Blind Estimation of Encoder and Interleaver Characteristics in a Non Cooperative Context

Gilles Burel, Roland Gautier

Blind Estimation of Encoder and Interleaver Characteristics in a Non Cooperative Context

Gilles BUREL and Roland GAUTIER
LEST, Université de Bretagne Occidentale
CS 93837, 29238 BREST cedex 3, FRANCE
Gilles.Burel@univ-brest.fr   http://www.univ-brest.fr/lest/tst/

ABSTRACT
In most digital transmission systems, the data stream is first encoded, then sent to an interleaver that rearranges the encoded symbols in order to provide protection against error bursts. In this paper, we consider analysis of an interleaved stream in a non cooperative context (i.e. military or spectrum surveillance application). In this context, the encoder and the interleaver used by the transmitter are unknown. We propose an approach that is able to estimate useful information, such as the interleaver period and the code rate. The method uses only the intercepted interleaved stream. Furthermore, we are able to perform a blind synchronization on the interleaver blocks.

The approach is based on linear algebra theory: we show that the normalized rank of a matrix Z (whose columns are analysis blocks taken from the interleaved stream) decreases when the size of the analysis blocks is a multiple of the interleaver period. Furthermore, we show that the maximum decrease is obtained when the analysis blocks are synchronized with the interleaver blocks, and that it is linked to the code rate.

KEY WORDS
Communication Systems, Interleaver, Blind Estimation, Non-Cooperative, Coding, Digital Transmissions,

1 Introduction

An interleaver is a device commonly used in conjunction with error correcting codes to counteract the effect of burst errors (Fig. 1). Indeed, in many digital communication systems [1], error correcting codes are not used alone, because the encoder provides protection against uniformly distributed errors, but it is not robust with respect to burst errors.

The information data, which is usually binary, is first encoded in order to add redundancy for protection against uniformly distributed random errors on the transmission channel. Then, the interleaver rearranges the encoded symbols so that the symbols from a codeword are separated by more than the typical length of a burst of errors [2]. As a consequence, the channel appears as a random-error channel to the decoder. The interleaved encoded data is then used to modulate a carrier.

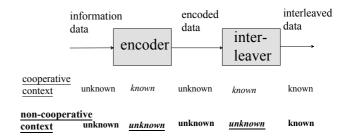|	information data	encoded data		interleaved data
cooperative context	unknown	*known*	unknown	*known*	known
non-cooperative context	**unknown**	***unknown***	**unknown**	***unknown***	**known**

Figure 1. Encoder and Interleaver

On the receiver side, the signal is demodulated, then sent to an inverse interleaver (which restores the initial order of the encoded symbols) and finally to the decoder. This requires that the receiver knows which encoder and which interleaver are used at the transmitter, in order to apply the appropriate inverse transformations. This knowledge is always available in the classical cooperative context. A good introduction to interleavers can be found in [3][4].

In this paper, we investigate the non-cooperative context, which is encountered in military applications or spectrum surveillance applications. In this context, a signal is intercepted, but we do not know which encoder and interleaver are used at the transmitter side. Only the interleaved data is known. Our objective, in this paper, is not to address the full problem of estimating the encoder and interleaver structures, but to provide an approach that helps to reach this objective. More precisely, what we propose is a method to:

- Estimate the interleaver period

- Perform a blind synchronization on the interleaver blocks

- Estimate the code rate.

The method is based on linear algebra. The paper is organized as follows. In Section 2, we present the mathematical model for the encoder and the interleaver. In Section 3, we propose a method for blind estimation of the interleaver period. An approach for blind synchronization on the interleaver blocks is proposed in Section 4, and the

estimation of the code rate is considered in Section 5. Finally, experimental results are provided in Section 6 and a conclusion is drawn in Section 7.

2 Mathematical model for the encoder and the interleaver

In the paper, for clarity of presentation, we will restrict the presentation to block encoders. Nevertheless, the method can be adapted to convolutional encoders by changing the sizes of the submatrices and their relative disposition.

