Two-letter Capacity Formula for Channels with Memory and Feedback

Christos K. Kourtellaris, Ioannis Tzortzis and Charalambos D. Charalambous

***Abstract*— For a class of channels with unit memory on previous channels outputs, we identify necessary and sufficient conditions, to test whether the capacity achieving channel input distributions with feedback are time-invariant, and whether feedback capacity is characterized by a two-letter expression, similar to that of memoryless channels. The method is based on showing that a certain dynamic programming equation, which in general, is a nested optimization problem over the sequence of channel input distributions, reduces to a non-nested optimization problem. We then apply these conditions to derive a two-letter expression for the feedback capacity of the Binary State Symmetric Channel, which is evaluated explicitly. Further, we derive computationally efficient upper bounds on the probability of maximum likelihood decoding error in the finite-blocklength regime.**

I. Introduction

Shannon in his landmark paper [1], showed that the capacity of Discrete Memoryless Channels (DMCs), $\{\mathbf{P}_{B|A}(b|a) : (a,b) \in \mathbb{A} \times \mathbb{B}\}$, is characterized by the two-letter formulae

$$C \stackrel{\triangle}{=} \max_{\mathbf{P}_A} I(A;B). \tag{I.1}$$

This is often shown by using the converse to the channel coding theorem, to obtain the equality

$$\max_{\mathbf{P}_{A^n}} I(A^n;B^n) = (n+1) \max_{\mathbf{P}_{A_i}} I(A_i;B_i) \tag{I.2}$$

where $A^n = (A_0,\ldots,A_n), B^n = (B_0,\ldots,B_n)$. The evaluation of the capacity is drastically simplified due to equality (I.2), since it transforms the nested[1] optimization problem, $\max_{\mathbf{P}_{A^n}} I(A^n;B^n)$, to a non-nested optimization problem $(n+1)\max_{\mathbf{P}_{A_i}} I(A_i;B_i) \equiv (n+1)C$.

For channels with memory, $\mathbf{P}(b_i|b^{i-1},a^i), i = 0,\ldots,n$, with $\mathbf{P}(b_0|b^{-1},a^0) = \mathbf{P}(b_0|b^{-1},a_0)$, i.e., b^{-1} is the initial state, and feedback codes with corresponding distributions $\mathscr{P}^{FB}_{[0,n]} \stackrel{\triangle}{=} \{\mathbf{P}_{A_i|A^{i-1},B^{i-1}} : i = 0,\ldots,n\}$ with $\mathbf{P}(a_0|a^{-1},b^{-1}) = \mathbf{P}_0(a_0|b^{-1})$, the information measure often employed to characterize capacity is the directed information [2]

$$I(A^n \to B^n) \stackrel{\triangle}{=} \sum_{i=0}^{n} I(A^i;B_i|B^{i-1}).$$

It is shown in [2], [3] that feedback capacity is given by

$$C^{FB}_{A^\infty \to B^\infty} = \lim_{n\to\infty} \frac{1}{n+1} C^{FB}_{A^n\to B^n} \tag{I.3}$$

where $C^{FB}_{A^n\to B^n} \stackrel{\triangle}{=} \sup_{\mathscr{P}^{FB}_{[0,n]}} I(A^n \to B^n)$.

In this paper, for certain channels with memory, we show the analog of (I.2) given by

$$\sup_{\mathscr{P}^{FB}_{[0,n]}} I(A^n \to B^n) \stackrel{?}{=} (n+1) \sup_{\mathbf{P}_{A_i|A^{i-1},B^{i-1}}} I(A^i;B_i|B^{i-1}) \tag{I.4}$$

Similar to memoryless channels, when (I.4) holds, then the evaluation of feedback capacity simplifies considerably. We identify necessary and sufficient conditions for a class of channels with memory on the previous channel output, so that the nested optimization problem associated with $C^{FB}_{A^n\to B^n}$, reduces to a non-nested optimization problem, and that (I.4) holds. These conditions imply that the capacity is achieved by a time-invariant channel input distribution, which transforms the expression of the capacity into a two-letter formula, analogous to (I.2). Then, we analyze the Binary State Symmetric Channel (BSSC) [4] (or POST channel [3]), and we show that the above conditions are met, and then we employ them to evaluate explicit expressions both for the capacity and the capacity achieving input distributions. The final expression of the capacity is a two-letter formula, analogous to (I.1). Moreover, we invoke the time-invariant nature of the capacity achieving input distribution and the side information (initial condition) which is available both to encoder and decoder to provide upper bounds on the probability of error for the ML decoder, and to evaluate the performance for finite-blocklength regime.

