Discrete Memoryless Channels with Memoryless Output Sequences

Marcelo S. Pinho
Department of Electronic Engineering
Instituto Tecnologico de Aeronautica
Sao Jose dos Campos, SP — 12228-900, Brazil
Email: mpinho@ieee.org

***Abstract*—This work studies what kind of input sequence a Discrete Memoryless Channel must have to produce a memoryless output. Actually, it is shown that there exist Discrete Memoryless Channels where an input sequence with memory can lead a memoryless output sequence. In these cases, the mutual information between the input and the output is equal to the sum of mutual informations between each pair of random variables drawn from the input and the output sequences, that is, $I(X^n;Y^n) = \sum_{i=1}^{n} I(X_i;Y_i)$. This result is not a guarantee that those input sequences achieve the channel capacity. This paper presents some examples of Discrete Memoryless Channels where input sequence with memory produces a memoryless output computing their mutual informations. This work also presents some results comparing the mutual information and the channel capacity for Discrete Memoryless Channels with Markovian input sequences and memoryless output sequences.**

I. Introduction

More than 60 years ago, the channel capacity theorem for a Discrete Memoryless Channel (DMC) was presented in the seminal paper of Claude Shannon [1]. The capacity of a DMC can be computed by the formulae

$$C^{\text{dmc}} = \max_{X} I\left(X;Y\right), \tag{1}$$

where X is the channel input, and Y is the channel output [2]. Distinct authors use different ways to define a general expression for the capacity of a discrete channel with memory, and each definition holds for a specific class of channels. A popular expression is given by

$$C = \lim_{n\to\infty} \sup_{X^n} \frac{1}{n} \left\{I\left(X^n;Y^n\right)\right\}, \tag{2}$$

where $X^n = (X_1, \cdots, X_n)$ is the channel input sequence, and $Y^n = (Y_1, \cdots, Y_n)$ is the channel output sequence. In [3], it was proved that this expression holds for the class of information stable channel. A formulae for a broader class of discrete channels can be found in [4].

Since the later equation is a generalization of the former, it is obvious that (2) is reduced to (1) when the channel is memoryless. A very simple way to show this result is by the use of a popular theorem which states that for every DMC,

$$I\left(X^n;Y^n\right) \leq \sum_{i=1}^{n} I\left(X_i;Y_i\right), \tag{3}$$

with equality if $X_1, \cdots, X_n$ are statistically independent, see [5, theorem 4.2.1]. Let $p\left(y_i|x_i\right)$ denote the single letter transition probabilities of the channel, which is constant for every $i = 1, \cdots, n$. Then using (1) it can be concluded that

$$I\left(X_i;Y_i\right) \leq C^{\text{dmc}}, \tag{4}$$

for every i, where C^{dmc} is computed considering that the relationship between X and Y in (1) is given by the channel transition probabilities. Furthermore (4) holds with equality if X_i has the probability distribution which achieves the maximal mutual information. Therefore, from (3) and (4) it can be concluded that

$$I\left(X^n;Y^n\right) \leq nC^{\text{dmc}} \tag{5}$$

with equality if $X_1, \cdots, X_n$ are independent and identically distributed (i.i.d. — with distribution which achieves the maximum mutual information). From (3) and (5), it can be concluded (as expected) that

$$C = C^{\text{dmc}},$$

for every DMC. Furthermore the channel capacity is achieved if $X_1, \cdots, X_n$ are i.i.d., with a known distribution.

It is interesting to observe that the above result presents only a sufficient condition to achieve the channel capacity. However, the theorem 4.2.1 of [5] can be slightly modified to lead a necessary condition too. Actually, as mentioned in [6, Problem 1.25], for a DMC, (3) holds with equality iff the output sequence $Y_1, \cdots, Y_n$ is memoryless. This result can be easily proved since

$$\begin{aligned}
I\left(X^n;Y^n\right) &= H(Y^n) - H(Y^n|X^n) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\
&\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\
&= \sum_{i=1}^{n} I\left(X_i;Y_i\right).
\end{aligned}$$

