21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Permutations within a code: variance of value and invariance of errors.

Marcelo Firer[1] and Jerry A. Pinheiro[2]

***Abstract*— Considering an $[n,k]_q$-linear code C and an ML-decoder a, we use a quadratic function for measuring (evaluating) errors occurred during the encoding-decoding process, we show (and with some restrictions also prove) that an encoding of C (that can actually be seen as a permutation of C) determined by a lexicographic order is sub-optimal.**

I. INTRODUCTION

In this paper we will use the theory introduced in [1] and [2] of expected loss in information theory to describe good encoders in a particular case, this theory permits to give a different treatment for different kind of errors, and that seems important when we are dealing with image, as showed in [3]. Given a $[n,k]_q$-linear code $\mathcal{C} \subset \mathbb{F}_q^n$, the well known error probability of $\mathcal{C}$ (which depends on the probability matrix of the channel and the decoder g) is defined by

$$P_e(\mathcal{C}) := \sum_{c\in\mathcal{C}} P_e(c)P(c)$$

where

$$P_e(c) := \sum_{x\notin D_g(c)} P(x|c)$$

and $D_g(c) := \{y \in \mathbb{F}_q^n : g(y) = c\}$ are the decision regions of $\mathcal{C}$. We will assume that the decoder is a function $g : \mathbb{F}_q^n \to \mathcal{C}$ such that $g(c) = c$ for all $c \in \mathcal{C}$. Let $f : \mathcal{I} \to \mathbb{F}_q^n$ be an encoder of the information set $\mathcal{I}$ into $\mathcal{C}$, so $f(\mathcal{I}) = \mathcal{C}$. Note that the error probability of $\mathcal{C}$ does not depend on the choice of the encoder f, in other words, given a decoder and a channel, the error probability of $\mathcal{C}$ is invariant under permutations of the encoder f (which is a bijection between $\mathcal{I}$ and $\mathcal{C}$). In [2], the authors generalized the concept of error probability to expected loss, in this case, the expect loss is not invariant over permutations of the encoder f because unlike of error probability, different errors have different importance in the encoding-decoding structure.

To introduce the expected error, we first need to define *loss function* which is a symmetric function $\mu : \mathcal{I}\times\mathcal{I} \to \mathbb{R}_+$ satisfying $\mu(i,i) = 0$ for all $i \in \mathcal{I}$. The loss $\mu(i,j)$ quantifies how much we loose by exchanging the information i by the information j. If f is an encoder of $\mathcal{I}$, let $\mu_f : \mathcal{C}\times\mathcal{C} \to \mathbb{R}_+$ be a function induced by μ in the code $\mathcal{C}$ defined by $\mu_f(c_i,c_j) := \mu(f^{-1}(c_i), f^{-1}(c_j))$.

*This work was partially supported by FAPESP, project 2007/56052-8

[1]Marcelo Firer is with Unicamp - State University of Campinas email: mfirer@ime.unicamp.br

[2]Jerry A. Pinheiro is with Unicamp - State University of Campinas e-mail: jerryapinheiro@gmail.com

The *expected loss* of a code $\mathcal{C}$ is defined by the mean over the expected error of the codewords, ie,

$$\mathbb{E}(\mathcal{C},\mu_f,g) := \sum_{c\in\mathcal{C}} \mathbb{E}(c,\mu_f,g)P(c)$$

where

$$\begin{aligned}\mathbb{E}(c,\mu_f,g) &:= \sum_{x\notin D_g(c)} \mu_f(g(y),c)P(y|c)\\ &= \sum_{y\in\mathbb{F}_q^n} \mu_f(g(y),c)P(y|c).\end{aligned}$$

If $\mu(i,j) = 1$ for all $i \neq j$, expected loss coincides with the definition of error probability. Note that with this new approach, the errors can be classified by relevance, and this classification is given by the loss function. From now on, given a fixed decoder, we will explore the characteristics of the expected loss looking for optimal or suboptimal characterizations of encoders. Looking at the expression for the expected loss function, it is clear that, in general, it is not invariant by the choice of the encoder $f : \mathcal{I} \to \mathbb{F}_q^n$. An initial study of the behaviour of the expected loss as a function of the encoder f is the goal of this work, as explained in the next session.