We chose the BSSC to illustrate the impact of this work, due to the simplicity of calculations. However, the conditions for non-nested optimization also hold for other channels with memory, such as, the symmetric erasure channel with memory [5].

II. Capacity of channels with memory on previous channels outputs

We begin our analysis with some necessary notation and preliminary results on channels with memory on previous channel outputs. The channel input and channel output spaces are finite alphabet spaces denoted by $\mathbb{A}$ and $\mathbb{B}$, respectively.

A. Feedback capacity for channels with arbitrary memory on previous channel outputs

A time-invariant channel with arbitrary memory on previous channel outputs is defined by

$$\mathscr{C}_{[0,n]} \stackrel{\triangle}{=} \{\mathbf{P}(b_i|b^{i-1},a_i) : i = 0,\ldots,n\}. \tag{II.5}$$

C K. Kourtellaris, I. Tzortzis and C. D. Charalambous, are with the Department of Electrical Engineering, University of Cyprus, Nicosia, Cyprus. E-mails: {kourtellaris.christos, tzortzis.ioannis,chadcha}@ucy.ac.cy.

[1]The concepts of nested and non-nested optimization are clarified in Section III.

978-1-5090-3097-2/17/$31.00

At time $i=0$ the conditional distribution is $\mathbf{P}(b_0|b^{-1},a_0)$, where $b^{-1}\in\mathbb{B}$ is the initial condition (or initial state) of the channel, with distribution $\mu(b^{-1})$, that is known both at the encoder and the decoder. The sequence of channel input distributions with feedback, for a general channel with memory, is defined by

$$\mathscr{P}_{[0,n]}^{FB} \triangleq \left\{\Pi_i(a_i|a^{i-1},b^{i-1}) : i=0,\ldots,n\right\}. \tag{II.6}$$

For $i=0$ the convention is $\Pi_0(a_0|a^{-1},b^{-1})=\Pi_0(a_0|b^{-1})$.

Given any channel input distribution, a channel distribution and a fixed initial distribution, the induced joint distribution is uniquely defined by

$$\mathbf{P}_i^{\Pi}(a^i,b^i)=\mu(b^{-1})\Big\{\prod_{j=0}^{i}\big(\mathbf{P}(b_j|b^{j-1},a_j)\Pi_j(a_j|a^{j-1},b^{j-1})\big)\Big\}$$

The conditional distribution of the output is $\mathbf{P}_i^{\Pi}(b_i|b^{i-1})$ and for $i=0$, the distribution is $\mathbf{P}_0^{\Pi}(b_0|b^{-1})$. The superscript Π emphasizes the dependence of the distributions on the channel input distribution. Directed information is given by

$$I(A^n\rightarrow B^n)\triangleq\sum_{i=0}^{n}I(A^i;B_i|B^{i-1})\overset{(\alpha)}{=}\sum_{i=0}^{n}I(A_i;B_i|B^{i-1})$$

where (α) holds due to the structure of the channel. Define the Finite Time Feedback Information (FTFI) quantity by

$$C_{A^n\rightarrow B^n}^{FB}\triangleq\sup_{\mathscr{P}_{[0,n]}^{FB}} I(A^n\rightarrow B^n). \tag{II.7}$$

Then, under the assumption that $\{B^{-1},A_0,B_0,A_1,B_1,\ldots\}$ is jointly ergodic or $\frac{1}{n+1}\sum_{i=0}^{n}\log\frac{\mathbf{P}(B_i|B^{i-1},A_i)}{\mathbf{P}_i^{\Pi}(B_i|B^{i-1})}$ is information stable [6], the feedback capacity is given by

$$C_{A^\infty\rightarrow B^\infty}^{FB}\triangleq\lim_{n\longrightarrow\infty}\frac{1}{n+1}C_{A^n\rightarrow B^n}^{FB}. \tag{II.8}$$

B. Feedback capacity for channels with unit memory on previous channel outputs

A special case of the channels defined by (II.5), is the channel with Unit Memory on previous Channel Output (UMCO), which is defined by

$$\mathscr{C}_{[0,n]}\triangleq\{\mathbf{P}(b_i|b_{i-1},a_i) : i=0,\ldots,n\}. \tag{II.9}$$

