A Lower Bound on the Feedback Capacity of Finite-State ISI Channels

Shaohua Yang, Aleksandar Kavčić
Division of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138

***Abstract*— By assuming a Markov-memory input process, we construct a Monte Carlo method for computing a lower bound on the feedback capacity of finite-state inter-symbol interference (ISI) channels. The transition probabilities of the Markov process are allowed to depend on previous observations of the channel outputs, where a suboptimal feedback strategy is chosen to implement the feedback. Generally, the bound is tight for low signal-to-noise ratios (SNRs). At high SNRs the tightness varies depending on the chosen channel. We provide 2 channel examples: one where the bound is not tight at high SNRs and the other where the bound is so tight that it surpasses the tightest known upper bounds on the feedforward capacity.**

I. Introduction

The computation of the capacity of a constrained-input finite-memory channel (such as the partial response channels) is a problem with no known solution. Recently, Monte Carlo methods have been proposed to compute tight lower and upper channel capacity bounds. Arnold and Loeliger [1], and, independently, Pfister, Soriaga and Siegel [2] have devised a method to compute the achievable information rate (i.e., a lower bound on the channel capacity) when a Markov input sequence is transmitted over the channel. In [3] an expectation-maximization method was proposed for optimizing the Markov transition probabilities to achieve a tight lower bound. In [4], an upper bound was constructed using the output distribution obtained by optimizing the lower bound [1]. The gap between the lower bound [1], [2], [3] and the upper bound [4] is often very tight as reported in [4].

Finite-state ISI channels are channels with memory. For channels with memory, feedback can increase the channel capacity [5]. If the channel input is unconstrained, several bounds on the feedback capacity are known [6], [7], [8]. When the channel input is constrained, as for partial response channels (with binary inputs), there are no known results (except those that trivially hold by applying the results for the unconstrained inputs [6], [7], [8]). In this paper we develop a lower bound on the feedback capacity of partial response channels when the input alphabet is binary. As in [3], the method is constructed by assuming the input process is a finite-memory Markov process, but here we allow the transition probabilities to depend on the observations of previous channel outputs. Apart from applications to partial response channels, the method can be used to lower-bound the feedback noisy capacities of constrained (say, run-length limited) codes, see [3] for this particular scenario.

This work was supported by the National Science Foundation under Grant No. CCR-9904458 and by the National Storage Industry Consortium.

Structure: We briefly describe the source/channel model in Section II. In Section III we manipulate the expression for the directed mutual information rate into a form suitable for implementing a feedback strategy. In the same section we formulate our feedback strategy as an optimization problem. In Section IV we derive a solution to the optimization problem and provide an algorithm to compute the resulting directed mutual information rate. Optimized directed mutual information rates for various partial response channels are shown and compared to some known capacity and capacity-bound curves in Section V. Section VI concludes the paper.

Notation: The superscript $^{\mathrm{T}}$ denotes matrix and vector transposition. Random variables are denoted by uppercase letters, while their realizations are denoted by lowercase letters. If a random variable is a member of a random sequence, an index "t" is used to denote time, e.g., X_t. A vector of random variables $[X_i, X_{i+1}, \cdots, X_j]^{\mathrm{T}}$ is shortly denoted by X_i^j, while its realization is shortly denoted by x_i^j. The letter $\mathcal{I}$ denotes the mutual information rate, while the letter h denotes the differential entropy.

II. Source/Channel Model

The source/channel model in this paper is essentially the same as the model used in a previous paper [3], where we considered the channel input to be a time-invariant Markov process. Here we relax the time-invariance constraint and allow the Markov process to be dependent on the past realizations of the channel output process. For the completeness of the text, we briefly describe the source/channel model along the same lines as in [3].

We assume that the source (channel input) is a discrete-time random process X_t whose realizations x_t take values from a finite-size source alphabet $\mathcal{X}$. The channel output process Y_t is statistically dependent on the input, where the realizations y_t may take values from a (possibly continuous) alphabet $\mathcal{Y}$. It is assumed that the channel input process has Markov memory length $L \geq 0$, where the memory may depend on the past realizations of the output process Y_t, i.e., we have for any integer $m \geq 0$

$$\Pr\left(X_{t+1}|X_{t-L-m}^t, Y_1^t\right) = \Pr\left(X_{t+1}|X_{t-L}^t, Y_1^t\right). \quad (1)$$

We consider an indecomposable finite-state machine channel [9]. The channel state at time t is denoted by the random variable S_t whose realization is $s_t \in \mathcal{S} = \{1, 2, \ldots, M\}$. We choose the state alphabet size M to be the minimum integer $M > 0$ such that S_t forms a Markov process of memory 1, i.e., for any integer $m \geq 0$

$$\Pr\left(S_{t+1}|S_{t-m}^{t}, Y_1^t\right) = \Pr\left(S_{t+1}|S_t, Y_1^t\right). \qquad (2)$$

For example, if the channel input X_t is a binary Markov process of memory 3 and the channel is PR4 (i.e., $1 - D^2$) of memory 2, then $M = 2^{\max(3,2)} = 8$ guarantees that the state sequence is a Markov process of memory 1.