II. BAYES ENCODERS

In coding theory, for a given channel model, given a code, there is a unique decoder (the maximum likelihood decoder) that minimizes the error probability. With our approach this is not true any more and we need to search for pairs of encoders-decoders maps minimizing the expected loss for a given loss function and code. We will call such pairs *Bayes Encoders-Decoders*. In many cases it is desirable (or at least justified) to restrict the encoders to a smaller family, mainly because of algorithm complexity of general decoders. Fixing a decoder, our main goal is characterize good encoders, in the optimal case, our goal is find the encoders f^* satisfying

$$\mathbb{E}(\mathcal{C},\mu_{f^*},g) = \min_f \ \mathbb{E}(\mathcal{C},\mu_f,g). \tag{1}$$

where the minimum is over all encoders of $\mathcal{I}$ with image $\mathcal{C}$. An encoder satisfying Equation (1) will be called *Bayes Encoder*. Considering a loss function is invariant by translations, that is, a function satisfying $\mu(x+y,z+y) = \mu(x,z)\ \forall\ x,y,z \in \mathcal{I}$, Panek, in [1], gave a characterization for the Bayes encoders. However, invariance by translation is an artificial condition for most possible applications and we look for such condition for the general case that is, as we shall see, much harder.

ISBN: 978-90-367-6321-9

Suppose $\mathcal{C} = \{c_1, \ldots, c_{q^k}\}$. It is straightforward to show that

$$\mathbb{E}(\mathcal{C}, \mu_f, g) = \sum_{i=1}^{q^k} \sum_{j=1}^{q^k} H_g(c_i, c_j)\mu_f(c_j, c_i) \tag{2}$$

where

$$H_g(c_i, c_j) = \sum_{y \in g^{-1}(c_j)} P(c_i|y)P(y).$$

Now using Equation (2), if A^g and B^f are matrix such that $A^g_{ij} = H_g(c_i, c_j)$ and $B^f_{ij} = \mu_f(c_i, c_j)$ then

$$\mathbb{E}(\mathcal{C}, \mu_f, g) = Tr(A^g B^f)$$

where $Tr(.)$ is the matrix trace function.

Fixing an encoder f, it is clear that if P is a permutation matrix, then $H = PB^f P^t$ is also a matrix constructed as above using an encoder h, ie, $H = B^h$, in addition, if f and h are two encoders, there is a permutation matrix P such that $B^h = PB^f P^t$, therefore we have the following theorem.

Theorem 1: Given a decoder g, an encoder f, if A^g and B^f are matrix defined as above, then

$$\min_P \ Tr(A^g P B^f P^t) = \min_f \mathbb{E}(\mathcal{C}, \mu_f, g),$$

where the left side minimum is over the permutation matrix set.

If σ is a permutation over $\{1, \ldots, q^k\}$, and $B^{f,\sigma}$ is the matrix with entries $B^{f,\sigma}_{ij} = \mu_f(c_{\sigma(i)}, c_{\sigma(j)})$, given any encoder f, as direct consequence of this theorem, the problem to find the Bayes encoder is equivalent to find a permutation σ of the codewords such that $B^{f,\sigma} = PB^f P^t$ for some P and P minimizes the trace. Now it is clear that the search space has $q^k!$ elements and characterizations for the general case better than given by the Theorem 1 is difficult. Despite the fact that this proposition brings the encoder problem into a class of well studied problems, minimizing the trace of matrices, the search for a solution still remains a very difficult problem. In the next section we will work with a very particular (the simplest) case.

Even if the decoder is fixed, the problem to find good encoders remains difficult, and to estimate the performance of a proposed encoder we need to know the distribution of the expected loss (as a function of the encoder f) or some indicative data, such as bounds. A good and natural bound is the mean of the expected loss which is simple to calculate when the decoder is fixed.

