Binary Puzzles as an Erasure Decoding Problem

Putranto Hadi Utomo Ruud Pellikaan
Eindhoven University of Technology
Dept. of Math. and Computer Science
PO Box 513. 5600 MB Eindhoven
p.h.utomo@tue.nl g.r.pellikaan@tue.nl

Appeared in:
Proceedings of the 36th WIC Symposium on Information Theory in the Benelux
Corrected version July 28, 2016

Abstract

Binary puzzles are interesting puzzles with certain rules. A solved binary puzzle
is an $n \times n$ binary array such that there are no three consecutive ones and also
no three consecutive zeros in each row and each column, the number of ones and
zeros must be equal in each row and in each column, every two rows and every
two columns must be distinct.

Binary puzzles can be seen as constrained arrays. Usually constrained codes
and arrays are used for modulation purposes. In this paper we investigate these
arrays from an erasure correcting point of view. We give lower and upper bound
for the rate of these codes, the probability of correct erasure decoding and erasure
decoding algorithms.

1 Introduction

Sudokus are nowadays very popular puzzles and they are studied for their mathematical
structure [2, 5, 18]. For instance the minimal number of entries that can be specified
in a single 9 $\times$ 9 puzzle to ensure a unique solution was in [14] conjectured to be
17, and this was proved by means of the chromatic polynomial of the Sudoku graph
[7]. Furthermore the erasure correcting capabilities and decoding algorithms of the
collection of $n \times n$ a Sudokus are considered [13, 16]. The asymptotic rate is still an
open problem [1, 7]. Solving an $n \times n$ Sudoku puzzle is an NP-hard problem [17].

The binary puzzle is also an interesting puzzle with certain rules and is the focus of
this paper. We look at the mathematical theory behind it. The solved binary puzzle
is an $n \times n$ binary array that satisfies:

1. no three consecutive ones and also no three consecutive zeros in each row and
each column,
2. every row and column is balanced, that is the number of ones and zeros must be
equal in each row and in each column,
3. every two rows and every two columns must be distinct.

Figure 1 is an example of a binary puzzle. There is only one solution satisfying
all three conditions. But there are 3 solutions satisfying (1) and (2). The solution
satisfying three all conditions is given in Figure 2. Figure 3 and 4 are solved puzzles
where the third constraint is excluded.
Figure 1: Unsolved Puzzle

Figure 2: Solved Puzzle

Figure 3: Solved Binario Puzzle with repetition of column/row allowed

Figure 4: Solved Binario Puzzle with repetition of column/row allowed
Binary and Sudoku puzzle can be seen as constrained arrays. Usually constrained codes and arrays are used for modulation purposes [8, 9]. We investigate these arrays from an erasure correcting point of view. We give lower and upper bound for the rate of these codes, the probability of correct erasure decoding and erasure decoding algorithms.

2 Constrained sequences and constrained array

Let \( C \) be a code in \( \mathbb{Q}^n \), where the alphabet \( \mathbb{Q} \) has \( q \) elements. Recall that the \( \) (information) rate \( R(C) \) of \( C \) is defined by

\[
R(C) = \frac{\log_2 |C|}{n}.
\]

In the following \( \mathbb{Q} = \mathbb{F}_2 \), \( n = lm \) and \( \mathbb{F}_2^{l \times m} \) is the set of binary \( l \times m \) arrays. Define:

\[
\begin{align*}
A_{l \times m} &= \{ X \in \mathbb{F}_2^{l \times m} \mid X \text{ satisfies (1)} \}; \\
B_{l \times m} &= \{ X \in \mathbb{F}_2^{l \times m} \mid X \text{ satisfies (2)} \}; \\
C_{l \times m} &= \{ X \in \mathbb{F}_2^{l \times m} \mid X \text{ satisfies (3)} \}; \\
D_{l \times m} &= \{ X \in \mathbb{F}_2^{l \times m} \mid X \text{ satisfies (1), (2) and (3)} \}.
\end{align*}
\]