At time $i=0$ the conditional distribution is $\mathbf{P}(b_0|b_{-1},a_0)$, where $B_{-1}=b_{-1}\in\mathbb{B}$ is the initial condition. For the UMCO channel, the channel input distribution associated with the maximization of (II.7) satisfies the conditional independence

$$\Pi_i(a_i|a^{i-1},b^{i-1})=\Pi_i(a_i|b_{i-1})\equiv\pi_i(a_i|b_{i-1}) \tag{II.10}$$

where $i=0,1,\ldots,n$ (see [7], [8] for derivation). Consequently, by (II.9), (II.10) we have the Markovian properties

$$\begin{aligned}&\mathbf{P}_i^{\pi}(a_i,b_i|a^{i-1},b^{i-1})=\mathbf{P}_i^{\pi}(a_i,b_i|a_{i-1},b_{i-1}),\\&\mathbf{P}_i^{\pi}(b_i|b_{i-1})=\sum_{a_i}\mathbf{P}(b_i|b_{i-1},a_i)\pi_i(a_i|b_{i-1}).\end{aligned} \tag{II.11}$$

Then (II.7) reduces to

$$\begin{aligned}C_{A^n\rightarrow B^n}^{FB,UMCO}&=\sup_{\overset{\circ}{\mathscr{P}}{}_{[0,n]}^{FB}}\sum_{i=0}^{n}I(A_i;B_i|B_{i-1})\\&\triangleq\sup_{\overset{\circ}{\mathscr{P}}{}_{[0,n]}^{FB}}\mathbf{E}_{\mu}^{\pi}\Big\{\sum_{i=0}^{n}\log\Big(\frac{\mathbf{P}(B_i|B_{i-1},A_i)}{\mathbf{P}_i^{\pi}(B_i|B_{i-1})}\Big)\Big\}\end{aligned} \tag{II.12}$$

where $\overset{\circ}{\mathscr{P}}{}_{[0,n]}^{FB}\triangleq\{\pi_i(a_i|b_{i-1}) : i=0,1,\ldots,n\}\subset\mathscr{P}_{[0,n]}^{FB}$, and the subscript in $\mathbf{E}_{\mu}^{\pi}$ indicates the expectation is for a fixed distribution of the initial state B_{-1}. The subscript, μ, will be omitted when no ambiguity arises.

III. Non-nested optimization of the UMCO channel

In this section we derive necessary and sufficient conditions so that the maximizing distribution in (II.12) is time-invariant. This gives rise to a two-letter expression for feedback capacity.

The dynamic programming recursion for $C_{A^n\rightarrow B^n}^{FB,UMCO}$ is obtained as follows. Let $V_t(b_{t-1})$ represent the value function, that is, the maximum expected total cost on the future time horizon $\{t,t+1,\ldots,n\}$ given output $B_{t-1}=b_{t-1}$ at time $t-1$, defined by

$$V_t(b_{t-1})=\sup_{\pi_i(a_i|b_{i-1}):i=t,t+1,\ldots,n}\mathbf{E}^{\pi}\Big\{\sum_{i=t}^{n}\log\Big(\frac{\mathbf{P}(B_i|B_{i-1},A_i)}{\mathbf{P}_i^{\pi}(B_i|B_{i-1})}\Big)\Big|B_{t-1}=b_{t-1}\Big\}. \tag{III.13}$$

Then (III.13) satisfies the dynamic programming recursions

$$V_n(b_{n-1})=\sup_{\pi_n(a_n|b_{n-1})}\sum_{a_n,b_n}\log\Big(\frac{\mathbf{P}(b_n|b_{n-1},a_n)}{\mathbf{P}_n^{\pi}(b_n|b_{n-1})}\Big)\times\mathbf{P}(b_n|b_{n-1},a_n)\pi_n(a_n|b_{n-1}), \tag{III.14}$$

$$\begin{aligned}V_t(b_{t-1})=\sup_{\pi_t(a_t|b_{t-1})}\Big\{\sum_{a_t}\Big[\sum_{b_t}\log\Big(\frac{\mathbf{P}(b_t|b_{t-1},a_t)}{\mathbf{P}_t^{\pi}(b_t|b_{t-1})}\Big)&\\\times\mathbf{P}(b_t|b_{t-1},a_t)+\sum_{b_t}V_{t+1}(b_t)\mathbf{P}(b_t|b_{t-1},a_t)\Big]&\\\times\pi_t(a_t|b_{t-1})\Big\},\quad t=0,1,\ldots,n-1.&\end{aligned} \tag{III.15}$$