Above, the following results are used. For a DMC, $H(Y^n|X^n) = \sum_{i=1}^{n} H(Y_i|X_i)$ and for every sequence Y^n, $H(Y^n) \leq \sum_{i=1}^{n} H(Y_i)$, with equality iff the components of Y^n are independent, see [2]. Therefore, a memoryless output sequence is a necessary condition to achieve the channel capacity. Here, it arises an interesting question. What kind of input sequence can produce a memoryless output? It is easy to see that if a memoryless sequence is applied in the input, a memoryless sequence will be observed in the output. Thus, the interesting point in the question is related to input sequences with memory. Is it possible to construct an input sequence with memory for a DMC which produces a memoryless output? This question is closed related to the problem 6-14 proposed in [7] and to the theorem 5.2.1 of [8]. Actually, in their book, Proakis and Salehi proposed the following problem.

> "Show that, for a DMC, the average mutual information between a sequence $X_1 X_2 \cdots X_n$ of channel inputs and the corresponding channel outputs satisfy the condition
>
> $$I\left(X_1, X_2, \cdots, X_n; Y_1, Y_2, \cdots, Y_n\right) \leq \sum_{i=1}^{n} I\left(X_i; Y_i\right)$$
>
> with equality if and only if the set of inputs symbols is statistically independent."

In the same way, Blahut in his book, presents the following theorem.

> "The mutual information between the block input and block output of a discrete memoryless channel satisfies
>
> $$I\left(X_1, \cdots, X_n; Y_1, \cdots, Y_n\right) \leq \sum_{\ell=1}^{n} I\left(X_\ell; Y_\ell\right)$$
>
> with equality if and only if the input random variables $X_1, \cdots, X_n$ are independent."

If the above statements are true, it can be concluded that it is not possible to construct an input sequence with memory for a DMC which produces a memoryless output. Actually, in the proof of theorem 5.2.1 presented in [8], it is affirmed that a necessary and sufficient condition to a memoryless output is a memoryless input (even though, it was not proved).

It is important to mention that the inequality (3) can be found in many books [2, lemma 8.9.2], [5, theorem 4.2.1], and [6, theorem 1.9]. However, it is only in [7, Problem 6-14] and in [8, theorem5.2.1] that is posted that the statistically independence of the input sequence is a necessary condition to equality.

The aim of this work, which is motivated by the problem 6-14 of [7], is to show that there are Discrete Memoryless Channels for which an input sequence with memory leads to a memoryless output (i.e., a counterexample for the statement of this problem and for the theorem 5.2.1 of [8]). For these channels, it is interesting to see if input sequences with memory can achieve the channel capacity. Even though this work does not stress this problem, it presents some results concerning the capacity of DMC's with input Markovian sequences. This paper is organized as follows. Section II is devoted to DMC's with the desired property. In Section III, the results concerning the capacity of DMC's with input Markovian sequences are shown. The last Section presents the final remarks.

II. Input sequence with memory producing a memoryless output in a DMC

As mentioned in the Introduction of this work, the statement of the problem 6-14 of [7] is equivalent to say that the output of a DMC is memoryless iff the input sequence is memoryless. In fact, the problem asks to the reader to prove that for a DMC,

$$I\left(X^n; Y^n\right) \leq \sum_{i=1}^{n} I\left(X_i; Y_i\right),$$

with equality iff the input sequence is memoryless. Since the equality holds iff the output is memoryless, if the statement is true, it can be concluded that the output is memoryless iff the input is memoryless too. The aim of this Section is to show that there are some DMC's for which an input sequence with memory can lead to a memoryless output. Actually, the most trivial example of this kind of Channel is the Binary Symmetric Channel (BSC) with null capacity. This channel has the property that the input is statistically independent of the output. Then $I\left(X^n; Y^n\right) = 0$ and $\sum_{i=1}^{n} I\left(X_i; Y_i\right) = 0$ for every input sequence (including any sequence $X_1, \cdots, X_n$ with memory), that is, $I\left(X^n; Y^n\right) = \sum_{i=1}^{n} I\left(X_i; Y_i\right) = 0$. In this case, it is also interesting to observe that any input sequence with memory achieves the channel capacity.