Proposition 1: If g is a fixed decoder of the $[n, k]_q$ linear code $\mathcal{C}$ and μ is a loss function, the mean expected loss is given by

$$\begin{aligned}\frac{\sum_f \mathbb{E}(\mathcal{C}, \mu_f, g)}{q^k!} &= \mathbb{E}(\mathcal{C}, \mu_{mean}, g) \\ &= \sum_{c \in \mathcal{C}} \sum_{y \in \mathbb{F}_q^n} \mu_{mean}(g(y), c)P(y|c)P(c)\end{aligned}$$

where $\mu_{mean}(x, x) = 0$ for all $x \in \mathcal{C}$ and

$$\mu_{mean}(x, y) = \frac{\sum_{s=1}^t (\#A_{j_s})j_s}{q^k(q^k - 1)}$$

for all $x \neq y$, furthermore, $\{j_1, \ldots, j_t\}$ are the all possible real values assumed by μ and $A_{j_k} = \{(x, y) \in \mathcal{I} \times \mathcal{I} | \mu(x, y) = j_k\}$.

Proof: First, note that

$$\frac{\sum_f \mathbb{E}(\mathcal{C}, \mu_f, g)}{q^k!} = \frac{1}{q^k!} \sum_{c \in \mathcal{C}} \sum_{y \in \mathbb{F}_q^n} P(y|c)P(c) \sum_f \mu_f(g(y), c). \tag{3}$$

where the left sum is over all encoders. If the image of μ_f is the set $Img = \{i_1, \ldots, i_s\}$, then

$$\begin{aligned}\sum_f \mu_f(x, y) = \sum_{t=1}^s \sum_{\substack{f \\ \mu_f(x,y)=i_t}} i_t &= \sum_{t=1}^s (q^k - 2)!(\#A_{i_t})i_t \\ &= (q^k - 2)! \sum_{t=1}^s (\#A_{i_t})i_t,\end{aligned}$$

where $A_{i_t} = \{(x, y) \in \mathcal{I} \times \mathcal{I} | \mu(x, y) = i_t\}$. By Equation 3,

$$\frac{\sum_f \mathbb{E}(\mathcal{C}, \mu_f, g)}{q^k!} = \frac{\sum_{t=1}^s (\#A_{i_t})i_t}{q^k(q^k - 1)} \sum_{c \in \mathcal{C}} \sum_{\substack{y \in \mathbb{F}_q^n \\ g(y) \neq c}} P(y|c)P(c).$$

Let $\lambda = (\sum_{t=1}^s (\#A_{i_t})i_t)/(q^k(q^k - 1))$ be the constant of the expression above and define $\mu_{mean}(i, j) = \lambda$ if $i \neq j$ and $\mu_{mean}(i, j) = 0$ otherwise. Therefore

$$\begin{aligned}\frac{\sum_f \mathbb{E}(\mathcal{C}, \mu_f, g)}{q^k!} &= \sum_{c \in \mathcal{C}} \sum_{y \in \mathbb{F}_q^n} \mu_{mean}(g(y), c)P(y|c)P(c) \\ &= \mathbb{E}(\mathcal{C}, \mu_{mean}, g).\end{aligned}$$

■

III. A particular Case

From now on, we will assume that the channel is a binary DSMC, therefore the conditional probabilities are given by

$$P(y|c) = (1 - p)^n \left(\frac{p}{1 - p}\right)^{d_H(y,c)}$$

where y and c are words of length n, $0 \leqslant p \leqslant 1/2$ is the error probability of each symbol and $d_H(.,.)$ is the well known Hamming distance function. If $\mathcal{C}$ is a $[n, n]_2$ linear code, there is a unique decoder g' of $\mathcal{C}$, indeed, there is a unique bijection $g' : \mathcal{C} \to \mathcal{C}$ satisfying $g'(c) = c$ for all $c \in \mathcal{C}$, namely, the identity function. This will actually be true for any channel satisfying $P(y|c) \leq P(c|c), \forall y \in \mathbb{F}_q^n$. In such a case, to find a Bayes encoder-decoder pair is equivalent to find Bayes encoder, since there is a unique decision to be made: accept as true whatever the message you receive. Codes like this ones (without redundancy) are rare but are used, for example, for image transmission in the satellite CBERS-2 (http://www.cbers.inpe.br/ingles/), in a cooperation program between Brazil an China. However, more then looking for possible applications, we believe that

understanding this instance may be a key for the general case.