Then, for a fixed initial distribution, $\mu(b_{-1})$, the FTFI capacity is given by

$$C_{A^n\rightarrow B^n}^{FB,UMCO}=\sum_{b_{-1}}V_0(b_{-1})\mu(b_{-1}). \tag{III.16}$$

Since for each time instant t, (i) the future channel input distributions$\{\pi_{t+1}(a_{t+1}|b_t),\ldots,\pi_n(a_n|b_{n-1})\}$ are fixed to their optimal strategies, and (ii) the dynamic programming recursion is concave functional in the input distribution $\pi_t(a_t|b_{t-1})$, the Kuhn-Tucker (KT) conditions hold. Hence, by the KT conditions (upon differentiation) we obtain the following necessary and sufficient conditions [5].

$$V_n(b_{n-1}) = \sum_{b_n} \log\Big(\frac{\mathbf{P}(b_n|a_n,b_{n-1})}{\mathbf{P}^{\pi}_n(b_n|b_{n-1})}\Big)\mathbf{P}(b_n|a_n,b_{n-1}),$$
$$\forall a_n \in \mathbb{A} \quad \text{if} \quad \pi_n(a_n|b_{n-1}) \neq 0, \tag{III.17}$$
$$V_t(b_{t-1}) = \sum_{b_t}\Big\{\log\Big(\frac{\mathbf{P}(b_t|a_t,b_{t-1})}{\mathbf{P}^{\pi}_t(b_t|b_{t-1})}\Big) + V_{t+1}(b_t)\Big\}$$
$$\times \mathbf{P}(b_t|a_t,b_{t-1}), \forall a_t \in \mathbb{A}, \text{ if } \quad \pi_t(a_t|b_{t-1}) \neq 0,$$
$$t = n-1, n-2\ldots,1,0. \tag{III.18}$$

If $\pi_t(a_t|b_{t-1}) = 0$ for some $t = n, n-1\ldots,1,0$, then the corresponding equalities in (III.17) or (III.18) are replaced by inequalities [5].

Before we proceed to our theorem regarding conditions for a two-letter characterization of feedback capacity, we state the definition of non-nested optimization.

Definition 3.1: (Non-nested optimization)
The optimization problem $C^{FB,UMCO}_{A^n\rightarrow B^n}$ defined by (III.16) is called non-nested if and only if the value function, (III.13), satisfies the following identity.

$$V_t(b_{t-1}) = \sum_{i=t}^{n} \sup_{\pi_i(a_i|b_{i-1})} \mathbf{E}^{\pi}\Big\{\log\Big(\frac{\mathbf{P}(B_i|B_{i-1},A_i)}{\mathbf{P}^{\pi}_i(B_i|B_{i-1})}\Big)\Big|B_{t-1} = b_{t-1}\Big\},$$
$$\forall\ (t, b_{t-1}) \in \{0,1,\ldots,n\}\times\mathbb{B}. \tag{III.19}$$

It is clear from the above definition, that if the optimization problem is non-nested, then the optimal channel input distribution that achieves the supremum of (III.19) is time-invariant, that is, $\pi_t(a_t|b_{t-1}) = \pi^{TI}(a_t|b_{t-1}), \forall\ t = n, n-1,\ldots 1,0$. Consequently, the conditional distribution of the output process is also time-invariant, $\mathbf{P}^{\pi}_i(B_i|B_{i-1}) = \mathbf{P}^{\pi^{TI}}(B_i|B_{i-1})$, and the value function, (III.19), satisfies the following identity

$$V_t(b_{t-1}) = (n-t+1) \sup_{\pi^{TI}(a_t|b_{t-1})} \mathbf{E}^{\pi}\Big\{\log\Big(\frac{\mathbf{P}(B_t|B_{t-1},A_t)}{\mathbf{P}^{\pi^{TI}}(B_t|B_{t-1})}\Big)\Big|$$
$$B_{t-1} = b_{t-1}\Big\}, \forall\ (t, b_{t-1}) \in \{0,1,\ldots,n\}\times\mathbb{B}. \tag{III.20}$$

Next, we state the theorem regarding necessary and sufficient conditions for non-nested optimization, which generalizes the non-nested and time-invariant properties of memoryless channels, to channels with memory.