Even though the BSC with null capacity is an example for a channel where an input sequence with memory produces a memoryless output sequence and achieves the channel capacity, it is not a useful channel (since it has no transmission capacity). In fact, it is possible to show some examples based on memoryless channels with non null capacity. Let $A_I = \{a_1, \cdots, a_{m_I}\}$ and $A_O = \{b_1, \cdots, b_{m_O}\}$ be the input alphabet and the output alphabet of the DMC, respectively. Let $\mathbf{\Pi} = \{p_{i,j}\}$ denote a $m_O \times m_I$ matrix of transition probabilities, that is, $p_{i,j} = \Pr\left[\mathrm{Y} = \mathrm{b_j} | \mathrm{X} = \mathrm{a_i}\right]$. To produce an output random variable Y with a probability vector given by

$$\mathbf{p}_O = \left(\Pr\left[\mathrm{Y} = \mathrm{b_1}\right], \Pr\left[\mathrm{Y} = \mathrm{b_2}\right], \cdots, \Pr\left[\mathrm{Y} = \mathrm{b_{m_O}}\right]\right)^T,$$

then the input random variable X must have a probability vector $\mathbf{p}$ such that

$$\mathbf{p}_O = \mathbf{\Pi}\mathbf{p}. \tag{6}$$

Analogously, let $\mathbf{\Pi}^{(n)}$ denote a $m_O^n \times m_I^n$ matrix of transition probabilities for vectors with n symbols, that is, each element of $\mathbf{\Pi}^{(n)}$ is a probability of Y^n is equal to a specific sequence with n symbols in A_O, given that X^n is a known sequence with n samples of A_I. Since the channel is memoryless, $\mathbf{\Pi}^{(n)}$ can be written as a Krönecker product of $\mathbf{\Pi}$, that is,

$$\mathbf{\Pi}^{(n)} = \underbrace{\mathbf{\Pi} \otimes \mathbf{\Pi} \otimes \cdots \otimes \mathbf{\Pi}}_{\text{n times}}.$$

In this case, to produce an output sequence Y^n with a probability vector given by

$$\mathbf{p}_O^{(n)} = \begin{pmatrix} \Pr\left[\mathrm{Y^n} = \mathrm{b_1 b_1 \cdots b_1}\right] \\ \Pr\left[\mathrm{Y^n} = \mathrm{b_2 b_1 \cdots b_1}\right] \\ \vdots \\ \Pr\left[\mathrm{Y^n} = \mathrm{b_{m_O} b_{m_O} \cdots b_{m_O}}\right] \end{pmatrix}$$

the input random sequence X^n must have a probability vector $\mathbf{p}^{(n)}$ such that

$$\mathbf{p}_O^{(n)} = \mathbf{\Pi}^{(n)}\mathbf{p}^{(n)}. \tag{7}$$

If Y^n is a sequence of i.i.d. random variables then

$$\mathbf{p}_O^{(n)} = \underbrace{\mathbf{p}_O^{(1)} \otimes \mathbf{p}_O^{(1)} \otimes \cdots \otimes \mathbf{p}_O^{(1)}}_{\text{n times}}.$$

From (6), it is obvious that $\mathbf{p}_O^{(1)} = \mathbf{\Pi}\mathbf{p}^{(1)}$. Thus

$$\begin{aligned}\mathbf{p}_O^{(n)} &= \mathbf{\Pi}\mathbf{p}^{(1)} \otimes \mathbf{\Pi}\mathbf{p}^{(1)} \otimes \cdots \otimes \mathbf{\Pi}\mathbf{p}^{(1)} \\ &= (\mathbf{\Pi} \otimes \cdots \otimes \mathbf{\Pi}) \left(\mathbf{p}^{(1)} \otimes \cdots \otimes \mathbf{p}^{(1)}\right) \qquad (8) \\ &= \mathbf{\Pi}^{(n)} \left(\mathbf{p}^{(1)} \otimes \cdots \otimes \mathbf{p}^{(1)}\right). \qquad (9)\end{aligned}$$

From (9) it can be seen that the probability vector $\mathbf{p}^{(n)} = \left(\mathbf{p}^{(1)} \otimes \cdots \otimes \mathbf{p}^{(1)}\right)$ satisfies (7). Since this probability vector is a measure of a memoryless sequence, it can be concluded that for every valid i.i.d. output sequence, there is a memoryless input sequence which can be used to yield this output (a valid i.i.d. output sequence is a sequence which can be produced in the channel output by the use of a specific input). Furthermore, for this memoryless input, (3) holds with equality.