A block encoder is defined by a full-rank generator matrix $\widetilde{G}_c$ which transforms each block of k_c information symbols into a block of n_c encoded symbols ($k_c < n_c$). Representing the information block and the encoded block by vectors $\widetilde{x}$ and $\widetilde{y}$, we have:

$$\widetilde{y} = \widetilde{G}_c \widetilde{x} \qquad (1)$$

The ratio below is called the code rate:

$$r = \frac{k_c}{n_c} \qquad (2)$$

Usually, the symbols are binary, hence $\widetilde{G}_c$ is a binary matrix and computations are performed modulo 2. A very simple example, which we will use throughout the paper to illustrate our approach, is the ($n_c = 3$, $k_c = 2$) parity code, which generator matrix is:

$$\widetilde{G}_c = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \qquad (3)$$

This encoder adds a parity bit to each block of 2 information bits.

As mentioned earlier, the encoded data is then sent to an interleaver which provides protection against bursts of errors. The interleaver can be modeled by a permutation matrix P, the size of which is $n_i \times n_i$, where n_i is called "the interleaver period". This means that the interleaver performs a permutation within each block of n_i encoded symbols. If we note y the vector representing a block of n_i encoded symbols and z the vector representing the corresponding interleaved block, we have:

$$z = Py \qquad (4)$$

In order to avoid useless complexity of the transmitter and receiver hardware, the interleaver period is a multiple of the size of the encoded block. That is:

$$n_i = b_c n_c \qquad (5)$$

where b_c is an integer. Now, if we consider a block x of $k_i = b_c k_c$ information symbols (that is the concatenation of b_c information blocks), we can write:
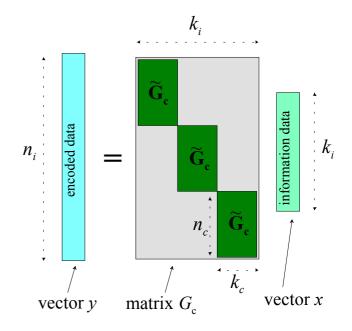
$$y = G_c x \qquad (6)$$

Figure 2. Structure of matrix G_c

where the structure of matrix G_c is shown on Figure 2 (the figure shows an example for $b_c = 3$). Note that, since $\widetilde{G}_c$ is full-rank, G_c is also full-rank.

Then, using Equations 4 and 6, we can write:

$$z = G_i x \qquad (7)$$

where G_i is the following $n_i \times k_i$ matrix:

$$G_i = P G_c \qquad (8)$$

Note that, since G_c and P are full-rank, G_i is also full-rank.

3 Blind estimation of the interleaver period

3.1 Principle of the approach

Our approach is based on dividing the interleaved stream into analysis blocks of an arbitrary size n_a, then on building a matrix Z whose columns are these blocks, and then examining the behavior of the ratio ρ defined below with respect to n_a:

$$\rho = \frac{rank\,(Z)}{n_a} \qquad (9)$$

Let us illustrate the approach on a simple example. Consider the ($n_c = 3$, $k_c = 2$) parity code defined above, and a random interleaver of period $n_i = 12$ (hence $b_c = 4$). From an interleaved stream containing 2400 symbols, we build matrix Z and compute ρ for increasing values of n_a. Figure 3 shows the obtained values of ρ as a function of n_a. We can note that ρ is equal to 1, except when n_a is

a multiple of the interleaver period n_i. As a consequence, the method allows to estimate the value of n_i. The dashed curves are the theoretical upper and lower bounds, that will be explained later.
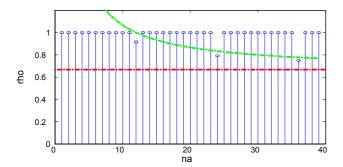

Figure 3. Estimation of the interleaver period (the dashed curves are lower and upper bounds).