Theorem 3.1: (Necessary and sufficient conditions for non-nested optimization)
The optimization problem $C^{FB,UMCO}_{A^n\rightarrow B^n}$ defined by (III.16) is non-nested if and only if there exists a constant $\overline{V}$ such that

$$V_n(b_{n-1}) = \overline{V}, \quad \forall b_{n-1} \in \mathbb{B} \tag{III.21}$$

where $V_n(b_{n-1})$ satisfies (III.17).

Proof: Suppose (III.21) holds and let

$$R_t(b_{t-1}) = \sum_{b_t\in\mathbb{B}_t} \log\Big(\frac{\mathbf{P}(b_t|a_t,b_{t-1})}{\mathbf{P}^{\pi}_t(b_t|b_{t-1})}\Big)\mathbf{P}(b_t|a_t,b_{t-1})$$
$$\forall a_t \in \mathbb{A} \quad \text{if} \quad \pi_n(a_n|b_{t-1}) \neq 0 \tag{III.22}$$

where $t = n, n-1,\ldots,1,0$. Consider the necessary and sufficient condition given in (III.18) at time $t = n-1$. Since $V_n(b_{n-1}) = \overline{V}, \forall\ b_{n-1}$, then by (III.18) we have

$$V_{n-1}(b_{n-2}) = R_{n-1}(b_{n-2}) + \overline{V}. \tag{III.23}$$

Moreover, since the channel is time-invariant, then

$$R_{n-1}(b_{n-2}) = V_n(b_{n-1}) = \overline{V}. \tag{III.24}$$

Substituting (III.24) to (III.23), we obtain

$$V_{n-1}(b_{n-2}) = 2\overline{V} \tag{III.25}$$

which satisfies (III.20). To complete the derivation, we use induction, that is, we assume validity of (III.20) for $t = n-2, n-3,\ldots,i$ and we show it also holds for $t = i-1$. This is similar to the case $t = n-1$, hence it is omitted. Conversely, if (III.20) holds and (III.21) is violated, then $\overline{V}$ is necessarily a function of b_{n-1}. Thus, at time $t = n-1$ the DP recursion is (III.18), which violates (III.20). ■

The following corollary follows from Theorem 3.1.

Corollary 3.1: If (III.21) holds, then

$$V_t(b_{t-1}) = (n-t+1)\overline{V}, \quad \forall b_{t-1} \in \mathbb{B}. \tag{III.26}$$

Moreover, if (III.21) holds, then by setting $t = 0$ in (III.26) and substituting the result in (III.16), we obtain

$$\begin{aligned} C^{FB,UMCO}_{A^n\rightarrow B^n} &= \sum_{b_{-1}\in\mathbb{B}} V_0(b_{-1})\mu(b_{-1}), \quad \forall b_{-1} \in \mathbb{B} \\ &= (n+1)\overline{V} \sum_{b_{-1}\in\mathbb{B}} \mu(b_{-1}), \quad \forall b_{-1} \in \mathbb{B} \\ &= (n+1)\overline{V}, \quad \forall b_{-1} \in \mathbb{B}. \end{aligned} \tag{III.27}$$

Thus, feedback capacity is given by

$$\begin{aligned} C^{FB,UMCO}_{A^\infty\rightarrow B^\infty} &\triangleq \lim_{n\longrightarrow\infty} \frac{1}{n+1} C^{FB}_{A^n\rightarrow B^n} \\ &= \lim_{n\longrightarrow\infty} \frac{1}{n+1}(n+1)\overline{V} = \overline{V} \\ &\overset{(\alpha)}{=} \sup_{\pi^{TI}(a_n|b_{n-1})} I(A_n;B_n|B_{n-1}) \\ &\overset{(\beta)}{=} \sup_{\pi^{TI}(a_0|b_{-1})} I(A_0;B_0|b_{-1}), \quad \forall b_{-1} \in \mathbb{B} \end{aligned} \tag{III.28}$$

where (α) results from the dynamic programming recursion of $C^{FB,UMCO}_{A^n\rightarrow B^n}$ and the fact that the value function at the terminal time is independent from b_{n-1}, and (β) holds due to the time-invariant property of the channel input distribution, and independence of dynamic programming recursion on b_{-1}. The above result is summarized in the following corollary.