At this point, it is interesting to notice that depending of $\mathbf{\Pi}^{(n)}$, (7) can have more than one solution. In fact, the critical point is the rank of the matrix defined by

$$\mathbf{\Pi}^{\text{ext}} = \begin{pmatrix} & \mathbf{\Pi} & & \\ 1 & 1 & \cdots & 1 \end{pmatrix} \tag{10}$$

If $\operatorname{rank}\left(\mathbf{\Pi}^{\text{ext}}\right) = \mathrm{m_I}$, then there is only one solution for the system of linear equations defined in (6) which is a valid probability vector. Furthermore, it can also be shown that there is only one valid solution for (7) which is obviously the memoryless solution. In this case, there is no statistically dependent input sequence such that $I\left(X^n;Y^n\right) = \sum_{i=1}^{n} I\left(X_i;Y_i\right)$. This is the case of any BSC with error probability different from $\frac{1}{2}$. Therefore, for every non null capacity BSC, the input sequence is memoryless iff the output is memoryless too. However, if the rank of $\mathbf{\Pi}^{\text{ext}}$ is less than m_I, then (6) has more than one solution and these solutions can be combined to construct an input sequence with memory for which (3) holds with equality (an input with memory leading to a memoryless output). The following example illustrates this combination and it serves as a counterexample for the statement of the problem 6-14 of [7].

Example 1: Let C be a DMC with a ternary input alphabet and a ternary output alphabet, that is $m_I = m_O = 3$, and with a matrix of transition probabilities given by

$$\mathbf{\Pi} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

It is possible to show that

$$\operatorname{rank}\left(\mathbf{\Pi}^{\text{ext}}\right) = 2.$$

and thus for every probability vector $\mathbf{p}_O$, (6) has more than one solution. For example, if $\mathbf{p}_O$ is set to $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^T$, then

$$\mathbf{p}_1 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^T \tag{11a}$$

$$\mathbf{p}_2 = \left(\frac{1}{2}, 0, \frac{1}{2}\right)^T \tag{11b}$$

$$\mathbf{p}_3 = (0,1,0)^T \tag{11c}$$

are valid solutions for (6). For a vector with two components, since the channel is memoryless, $\mathbf{\Pi}^{(2)} = \mathbf{\Pi} \otimes \mathbf{\Pi}$. If the output probability vector is set to

$$\mathbf{p}_O^{(2)} = \left(\frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}\right)^T$$

then there are many solutions for (7) and some of them will produce a pair of statistically dependent random variables. In fact, it is easy to see that

$$\mathbf{p}^{(2)} = \left(\frac{1}{6}, 0, \frac{1}{6}, 0, \frac{1}{3}, 0, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}\right)^T \tag{12}$$

is a solution. It is also easy to observe that this random vector is not independent. For this solution, the mutual informations are given by

$$I(X_1;Y_1) = 0.4444 \tag{13a}$$

$$I(X_2;Y_2) = 0.3704 \tag{13b}$$

$$I(X^2;Y^2) = 0.8148 \tag{13c}$$

■

In Example 1, it is interesting to observe that even though $I(X^2;Y^2) = I(X_1;Y_1) + I(X_2;Y_2)$, the mutual information between the input and output vectors does not achieve its maximum value. Actually, if the solution

$$\mathbf{p}^{(2)} = \left(\frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}\right)^T$$

was chosen, then $I(X^2;Y^2) = 0.8888$. Example 1 is a particular case of the following example.

Example 2: Let C be a ternary-input-ternary-output channel with a matrix transition probabilities given by

$$\mathbf{\Pi} = \begin{pmatrix} 1-\varepsilon_1 & \varepsilon_1 & 0 \\ \varepsilon_1 & 1-\varepsilon_1-\varepsilon_2 & \varepsilon_2 \\ 0 & \varepsilon_2 & 1-\varepsilon_2 \end{pmatrix}, \tag{14}$$