3.2 Mathematical details

Now, let us detail the mathematical considerations that allowed us to propose this method and let us show why it works.

Let us consider a long interleaved data stream obtained from an intercepted signal. Figure 4 shows the relation between this interleaved data stream and the corresponding information stream. We recall that, in the non-cooperative context, only the interleaved data stream is known. Everything else (the large matrix, the submatrices G_i, their sizes, the information stream, etc.) is unknown.

If we divide the interleaved stream into analysis blocks of an arbitrary size n_a, as shown on the figure, the submatrix G_a which represents the transformation between information data and an analysis block varies from one block to another. However, if n_a is a multiple of the interleaver period n_i, the submatrix G_a is the same matrix for all analysis blocks, as shown on Figure 5.

Therefore, if n_a is a multiple of the interleaver period (i.e. $n_a = b_i n_i$, where b_i is an integer), we can write:

$$Z = G_a X \tag{10}$$

where Z is a matrix whose columns are the interleaved blocks and X a matrix whose columns are the corresponding (overlapping) information blocks. The length of the intercepted stream is assumed sufficient to have the number of columns in matrix Z greater than the number of rows. Due to the equation above, the rank of matrix Z is subject to the following inequality:

$$rank\,(Z) \le \min\,(k_a, n_a) \tag{11}$$

because, due to equation 10, the rank of Z cannot be greater than the rank of G_a. And the rank of G_a, which
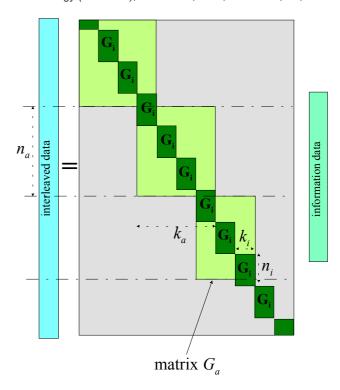

Figure 4. Structure of matrices G_a (when n_a is not a multiple of n_i)

is an $n_a \times k_a$ matrix, cannot be higher than its smallest dimension.

If we examine in more details the structure of matrix G_a (see Figure 5), we can obtain a closer upper bound:

$$rank(Z)$$
$$\le \min\{\,(b_i - 1)\,k_i + \min\,(d, k_i) + \min\,(n_i - d, k_i)\ ,\ n_a\} \tag{12}$$

where d is an integer which represents the unknown desynchronization between interleaver blocks and analysis blocks ($0 \le d \le n_i - 1$). In practice, if the length of the observed interleaved stream is sufficient, this close upper bound is often reached because there is enough diversity in the data.

Then, since $k_i / n_i = k_c / n_c = r$, the upper bound for the ratio $\rho = rank(Z)/n_a$ is:

$$\rho \le \min\left(r + \frac{1}{b_i}\left(-r + \min\,(\alpha, r) + \min\,(1 - \alpha, r)\right), 1\right) \tag{13}$$

where

$$\alpha = \frac{d}{n_a} \tag{14}$$

and $0 \le \alpha \le 1/b_i - 1/n_a$. Please note that the obtained value of ρ is the same for α and $1 - \alpha$, hence we can restrict the study to $0 \le \alpha \le 1/2$.
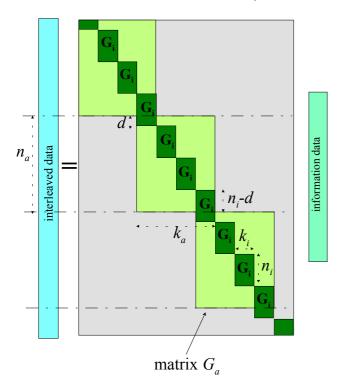
matrix G_a

Figure 5. Structure of matrix G_a when n_a is a multiple of n_i (here $n_a = 3n_i$)

To avoid boring details, let us examine only the case where $r \geq 1/2$ (results for $r \leq 1/2$ are similar and trivial to obtain). Then, we have two intervals for α to consider:

- For $0 \leq \alpha \leq 1 - r$ (hence $r \leq 1 - \alpha \leq 1$), equation 13 becomes:

$$\rho \leq r + \frac{\alpha}{b_i} \qquad (15)$$

- For $1 - r \leq \alpha \leq 1/2$ (hence $1/2 \leq 1 - \alpha \leq r$), equation 13 becomes:

$$\rho \leq r + \frac{1-r}{b_i} \qquad (16)$$

Finally, whichever the unknown desynchronization is, we can always write:

$$\rho \leq r + \frac{1-r}{b_i} \qquad (17)$$

This "desynchronization independent" upper bound is shown on Figure 3. It is easy to see that, when b_i is sufficiently large, this upper bound is less than one, hence we are sure that noticeable decreases of ρ will be observed. Sometimes (depending on the desynchronization between the analysis blocks and the interleaver blocks) this "desynchronization independent" upper bound is reached, as shown on Figure 6.
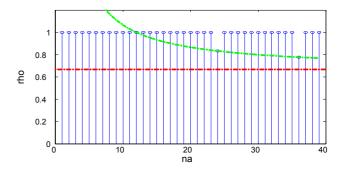

Figure 6. Estimation of the interleaver period: in this example, the upper bound is reached.

4 Blind synchronization with the interleaver blocks.

Once the interleaver period n_i is estimated, we set the size of the analysis blocks to $n_a = n_i$, and we skip the first $\widehat{d}$ symbols in the interleaved stream ($0 \leq \widehat{d} \leq n_i - 1$). If we plot ρ versus $\widehat{d}$, we obtain the result shown on Figure 7.
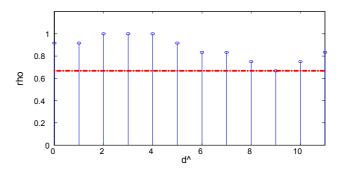

Figure 7. Blind synchronization with the interleaver blocks (the dashed horizontal line shows the code rate).

Here, the initial desynchronization of the intercepted interleaved stream was $d = 9$, and we can note that the minimum of ρ is obtained for $\widehat{d} = d$. Hence, locating the minimum of ρ allows to estimate d, and then to synchronize the analysis blocks with the interleaver blocks.

This result is not difficult to explain. Indeed, when we skip the first $\widehat{d}$ symbols, Equation 14 becomes:

$$\alpha = \mathrm{mod}\left(\frac{d - \widehat{d}}{n_i}, 1\right) \qquad (18)$$

and, since $b_i = 1$ (because $n_a = n_i$), equations 15 and 16 become:

- For $0 \leq \alpha \leq 1 - r$ (hence $r \leq 1 - \alpha \leq 1$):

$$\rho \leq r + \alpha \qquad (19)$$

- For $1 - r \leq \alpha \leq 1/2$ (hence $1/2 \leq 1 - \alpha \leq r$):

$$\rho \leq 1 \qquad (20)$$

Hence, since in practice the close upper bound is usually reached, we can note that the minimum of ρ is expected for $\alpha = 0$, that is for $\widehat{d} = d$. We will see, in the next Section, that when $\widehat{d} = d$, we also have $\rho = r$ (this result can be used to estimate the code rate r).

5 Blind estimation of the code rate

Once synchronization is done, let us skip the first d symbols of the interleaved stream, and compute ρ for increasing values of n_a again. We obtain the result shown on Figure 8.
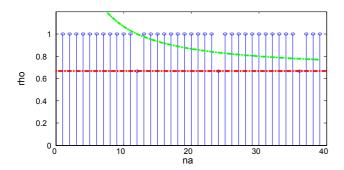

Figure 8. ρ versus n_a when the analysis blocks are synchronized

We can note that when n_a is a multiple of the interleaver period n_i, the value of ρ becomes 2/3, which, in fact, is exactly the code rate ($r = k_c/n_c = 2/3$). Hence, we obtain the value of the code rate.