With this choice of the states, it is apparent that the input sequence X_t and the state sequence S_t uniquely determine each other. Hence, from this point on, the term Markov process will be reserved for the state sequence S_t which is a Markov process of memory 1. The Markov process transition probabilities (which may change with time because we allow them to be dependent on the past realizations of the output sequence y_1^{t-1}) are denoted by

$$P_{ij}^{(t)} = \Pr\left(S_t = j|S_{t-1} = i, Y_1^{t-1} = y_1^{t-1}\right), \qquad (3)$$

where $1 \leq i \leq M$ and $1 \leq j \leq M$. Clearly, we must have $\sum_{j=1}^{M} P_{ij}^{(t)} = 1$ for any $i \in \mathcal{S}$ and for any realization y_1^{t-1}. A transition from state i to state j is considered to be *invalid* if the Markov state sequence cannot be taken from state i to state j. The transition probability $P_{ij}^{(t)}$ for an invalid transition is thus zero. A *valid* transition is a transition that is not invalid. A trellis section, denoted by $\mathcal{T}$, is the set of all valid transitions. Thus, a valid transition (i, j) satisfies $(i, j) \in \mathcal{T}$.

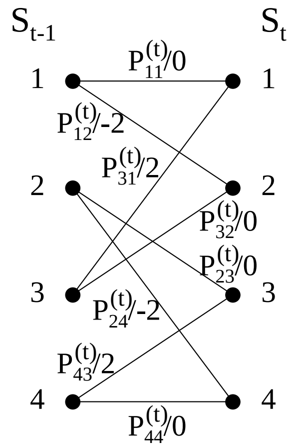
The symbol $P_{ij}^{(t)}$ denotes the (conditional) transition probability of the process S_t entering state $s_t = j$ at time t, when the state at time $t-1$ is $s_{t-1} = i$. To denote the transition probability mass function at time t, we use the symbol $P^{(t)}$. Thus, $P^{(t)}$ denotes a collection of transition probabilities $P_{ij}^{(t)}$ (for $(i, j) \in \mathcal{T}$) which satisfy $\sum_{j=1}^{M} P_{ij}^{(t)} = 1$ for any $i \in \mathcal{S}$.

The channel output Y_t is a *hidden* Markov sequence induced by the state sequence S_t, i.e., for a discrete random variable Y_t, the probability mass function of Y_t satisfies

$$\Pr\left(Y_t|S_{-\infty}^{t}, Y_{-\infty}^{t-1}\right) = \Pr\left(Y_t|S_{t-1}, S_t\right). \qquad (4)$$

If Y_t is a continuous random variable, replace the probability mass functions in (4) by the corresponding probability density functions (pdf's). In this paper, we will adopt the standard partial response channel model, where we assume that the channel output is corrupted by additive white Gaussian noise whose variance is σ^2. Then, the conditional probability mass functions in (4) are replaced by Gaussian conditional pdf's

$$f_{ij}(y_t) = f_{Y_t|S_{t-1},S_t}(y_t|S_{t-1} = i, S_t = j). \qquad (5)$$



$$\mathcal{T} = \left\{ \begin{array}{c} (1,1) \\ (1,2) \\ (2,3) \\ (2,4) \\ (3,1) \\ (3,2) \\ (4,3) \\ (4,4) \end{array} \right\}$$

$$f_{43}(y_t) = \frac{\mathrm{e}^{-\frac{1}{2\sigma^2}(y_t-2)^2}}{\sqrt{2\pi\sigma^2}}$$

Fig. 1. A trellis section for the $1 - D^2$ channel, illustrating the notation used in this paper.

For $(i, j) \in \mathcal{T}$ the Gaussian pdf $f_{ij}(y_t)$ in (5) is parameterized by a mean that depends on the state transition (i, j) and by the variance σ^2. Figure 1 illustrates the notation on the example of the PR4 ($1 - D^2$) channel.