Corollary 3.2: If there exists a $\overline{V}$ which satisfies (III.21), then

$$C^{FB,UMCO}_{A^\infty\rightarrow B^\infty} = \sup_{\pi^{TI}(a_0|b_{-1})} I(A_0;B_0|b_{-1}), \quad \forall b_{-1} \in \mathbb{B}. \tag{III.29}$$

Note that (III.29) is a two letter formula, since $I(A_0;B_0|b_{-1})$ is independent of $b_{-1} \in \mathbb{B}$.

IV. The Binary State Symmetric Channel: Feedback Capacity and Error Exponents

In this section we employ Theorem 3.1, to show that the feedback capacity of the Binary State Symmetric Channel (BSSC) is characterized by (III.29). Then, we compute both the feedback capacity and the corresponding capacity achieving input distribution. Subsequently, we derive an upper bound on the error probability of maximum likelihood decoder which is easy to compute, and we evaluate its performance on the finite-blocklength regime.

A. Feedback capacity

The Binary State Symmetric Channel (BSSC) is defined by

$$\mathbf{P}(b_i|a_i,b_{i-1}) = \begin{array}{c} \\ 0 \\ \\ 1 \end{array}\begin{array}{c} \begin{array}{cccc} 0,0 & 0,1 & 1,0 & 1,1 \end{array} \\ \left[\begin{array}{cccc} \alpha & \beta & 1-\beta & 1-\alpha \\ & & & \\ 1-\alpha & 1-\beta & \beta & \alpha \end{array}\right] \end{array} \quad \text{(IV.30)}$$

where $i=0,1,2,\ldots,n$ and $(\alpha,\beta)\in[0,1]\times[0,1]$. It can be shown that the BSSC is within a transformation equivalent to the Previous Output STate (POST) channel [3], in which the state of the channel is defined as the modulo2 addition of the current input symbol and the previous output symbol. The authors in [3] derived the feedback capacity for the POST channel using asymptotic analysis; however, they did determine the capacity achieving distribution, neither it is recognized this is time-invariant.

The following theorem, compared to [3], [4], gives the following new results. It proves that feedback capacity of the BSSC is reduced to a two-letter expression, and that the capacity achieving input distribution, which is computed explicitly, is time-invariant.

Theorem 4.1: (Feedback capacity of the BSSC)
The feedback capacity of the BSSC, defined by (IV.30), is given by

$$C^{FB,BSSC}_{A^\infty \to B^\infty} \overset{(\alpha)}{=} \max_{\pi^{TI}(a_0|b_{-1})} I(A_0;B_0|B_{-1}=b_{i-1}),\ \forall b_{i-1}\in\{0,1\}$$
$$\overset{(\beta)}{=} H(\lambda)-\nu H(\alpha)-(1-\nu)H(\beta) \quad \text{(IV.31)}$$

where

$$\lambda=\frac{1}{1+2^{\mu}},\ \mu=\frac{H(\beta)-H(\alpha)}{1-\alpha-\beta},\ \nu=\frac{1-(1-\beta)(1+2^{\mu})}{(\alpha+\beta-1)(1+2^{\mu})}.$$

The capacity achieving input distribution is time invariant and is given by

$$\pi^{TI}(a_0|b_{-1}) = \begin{bmatrix} \nu & 1-\nu \\ 1-\nu & \nu \end{bmatrix}. \quad \text{(IV.32)}$$

Proof: The complete proof is given in [9]. Next, we outline the main steps of the proof. Condition (III.21), guaranties the validity of (III.29), which is identical to equation (α). Thus, to prove equation (α) it is sufficient to show that (III.21) holds. The first step involves the evaluation of $V_n(b_{n-1}=0)$ from (III.17). We substitute $b_{n-1}=0, a_n=0$ in (III.17) to obtain the first equation, and $b_{n-1}=0, a_n=1$, to obtain the second equation. Solving the resulted system of equation, we have

$$\mathbf{P}^{\pi}_n(b_n=0|b_{n-1}=0)=\frac{1}{1+2^{\mu}}\equiv\lambda. \quad \text{(IV.33)}$$

Solving (IV.33) with respect to the channel input distribution (i.e. (II.11)) yields

$$\pi_n(a_n=0|b_{n-1}=0)=\nu. \quad \text{(IV.34)}$$

Repeating the same procedure for $V_n(b_{n-1}=1)$, we obtain the following results

$$\mathbf{P}^{\pi}_n(b_n=1|b_{n-1}=1) = \lambda, \quad \text{(IV.35)}$$
$$\pi_n(a_n=1|b_{n-1}=1) = \nu. \quad \text{(IV.36)}$$