where $\varepsilon_2 = \frac{1-2\varepsilon_1}{2-3\varepsilon_1}$. Let r_1, r_2 and r_3 be the rows of this matrix. It is easy to see that $r_1 + r_2 + r_3 = (1,1,1)$. Furthermore, it can shown that $\alpha r_1 + (1-\alpha) r_3 = r_2$, if $\alpha = \frac{\varepsilon_1}{1-\varepsilon_1}$. Therefore, $\operatorname{rank}\left(\mathbf{\Pi}^{\text{ext}}\right) = 2$ and (6) has more than one solution. These solutions can be used to build an input sequence with memory which produces a memoryless output and for which

$$I\left(X^n;Y^n\right) = \sum_{i=1}^{n} I\left(X_i;Y_i\right),$$

If $\varepsilon_1 = \frac{1}{3}$, then this example is identical to Example 1. It is interesting to observe that in this case ($\varepsilon_1 = \frac{1}{3}$), $\frac{1}{2}r_1 + \frac{1}{2}r_3 = r_2$.

■

In Example 1, the input X^2 was built using different solutions for (6). It is easy to see that this procedure can also be used to construct a Markovian sequence which produces a memoryless output. This result is presented in the following theorem.

Theorem 1: Let C be a DMC with a matrix of transition probabilities $\mathbf{\Pi}$. If $\mathrm{rank}\left(\mathbf{\Pi}^{\text{ext}}\right) < \mathrm{m_I}$, then there is a Markovian input sequence which produces a memoryless output.

Proof:

Let $\mathbf{q}_i = (q_{1,i}, q_{2,i}, q_{3,i})^T$, $i = 1, 2, 3$ denote three different probability vectors such that $\mathbf{\Pi q}_i = \mathbf{p}_O$. Let $\mathbf{\Pi}_j$ be the j^{th} row of $\mathbf{\Pi}$ and let p_{O_j} denote the j^{th} component of $\mathbf{p}_O$. Therefore $\mathbf{\Pi}_j \mathbf{q}_i = p_{O_j}$.

Let $X_1, \cdots, X_n$ be a Markovian sequence with matrix of conditional probabilities defined by $\mathbf{Q} = \{q_{i,j}\}$, where $q_{i,j} = \Pr\left[\mathrm{X_k = a_i | X_{k-1} = a_j}\right]$.

If the input sequence is such that $\mathbf{p}^{(1)} = \mathbf{q}_1$, then from (7) it can be concluded that

$$\begin{aligned}
\mathbf{p}_O^{(n)} &= \mathbf{\Pi}^{(n)}\mathbf{p}^{(n)}. \\
&= (\underbrace{\mathbf{\Pi} \otimes \mathbf{\Pi} \otimes \cdots \otimes \mathbf{\Pi}}_{n \text{ times}}) \begin{pmatrix} q_{1,1}q_{1,1}\cdots q_{1,1}q_{1,1} \\ q_{1,1}q_{1,1}\cdots q_{1,1}q_{2,1} \\ q_{1,1}q_{1,1}\cdots q_{1,1}q_{3,1} \\ \vdots \\ q_{2,1}q_{1,2}\cdots q_{1,1}q_{1,1} \\ q_{2,1}q_{1,2}\cdots q_{1,1}q_{2,1} \\ \vdots \\ q_{3,1}q_{3,3}\cdots q_{3,3}q_{3,3} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{\Pi}_1\mathbf{q}_1\mathbf{\Pi}_1\mathbf{q}_1\cdots\mathbf{\Pi}_1\mathbf{q}_1\mathbf{\Pi}_1\mathbf{q}_1 \\ \mathbf{\Pi}_1\mathbf{q}_1\mathbf{\Pi}_1\mathbf{q}_1\cdots\mathbf{\Pi}_1\mathbf{q}_1\mathbf{\Pi}_2\mathbf{q}_1 \\ \mathbf{\Pi}_1\mathbf{q}_1\mathbf{\Pi}_1\mathbf{q}_1\cdots\mathbf{\Pi}_1\mathbf{q}_1\mathbf{\Pi}_3\mathbf{q}_1 \\ \vdots \\ \mathbf{\Pi}_2\mathbf{q}_1\mathbf{\Pi}_1\mathbf{q}_2\cdots\mathbf{\Pi}_1\mathbf{q}_1\mathbf{\Pi}_1\mathbf{q}_1 \\ \mathbf{\Pi}_2\mathbf{q}_1\mathbf{\Pi}_1\mathbf{q}_2\cdots\mathbf{\Pi}_1\mathbf{q}_1\mathbf{\Pi}_2\mathbf{q}_1 \\ \vdots \\ \mathbf{\Pi}_3\mathbf{q}_3\mathbf{\Pi}_3\mathbf{q}_3\cdots\mathbf{\Pi}_3\mathbf{q}_3\mathbf{\Pi}_3\mathbf{q}_3 \end{pmatrix} \\
&= \begin{pmatrix} p_{O_1}^n \\ p_{O_1}^{n-1}p_{O_2} \\ p_{O_1}^{n-1}p_{O_3} \\ \vdots \\ p_{O_2}p_{O_1}^{n-1} \\ p_{O_2}p_{O_1}^{n-2}p_{O_2} \\ \vdots \\ p_{O_3}^n \end{pmatrix} \qquad (15)
\end{aligned}$$