Motivated by applications in image transmission, suppose $n = 8$ and that $\mathcal{I} = \{0, 1, \ldots, 255\}$ is the set of all gray colors in RGB. Also, suppose $i_1 = 0, i_2 = 1, \ldots, i_{256} = 255$ an ordering in $\mathcal{I}$. There is a natural normalized loss function μ' in $\mathcal{I}$, namely

$$\mu'(i_s, i_t) = \frac{1}{255^2}|i_s - i_t|^2 = \frac{1}{255^2}|s - t|^2 \quad \forall\ i_s, i_t \in \mathcal{I}.$$

Suppose $\mathcal{C} = \{c_1, \ldots, c_{256}\}$ where the codewords are ordered by the lexicographic order and define $f'(i_j) = c_j$ an encoder of $\mathcal{I}$. What we will do next is show that the expected loss in this case is better than the mean of the expected loss, also we have reasons to believe that the encoder f' as constructed is a Bayes encoder.

Note that in the gray scale case, $j_s = (s/255)^2$ for all $s \in \{1, \ldots, 255\}$ and $\#A_{j_s} = 2(256 - s)$, so

$$\mathbb{E}(\mathcal{C}, \mu_{mean}, g') = \sum_{i=1}^{256}\sum_{\substack{j\neq i\\ j=1}}^{256} \frac{\sum_{s=1}^{255}(2^8 - s)s^2}{128(255)^3} r^{d(c_i,c_j)} = p(r) \tag{4}$$

where $r = p/(1-p)$. By other side,

$$\mathbb{E}(\mathcal{C}, \mu'_f, g') = \sum_{i=1}^{256}\sum_{\substack{j\neq i\\ j=1}}^{256} \frac{1}{255^2}(i-j)^2 r^{d(c_i,c_j)} = q(r). \tag{5}$$

In this case, the coefficients of the polynomials p and q are known, indeed,

$$\begin{aligned} p(r) = {} & \frac{65792}{765}r + \frac{460544}{765}r^2 + \frac{460544}{255}r^3 + \frac{460544}{153}r^4 \\ & + \frac{460544}{153}r^5 + \frac{460544}{255}r^6 + \frac{460544}{765}r^7 + \frac{65792}{765}r^8 \end{aligned}$$

and

$$\begin{aligned} q(r) = {} & \frac{67371008}{195075}r + \frac{235798528}{195075}r^2 + \frac{471597056}{195075}r^3 \\ & + \frac{117899264}{39015}r^4 + \frac{471597056}{195075}r^5 + \frac{235798528}{195075}r^6 \\ & + \frac{67371008}{195075}r^7 + \frac{8421376}{195075}r^8. \end{aligned}$$

All the coefficients of these polynomials are positive, so they do not have root in the interval $(0, 1]$, furthermore, for all $r \in (0, 1)$,

$$h(r) = \frac{p(r)}{q(r)} = \frac{255(1+r)^7}{128(8 + 28r + 56r^2 + 70r^3 + 56r^4 + 28r^5 + 8r^6 + r^7)}.$$

Therefore

$$h'(r) = \frac{255(1+r)^6(28 + 56r + 70r^2 + 56r^3 + 28r^4 + 8r^5 + r^6)}{128(8 + 28r + 56r^2 + 70r^3 + 56r^4 + 28r^5 + 8r^6 + r^7)^2}$$