Substituting equations (IV.33)-(IV.36) in (III.17), yields (i) $V_n(b_{n-1}=0)=V_n(b_{n-1}=1)=\overline{V}=H(\lambda)-\nu H(\alpha)-(1-\nu)H(\beta)$, thus (α) holds, and (ii) the explicit evaluation of feedback capacity since $C^{FB,BSSC}_{A^\infty\to B^\infty}=\overline{V}$. Moreover, since (III.21) holds, then the capacity achieving input distribution is time invariant and is given by (IV.34) and (IV.36). ■

The above theorem, and in particular its proof, shows that the sufficiency of (III.21) simplifies the evaluation of feedback capacity, since the optimization problem is performed in one-shot, that is, only for the terminal condition.

B. Bounds on the probability of error

Bounds on the probability of error, using ML decoder, for general finite state channels with feedback are given in [10]. The BSSC is a special case of the finite state channels, where the state is the previous channel output, with the additional particularity; at each time instant the current state of the channel is known both to the encoder and the decoder. Recall, that we have already assumed that the initial state of the channel, b_{-1}, is known both to the encoder and the decoder.

In this section, by exploiting the known state at the encoder and the decoder, we expand the results on finite state channels with known state at the decoder [11, Theorem 5.9.3] and the results on finite state channels with feedback [10], and provide computationally efficient bounds of the probability of error for the BSSC. These bounds are evaluated on the capacity achieving channel input distribution. Upper bounds for the general UMCO channel are given in [9].

Consider the BSSC channel and let $\mathbf{P}^{(n)}_{e,m}(b_{-1})$ denote the probability of error for an arbitrary message $m\in\mathscr{M}_n \overset{\triangle}{=} \{1,2,\ldots,M_n=\lfloor 2^{nR}\rfloor\}$, given the initial state $b_{-1}\in\mathbb{B}$. Define the term

$$\Lambda^{\pi^{TI}}(b_i,b_{i-1}) \overset{\triangle}{=} \sum_{b_i}\left[\sum_{a_i}\pi^{TI}(a_i|B_{i-1})\mathbf{P}(b_i|a_i,b_{i-1})^{\frac{1}{1+\rho}}\right]^{1+\rho} \quad \text{(IV.37)}$$

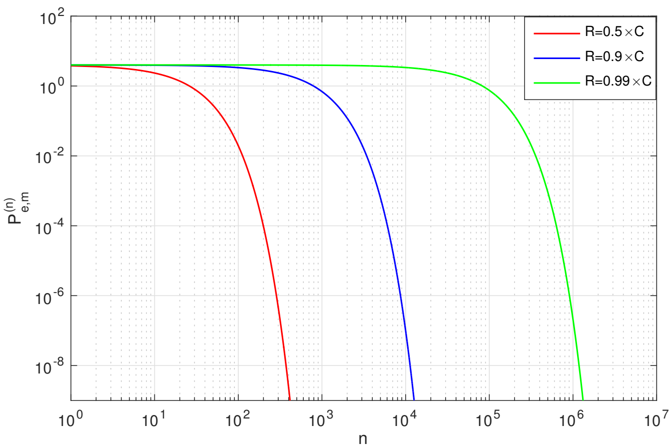

Fig. IV.1: Upper bound on the probability of error for the BSSC with parameters $\alpha = 0.95$, $\beta = 0.85$, where C denotes feedback capacity given by (IV.31).

and let $\left[\Lambda^{\pi^{TI}}(b_i, b_{i-1})\right]$ denote the matrix with elements identified by $\Lambda^{\pi^{TI}}(b_i, b_{i-1})$, $\lambda^{\pi^{TI}}_{max}(\rho)$ the largest eigenvalue of the matrix $\left[\Lambda^{\pi^{TI}}(b_i, b_{i-1})\right]$, and v_{max} and v_{min} the maximum and minimum components, respectively, of the positive eigenvector that corresponds to the largest eigenvalue. Then, the following theorem holds.