The theory of constrained sequences, that is for \( l = 1 \), is well established and uses the theory of graphs and the eigenvalues of the incidence matrix to give a linear recurrence. An explicit formula for the number of such sequences of a given length \( m \) can be expressed in terms of the eigenvalues. The asymptotical rate is equal to \( \log_q (\lambda_{\text{max}}) \), where \( \lambda_{\text{max}} \) is the largest eigenvalue. See [8, 9]. Shannon [15] showed already that the following relation holds for \( m \geq 1 \):

\[
|A_{1 \times (m+2)}| = |A_{1 \times (m+1)}| + |A_{1 \times m}|.
\]

Asymptotically this gives

\[
R(A_{1 \times m}) \approx \log_2 \left( \frac{1}{2} + \frac{1}{2} \sqrt{5} \right), \quad \text{for } m \to \infty
\]

The number of balanced sequence is equal to a number of combination of ones, that is \( B_{1 \times 2m} = \binom{2m}{m} \) and asymptotically \( R(B_{1 \times 2m}) \approx 1 \), for \( m \to \infty \). It was shown [6, 10, 11] that the balanced property does not influence the asymptotic rate of constrained sequences. So \( R(A_{1 \times 2m} \cap B_{1 \times 2m}) \approx \log_2 \left( \frac{1}{2} + \frac{1}{2} \sqrt{5} \right) \), for \( m \to \infty \). We expect that a similar result holds for balanced constrained arrays.

For arrays we know that \( \binom{2m}{m} \leq |B_{2l \times 2m}| \leq \binom{2m}{m} \). From these inequalities it is can be shown that, asymptotically:

\[
\frac{1}{2} \leq R(B_{2m \times 2m}) \leq 1, \quad \text{for } m \to \infty
\]

Four arbitrary elements of \( B_{2m \times 2m} \) gives an element of \( B_{4m \times 4m} \). So \( |B_{4m \times 4m}| \geq |B_{2m \times 2m}|^4 \). Therefore \( R(B_{2m \times 2m}) \) is increasing in \( m \).
Now, consider $C_{l \times m}$. We clearly have that $|C_{l \times m}| \leq 2^n(2^m - 1) \cdots (2^m - n + 1)$. Furthermore, if $m = n$, $|C_{(n+1) \times (n+1)}| \geq |C_{n \times n}| \cdot (2^{2n+1} - 2n2^n + n^2)$. This implies that, asymptotically:

$$R(C_{2m \times 2m}) \approx 1, \text{ for } m \to \infty$$

The size of $D_{2m \times 2m}$ can be approximated by smaller building blocks such that the conditions are still satisfied [4]. There are exactly two building block of size $2 \times 2$. Hence, $R(D_{2m \times 2m}) \geq \frac{1}{(2m)^2} \log_2(2^{2m^2}) = \frac{1}{4}$, for $m \geq 1$.

Numerically, we have

<table>
<thead>
<tr>
<th>$m$</th>
<th>$A_{2m \times 2m}$</th>
<th>$B_{2m \times 2m}$</th>
<th>$C_{2m \times 2m}$</th>
<th>$D_{2m \times 2m}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Size</td>
<td>Rate</td>
<td>Size</td>
<td>Rate</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>1</td>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>2030</td>
<td>0.69</td>
<td>90</td>
<td>0.41</td>
</tr>
<tr>
<td>3</td>
<td>3858082</td>
<td>0.61</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

### 3 Erasure Channel

Suppose $Q$ is a set of an alphabet and $C$ is a code in $Q^n$. Define $\hat{Q} = Q \cup \{-\}$, where the symbol "-" denotes a blank, that is an erasure, and $\hat{C} = \{ r \in \hat{Q}^n | r \text{ is obtained from a } c \in C \text{ by erasures} \}$.

Suppose $r$ is the received word given that $c$ is sent. We have $d(r, c)$ is the Hamming distance between $r$ and $c$. Since the errors are only blanks, $d(r, c)$ equal to the number of blanks in $r$. Let $c(r)$ be a closest codeword to $r$, then $d(r, C) = d(r, c(r))$. Let $p$ be the probability that a symbol is erased, and let $P_{ed,C}(p)$ denote the probability of correct erasure decoding. Then

$$P_{ed,C}(p) = \sum_{c \in C} P(c) \sum_{\substack{r \in \hat{C} \\text{c}(r)=c}} P(r|c)$$

Suppose $E_i(C) = \{ r \in \hat{C} | d(r, C) = i \}$ and $E_i(C) = |E_i(C)|$.