Again, this result is not difficult to explain. If we skip the first d symbols of the interleaved stream, we are synchronized on the interleaver blocks: the structure of the submatrices G_a is shown on figure 9.

From the structure of G_a, we can see that G_a is full-rank (because G_i is full-rank, see Section 2). Hence:

$$rank\,(Z) = k_a \qquad (21)$$

and

$$\rho = \frac{k_a}{n_a} \qquad (22)$$

Since $k_a = b_i k_i = b_i b_c k_c$ and $n_a = b_i n_i = b_i b_c n_c$, we obviously have:

$$\rho = r \qquad (23)$$

In practice, in order to save computation time, it is not necessary to perform these computations (unless for verification), nor to draw figure 8, because during the blind synchronization process, the minimum value obtained for ρ was $\rho = r$ (see figure 7).
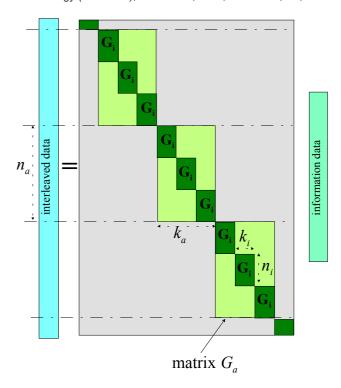

Figure 9. Structure of matrix G_a when the analysis blocks are synchronized

6 Experimental results

In this section we show experimental results obtained with a ($n_c = 7, k_c = 4$) Hamming code and a random interleaver of period $n_i = 28$. We used an interleaved stream of 3584 symbols for analysis. The desynchronization was $d = 22$.

Figure 10 shows the obtained values of ρ with respect to the size of the analysis blocks (n_a). As expected, we can note that the value of ρ decreases when n_a is a multiple of the interleaver period.
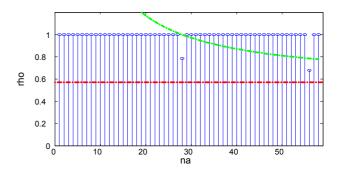

Figure 10. Estimation of the interleaver period (Hamming code)

Once n_i is estimated, we set $n_a = n_i$ and we compute

the value of ρ with respect to $\widehat{d}$, which is the number of skipped symbols (Fig. 11). As expected, the minimum value is equal to the code rate ($r = 4/7$) and it is obtained for $\widehat{d} = d = 22$.
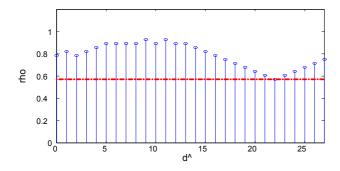

Figure 11. Blind synchronization with the interleaver blocks (Hamming code). The dashed horizontal line shows the code rate.

7 Conclusion

In this paper, we have proposed an approach to estimate some encoder and interleaver characteristics in a non cooperative context. The approach is based on linear algebra.

Thanks to the structure of transformation matrices, we have shown that the behavior of the normalized rank of matrix Z, whose columns are analysis blocks taken from the intercepted interleaved stream, provides a lot of information. This information allows to estimate the interleaver period and the code rate, as well as to perform a blind synchronization with the interleaver blocks.

Further work will include extraction of more information (interleaver permutation matrix, encoder generator matrix) from the synchronized analysis blocks.

References

[1] John G. Proakis, *Digital Communications* (Mc Graw Hill Eds, Third Edition, 1995)

[2] J.L.Ramsey, Realization of Optimum Interleavers, *IEEE Trans. on Information Theory*, Vol. 16, No. 3, May 1970, pp. 338-345

[3] K. Andrews, C. Heegard, D. Kozen, A theory of Interleavers, *Technical Report 97-1634*, Computer Science Department, Cornell University, June 1997

[4] C. Heegard, S.B. Wicker, *Turbo Coding* (Kluwer Academic Publishers, 1st Edition, January 1999)