Theorem 4.2: For the BSSC, any positive n and any $R > 0$, there exist an (n,R) feedback code, such that for each $m \in \mathscr{M}_n$, each initial condition b_{-1} known both at the encoder and the decoder, and for all $0 \leq \rho \leq 1$, the probability of error is bounded above by

$$\mathbf{P}^{(n)}_{e,m} \leq 4\frac{v_{max}}{v_{min}} 2^{\left\{-n\left[-\rho R - \log \lambda^{\pi^{TI}}_{max}(\rho)\right]\right\}} \tag{IV.38}$$

where $v_{max} = v_{min} = 1$ and

$$\lambda^{\pi^{TI}}_{max}(\rho) = \left[v\alpha^{\frac{1}{1+\rho}} + (1-v)(1-\beta)^{\frac{1}{1+\rho}}\right]^{1+\rho} + \left[v(1-\alpha)^{\frac{1}{1+\rho}} + (1-v)\beta^{\frac{1}{1+\rho}}\right]^{1+\rho} \tag{IV.39}$$

Proof: The analytical proof is given in [9], and is omitted due to space limitation. Roughly, the proof follows Gallagers proof [11, Theorem 5.9.3] and employs the time-invariant nature of capacity achieving input distribution to provide computationally efficient bound of the probability of error evaluated on the capacity achieving input distribution. The expression for $\lambda^{\pi^{TI}}_{max}(\rho)$ is obtained by employing properties of Toeplitz matrices. ■

If the initial state of the channel is not known at the decoder, then (IV.38) is replaced by (see [9, Section IV.A.3] or [11, Problem 5.36])

$$\mathbf{P}^{(n)}_{e,m} \leq 4|\mathbb{B}|\frac{v_{max}}{v_{min}} 2^{\left\{-n\left[-\rho R - \log \lambda^{\pi^{TI}}_{max}(\rho)\right]\right\}}.$$

Moreover, better bounds can be obtained by optimizing with respect to the channel input distribution. However, even for DMC's, the error exponents and the bounds on the probability of error are often evaluated on the capacity achieving distribution. The upper bound of the error probability, (IV.38), for fixed α, β, various values of the rate R, and optimized with respect to ρ, is depicted in Fig. IV.1.

V. Conclusions

This paper gives the necessary and sufficient conditions such that the feedback capacity of channels with memory on previous channel output is characterized by a two-letter expression, similar to that of memoryless channels. We employ these conditions to provide a two-letter formula for the feedback capacity of the BSSC channel, and to show the capacity achieving channel input distribution is time-invariant. Further, we derive bounds on the ML probability of error at the finite-blocklength regime. The adopted approach can be employed for more general channels with memory and feedback.

References

[1] C. E. Shannon, "A mathematical theory on communication," *Bell System Technical Journal*, no. 27, pp. 379–423, October 1948.

[2] J. Massey, "Causality, feedback and directed information," *IEEE International Symposium on Information Theory and its Applicationss*, vol. 72, pp. 303–305, November 2001.

[3] H. Permuter, H. Asnani, and T. Weissman, "Capacity of a post channel with and without feedback," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6041–6057, Oct 2014.

[4] C. K. Kourtellaris and C. D. Charalambous, "Capacity of binary state symmetric channel with and without feedback and transmission cost," in *Information Theory Workshop (ITW), 2015 IEEE*, April 2015, pp. 1–5.

[5] P. A. Stavrou, C. D. Charalambous, and C. K. Kourtellaris, "Sequential necessary and sufficient conditions for capacity achieving distributions of channels with memory and feedback," *IEEE Transactions on Information Theory, accepted for publication, DOI: 10.1109/TIT.2017.2724552*, 2017.

[6] R. L. Dobrushin, "Information transmission in channel with feedback," *Theory of Probability and its Applications*, vol. 3, no. 4, pp. 367–383, 1958.

[7] C. K. Kourtellaris and C. D. Charalambous, "Information structures of capacity achieving distributions for feedback channels with memory and transmission cost: Stochastic optimal control & variational equalities-part i," *arXiv preprint arXiv:1512.04514*, 2015.

[8] ——, "Information structures of capacity achieving distribution for channels with memory and feedback," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1287–1291.

[9] C. K. Kourtellaris, C. D. Charalambous, and I. Tzortzis, "Single letter expression of capacity for a class of channels with memory," *submitted to IEEE Transactions on Information Theory, Paper no. IT-16-0909*, no. arXiv:1701.01007, Nov. 2016. [Online]. Available: http://arxiv.org/pdf/arXiv:1701.01007.pdf

[10] H. Permuter, T. Weissman, and A. Goldsmith, "Capacity of finite-state channels with time-invariant deterministic feedback," in *2006 IEEE International Symposium on Information Theory*, July 2006, pp. 64–68.

[11] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.