From (15) it can be concluded that the output sequence is memoryless.

■

Theorem 1 shows a Markovian input sequence leading a memoryless output in a DMC. Actually Example 1 uses the same procedure of the proof of this Theorem. However, in this Example, it is also shown that the input with memory does not achieve the channel capacity. In the following Section, some results for Markovian inputs are presented.

III. DMC capacity and an input with memory

As mentioned in the Introduction, the inequality presented in (3) plays an important role in the calculus of DMC capacity. Since an input with memory can lead a memoryless output in a DMC, producing an input-output pair such that (3) holds with equality, nothing more natural than ask if an input with memory can achieve the channel capacity. Even though this work does not answer completely this question, this Section shows some interesting results concerning a Markovian input sequence.

In the channel of Example 1, the second row of $\mathbf{\Pi}$ is the arithmetic mean of the first and third rows. The next theorem shows that under this condition, the Markovian input sequence which achieves the channel capacity is not a truly sequence with memory.

Theorem 2: Let C be a DMC where $\mathbf{\Pi}_i$, $i = 1, \cdots, m_O$ are the rows of the matrix of transition probabilities $\mathbf{\Pi}$. Suppose that C has the following properties.

i. $m_I = m_O$,
ii. $\mathrm{rank}\left(\mathbf{\Pi}^{\text{ext}}\right) < \mathrm{m_I}$,
iii. $\mathrm{rank}\left(\mathbf{\Pi}\right) = \mathrm{m_I} - \ell$,
iv. $\mathbf{\Pi}$ has no two identical rows.

Let X^n be a stationary Markovian input sequence which produces a memoryless output with maximum entropy and suppose that X^n has a matrix of conditional probabilities given by

$$\mathbf{Q} = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_{m_I} \end{pmatrix}$$

Then the matrix $\mathbf{Q}$ which maximizes the mutual information has $m_I - \ell$ identical columns and the others columns are related to symbols with null probability.

Proof:

Since the output is memoryless with maximal entropy, the mutual information between the input and the output can be written as

$$\begin{aligned}
\frac{1}{n} I\left(X^n; Y^n\right) &= H\left(Y\right) - H\left(Y|X\right) \\
&= H_{\text{max}} - \sum_{i=1}^{m_I} p_i H(\mathbf{q}_i) \qquad (16)
\end{aligned}$$

where $p_i = \Pr\left[\mathrm{X = a_i}\right]$ and $H(\mathbf{q}_i) = -\sum_{j=1}^{m_I} q_{j,i}\log_2\left(q_{j,i}\right)$. Since $\mathrm{rank}\left(\mathbf{\Pi}\right) = \mathrm{m_I} - \ell$ and there are no two identical rows in $\mathbf{\Pi}$, there are ℓ rows in $\mathbf{\Pi}$ which are means of the others rows. Let L be the set of all rows in $\mathbf{\Pi}$ which can be written as a mean of others rows. By the concavity of the entropy, it can be concluded that $\mathbf{q}_i \in L$, then there is $j \neq i$ such that $\mathbf{q}_j \notin L$ and $H(\mathbf{q}_j) < H(\mathbf{q}_i)$. Therefore the distribution $p_1, \cdots, p_{m_I}$ which maximizes the mutual information is such that $p_i = 0$ for every $\mathbf{q}_i \in L$ (symbols with null probability). Since X^n is stationary,

$$\mathbf{Q} \begin{pmatrix} p_1 \\ \vdots \\ p_{m_I} \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_{m_I} \end{pmatrix} \tag{17}$$

