$C_{Proj2} = O \left( \binom{N + \frac{N^2}{2k}}{k} \frac{\binom{2m}{k}}{2m(1-\gamma)} \right)$$

And if we concatenate $L$ with itself, the same for $R$?

That won't work: some of the $k$ columns drawn can be identical

$\implies$ Filtering on less than $k$ columns
But let see the function $f(t) = t.h(\frac{\lambda}{t}, \gamma)$
But let see the function $f(t) = t \cdot h\left(\frac{\lambda}{t}, \gamma\right)$

In fact, $f(t)$ is decreasing
At a fixed list size $N$ and a fixed $\gamma$, what happens if the vectors are $k$ times longer?

**Complexity**

$$C_{Proj_k} = O\left(\left( N + \frac{N^2}{2^k} \right) \frac{\binom{km}{k}}{\binom{km(1-\gamma)}{k}} \right)$$

If we choose $k = \lambda m$ then

$$C_{Proj_k} = O\left(2^m(\lambda+k.\text{h}(\frac{\lambda}{k},\gamma))\right)$$
At a fixed list size $N$ and a fixed $\gamma$, what happens if the vectors are $k$ times longer?

### Complexity

$$C_{Projk} = O \left( \left( N + \frac{N^2}{2^k} \right) \frac{\binom{km}{k}}{\binom{km(1-\gamma)}{k}} \right)$$

If we choose $k = \lambda m$ then

$$C_{Projk} = O \left( 2^m (\lambda + k \cdot h(\frac{\lambda}{k}, \gamma)) \right)$$

Concatenate $L$ with itself $k$ times, the same for $R$?
At a fixed list size N and a fixed \( \gamma \), what happens if the vectors are \( k \) times longer?

**Complexity**

\[
C_{Projk} = O \left( \left( N + \frac{N^2}{2^k} \right) \frac{\binom{km}{k}}{\binom{km(1-\gamma)}{k}} \right)
\]

If we choose \( k = \lambda m \) then

\[
C_{Projk} = O \left( 2^m (\lambda+k \cdot h\left( \frac{\lambda}{k}, \gamma \right)) \right)
\]

Concatenate \( L \) with itself \( k \) times, the same for \( R \)?

Again columns drawn can be identical
At a fixed list size $N$ and a fixed $\gamma$, what happens if the vectors are $k$ times longer?

**Complexity**

$C_{Projk} = O \left( \left( N + \frac{N^2}{2^k} \right) \left( \frac{km}{k} \right) \left( \frac{km(1-\gamma)}{k} \right) \right)$

If we choose $k = \lambda m$ then

$C_{Projk} = O \left( 2^m(\lambda+k.h(\frac{1}{k},\gamma)) \right)$

Concatenate $L$ with itself $k$ times, the same for $R$?

Again columns drawn can be identical

It looks like drawing with replacement
The projection method drawing columns with replacement

**Complexity**

\[ C_{ProjR} = O \left( \left( N + \frac{N^2}{2^m(1-(1-\frac{1}{m})^k)} \right)(1-\gamma)^{-k} \right) \]
The projection method drawing columns with replacement

### Complexity

\[
C_{ProjR} = O\left(\left(N + \frac{N^2}{2^m(1-(1-\frac{1}{m})^k)}\right)(1-\gamma)^{-k}\right)
\]

If we choose \( k = \frac{\ln(1-\lambda)}{\ln(1-\frac{1}{m})} \) then
The projection method drawing columns with replacement

**Complexity**

\[ C_{ProjR} = O \left( \left( N + \frac{N^2}{2^m(1-(1-\frac{1}{m})^k)} \right) (1-\gamma)^{-k} \right) \]

If we choose \( k = \frac{\ln(1-\lambda)}{\ln(1-\frac{1}{m})} \) then

\[ C_{ProjR} = O \left( N(1-\gamma)^{-k} \right) = O \left( 2^m(\lambda+\log_2(1-\gamma)\ln(1-\lambda)) \right) \]
The projection method drawing columns with replacement

**Complexity**

\[
C_{ProjR} = O \left( \left( N + \frac{N^2}{2^m(1-(1-\frac{1}{m})^k)} \right) (1 - \gamma)^{-k} \right)
\]

If we choose \( k = \frac{\ln(1-\lambda)}{\ln(1-\frac{1}{m})} \) then

\[
C_{ProjR} = O \left( N(1 - \gamma)^{-k} \right) = O \left( 2^m(\lambda + \log_2(1-\gamma \ln(1-\lambda))) \right)
\]

\[
\log_2(1 - \gamma \ln(1 - \lambda)) \leq h(\lambda, \gamma)
\]
Work in progress:

- Drawing columns with replacement in practice?
- Concatenate lists seems to improve complexity of Esser, Kübler and Zweydinger algorithm

\[ 2y(\lambda/2, \gamma) \leq y(\lambda, \gamma) \]
Thank you