Define the homogenous erasure distance enumerator for code $C$ by

$$E_C(X, Y) = \sum_{i=0}^{n} E_i(C) X^{n-i}Y^i$$

**Proposition 3.1**

$$P_{ed,C}(p) = \frac{1}{|C|} E_C(1 - p, p)$$
Proposition 3.2 Let $C \subseteq Q^m$ and $D \subseteq Q^n$. We have
\[ E_{C \times D}(X,Y) = E_C(X,Y) \cdot E_D(X,Y) \]

Corollary 3.3
\[ P_{ed,C \times D}(p) = (P_{ed,C}(p)) \cdot (P_{ed,D}(p)) \]

Corollary 3.4
\[ P_{ed,C^n}(p) = (P_{ed,C}(p))^n \]

4 Binary Puzzle Solver

Binary puzzle can be seen as a SAT problem. Since each cell in the binary puzzle can only take the values ‘0’ and ‘1’, we can express the puzzle as an array of binary variables, where false corresponds to ‘0’ and true to ‘1’. Next, we express each condition in terms of a logical expression.

Suppose we have an $2^m \times 2^m$ array in the variables $x_{ij}$. The array satisfies the first condition, that there are no three consecutive ones and also no three consecutive zeros in each row and each column, if and only if the expression below is true:
\[
\left( \bigwedge_{j=1}^{2m} \left\{ \bigwedge_{k=1}^{2m-2} \left( \neg \left( \bigwedge_{i=k}^{k+2} x_{ij} \right) \right) \wedge \neg \left( \bigwedge_{i=k}^{k+2} \neg x_{ij} \right) \right) \right) \wedge \\
\left( \bigwedge_{i=1}^{2m} \left\{ \bigwedge_{k=1}^{2m-2} \left( \neg \left( \bigwedge_{j=k}^{k+2} x_{ij} \right) \right) \wedge \neg \left( \bigwedge_{j=k}^{k+2} \neg x_{ij} \right) \right) \right) 
\]

For satisfying the second condition on balancedness, the following expression must be true
\[
\left( \bigwedge_{j=1}^{2m} \left\{ \bigwedge_{1\leq i_1<\cdots<i_{m+1}\leq 2m} \left( \bigvee_{k=1}^{m+1} x_{i_k,j} \right) \right\} \right) \wedge \\
\left( \bigwedge_{i=1}^{2m} \left\{ \bigwedge_{1\leq j_1<\cdots<j_{m+1}\leq 2m} \left( \bigvee_{k=1}^{m+1} x_{i,j_k} \right) \right\} \right) \\
\left( \bigwedge_{j=1}^{2m} \left\{ \bigwedge_{1\leq i_1<\cdots<i_{m+1}\leq 2m} \left( \bigvee_{k=1}^{m+1} \neg x_{i_k,j} \right) \right\} \right) \wedge \\
\left( \bigwedge_{i=1}^{2m} \left\{ \bigwedge_{1\leq j_1<\cdots<j_{m+1}\leq 2m} \left( \bigvee_{k=1}^{m+1} \neg x_{i,j_k} \right) \right\} \right) .
\]

Note that the complexity of this expression grows as $\binom{2m}{m}$ which is exponentially in $m$. An alternative polynomial expression can be obtained.

The satisfiability of the third condition, that every two rows and every two columns must be distinct, is equal to
\[
\left( \bigwedge_{1\leq j_1<j_2\leq 2m} \left\{ \bigwedge_{i=1}^{2m} \neg \left[ (x_{i,j_1} \land x_{i,j_2}) \lor (\neg x_{i,j_1} \land \neg x_{i,j_2}) \right] \right\} \right) \wedge \\
\left( \bigwedge_{1\leq i_1<i_2\leq 2m} \left\{ \bigwedge_{j=1}^{2m} \neg \left[ (x_{i_1,j} \land x_{i_2,j}) \lor (\neg x_{i_1,j} \land \neg x_{i_2,j}) \right] \right\} \right) .
\]

It is shown in [3] that the binary puzzle is NP-complete.
References


http://nearly42.org