Let $\mathbf{q}_i \in L$. From (17), $\sum_{k=1}^{m_I} q_{i,k} p_k = p_i = 0$. Therefore, in this sum, if $p_k \neq 0$, then $q_{i,k}$ must be null. Thus if p_k is respected to a symbol with non null probability, $q_{i,k}$ must be null. Then in the k^{th} column of $\mathbf{Q}$, $\mathbf{q}_k$ has ℓ zeros — all $q_{i,k}$ such that $\mathbf{q}_i \in L$. However, for every k, $\mathbf{q}_k$ is a solution of (6). Since $\mathrm{rank}\,(\mathbf{\Pi}) = \mathrm{m_I} - \ell$, there is only one solution with ℓ zeros. Thus all $\mathbf{q}_k \notin L$ are the same vector.

■

Theorem 2 shows that for some DMC's the unique Markovian input sequence that can achieve the channel capacity is a sequence such that all states with non null probability has the same conditional probabilities, that is, the sequence is a truly memoryless sequence. This is the case of the channel presented in Example 2 when $\varepsilon_1 \in \left(0, \frac{1}{2}\right)$.

The next theorem, which is the last result of this work, deals with a channel where $\mathbf{\Pi}$ has identical rows.

Theorem 3: Let C be a DMC where $\mathbf{\Pi}$ has two identical rows. Then there is a non memoryless input sequence which achieves the capacity.

Proof:

Actually, the proof of this theorem is quite simple. Let $\mathbf{p} = (p_1, \cdots, p_{m_I})^T$ be an input vector probability which leads to a maximum of $I(X;Y)$. Let p_i and p_j be the probabilities of symbols which produces the identical rows in $\mathbf{\Pi}$. It is easy to see that if a new vector probability $\mathbf{p}'$ such that $p'_i = p_i - \delta$, $p'_j = p_j + \delta$ and $p'_k = p_k$ for all $k \neq i, j$, then $\mathbf{p}'$ produces the *same* output and achieves the channel capacity. Using Theorem 1 these two vector probabilities can be used to build a Markovian input sequence which achieves the capacity.

■

The BSC with null capacity and the channel of Example 2 with $\varepsilon_1 = \frac{1}{2}$ are special cases of channels studied in Theorem 3. It is interesting to notice that the channel of Example 2 has non null capacity. However, it is obvious that two input symbols with the same transition probability distribution do not improve the channel capacity (i.e., one of the input symbol can be discarded with no penalty in capacity). Therefore, in practice Theorems 2 and 3 do not show any useful case of a DMC with input sequence with memory.

IV. Final Remarks

This work was strongly motivated by the problem 6-14 proposed in [7]. Reading this problem, someone who knows that (3) holds with equality iff the output sequence is memoryless can be led to believe that a Discrete Memoryless Channel with a memoryless output must have a memoryless input sequence. In fact, even though it is not valid, this statement sounds very intuitive and it is the error in the proof of theorem 5.2.1 presented in [8]. In this work it was shown some examples of DMC's where an input with memory produces a memoryless output. In this context, an important issue is the relationship between the channel capacity and the maximum mutual information when the input sequence has memory. In the examples presented here, input sequences with memory achieves the channel capacity only in some special cases which are not useful.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal,* vol. 27, pp. 379-423, 623-656, 1948.

[2] T. M. Cover & J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1991.

[3] R. L. Dobrushin, "General formulation of Shannon's main theorem in information theory," *American Mathematical Society Translations,* vol. 33, pp. 323-438, 1963.

[4] S. Verdu & T. S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory,* vol. 40, pp. 1147-1157, 1994.

[5] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons, 1968.

[6] R. J. McEliece, *The Theory of Information and Coding (Encyclopedia of Mathematics and its Applications)*, 2^{nd} Ed., Cambridge University Press, 2002.

[7] J. G. Proakis & M. Salehi, *Digital Communications*, 5^{th} Ed., McGraw-Hill, 2008.

[8] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, 1987.