Notice that $h'(r) > 0$ for all $r \in (0, 1)$ which implies that h is increasing in this interval, we also have $h(1) = 1$, so $h(r) < 1$ for all $r \in (0, 1)$, as $p(0) = q(0) = 0$, we conclude that $\mathbb{E}(\mathcal{C}, \mu'_{f'}, g') \leqslant \mathbb{E}(\mathcal{C}, \mu_{mean}, g')$ for all $p \in (0, 1/2)$ and in particular, the inequality is strict when $p \in (0, 1)$. Therefore, we had proved the following theorem:

Theorem 2: Under the hypothesis used in the construction above, for every $p \in (0, 1/2)$, the expected loss of the 8-dimensional code over $\mathbb{F}_2^8$ is less than the mean of the expected loss, ie,

$$\mathbb{E}(\mathcal{C}, \mu'_{f'}, g') \leqslant \mathbb{E}(\mathcal{C}, \mu_{mean}, g').$$

The encoder constructed seems to be a very good choice because its expected loss is smaller than the mean and also given a finite subset S of $[0, 1/2]$, if for each $p \in S$ we choose uniformly a permutation matrix P, the graphic of the expected loss has a nice behaviour, indeed see Figure 1.

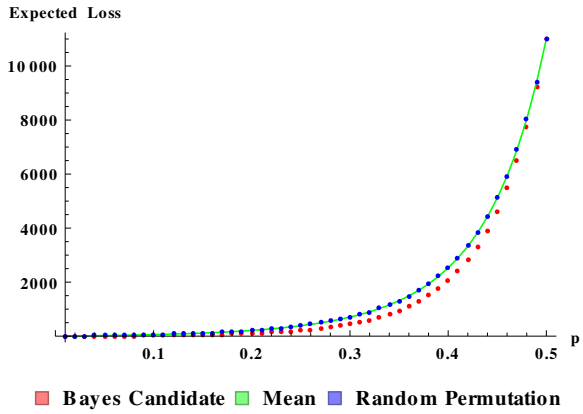

Fig. 1. p in the interval $[0, 1]$.

To be clear, the Graphic 1 for the values of p in the interval $(0, 0.25)$ is represented in the Figure 2.

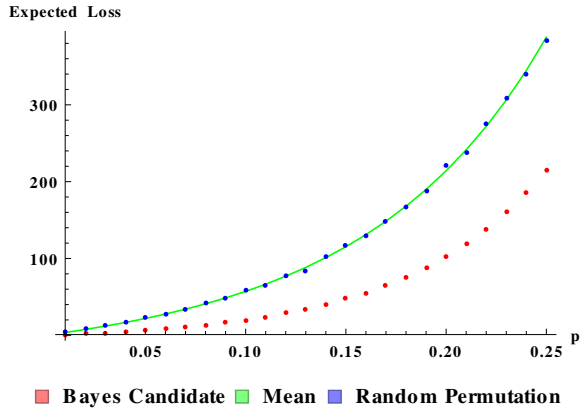

Fig. 2. p in the interval $[0, 0.25]$.

The construction made for the 8-dimensional case is similar for other dimensions, but show for a n-dimensional code that this construction is better than the mean is still an open problem. For the cases where we can calculate the minimum of the expected loss ($n = 2$ and $n = 3$), by similar construction it is possible verify that the encoder f' is indeed a Bayes encoder. By this results in low dimension and by the

nice behaviour of the function plotted in the Graphic 1, we did not proof yet, but we would be surprise if the encoder f' is not a Bayes encoder for the 8-dimensional case.

REFERENCES

[1] Luciano Panek. *Codificação na Presença do Valor Semântico da Informação*. PhD thesis, Universidade Estadual de Maringá, 2012.

[2] M. Firer, L. Panek and L. L. Ramos Rifo, *Coding in the Presence of Semantic Value of Information: Unequal Error Protection Using Poset Decoders*, http://arxiv.org/pdf/1108.3832.pdf.

[3] M. Firer, L. Panek and L. L. Ramos Rifo, *Coding and decoding schemes tailor made for image transmission*, Information Theory and Applications Workshop (ITA), 2013.