III. Information Rate and Feedback Strategy

A. Directed Mutual information rate

The information rate measure for channels with feedback is the directed mutual information rate $\mathcal{I}(X \to Y)$, see [12]. For finite state machine channels in Section II, the directed information rate is

$$\begin{aligned}
\mathcal{I}(X \to Y) &\overset{\text{(i)}}{=} \lim_{n\to\infty} \frac{1}{n} I\left(X_1^n \to Y_1^n\right) \\
&\overset{\text{(ii)}}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I\left(X_1^t; Y_t|Y_1^{t-1}\right) \\
&\overset{\text{(iii)}}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \left[h\left(Y_t|Y_1^{t-1}\right) - h\left(Y_t|S_1^t, Y_1^{t-1}\right)\right] \\
&\overset{\text{(iv)}}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \left[h\left(Y_t|Y_1^{t-1}\right) - h\left(Y_t|S_{t-1}, S_t\right)\right] \\
&\overset{\text{(v)}}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} h\left(Y_t|Y_1^{t-1}\right) - \frac{1}{2}\log\left(2\pi\mathrm{e}\sigma^2\right), \qquad (6)
\end{aligned}$$

where (i) and (ii) are the definitions of the directed mutual information rate, (iii) is the chain rule of the entropy, (iv) follows from the hidden Markov channel assumption (4), and (v) follows from the assumption that the channel noise is Gaussian with variance σ^2. As was shown in [1], [2], the first term on the right-hand side of (6) can be computed using the normalization coefficients in the forward recursion of the sum-product (BCJR, Baum-Welch) algorithm. Following the notation in [1], we denote the realization of the forward normalization coefficient at time t by λ_t, while we denote the random variable representing this coefficient by Λ_t. As in [1], [2], using Baron's [10] extension of the Shannon-McMillan-Breiman theorem, we have the following sequence of equalities (where $\overset{\text{wp1}}{=}$ denotes convergence with probability 1)

$$\begin{aligned}\lim_{t\to\infty} h\left(Y_t|Y_1^{t-1}\right) &\overset{\text{wp1}}{=} \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\log\lambda_t \\ &= \lim_{t\to\infty}\mathrm{E}\left[\log\Lambda_t\right] \\ &= \lim_{t\to\infty}\mathrm{E}\left(\mathrm{E}\left[\log\Lambda_t|Y_1^{t-1}\right]\right) \\ &\overset{\text{wp1}}{=} \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\left[\log\Lambda_t|y_1^{t-1}\right]. \end{aligned}\tag{7}$$

We next find an expression for $\mathrm{E}\left[\log\Lambda_t|y_1^{t-1}\right]$. Denote the *normalized* α-coefficients of the forward run of the BCJR algorithm at time $t-1$ by (see [11] for details)

$$\alpha_{t-1}(i) = \Pr\left(S_{t-1}=i|y_1^{t-1}\right). \tag{8}$$

The vector collecting all normalized α-coefficients at time $t-1$ is denoted by the α-vector

$$\underline{\alpha}_{t-1} = \left[\alpha_{t-1}(1)\ ,\ \alpha_{t-1}(2)\ ,\ \cdots\ ,\alpha_{t-1}(M)\right]^{\mathrm{T}}. \tag{9}$$

Using the notation introduced in (3), (5) and (8), we may express the conditional pdf of the channel output Y_t, given the past observations y_1^{t-1} as

$$f_{\underline{\alpha}_{t-1}}(y_t) = \sum_{i,j}\alpha_{t-1}(i)\cdot P_{ij}^{(t)}\cdot f_{ij}(y_t). \tag{10}$$

The index $\underline{\alpha}_{t-1}$ in $f_{\underline{\alpha}_{t-1}}(y_t)$ denotes that the pdf is parameterized by the vector $\underline{\alpha}_{t-1}$ in (9). If the channel is corrupted by Gaussian noise, then the conditional pdf's $f_{ij}(y_t)$ are all Gaussian, and the pdf $f_{\underline{\alpha}_{t-1}}(y_t)$ in (10) is a mixture of Gaussians. Further, from the forward recursion of the BCJR algorithm [11], and from the definition of the normalizing constants λ_t in [1], it is straightforward to derive

$$\frac{1}{\lambda_t} = \sum_{i,j}\alpha_{t-1}(i)\cdot P_{ij}^{(t)}\cdot f_{ij}(y_t). \tag{11}$$

Combining (10) and (11), we get

$$\mathrm{E}\left[\log\Lambda_t|y_1^{t-1}\right] = \int_{-\infty}^{\infty} f_{\underline{\alpha}_{t-1}}(y_t)\log\frac{1}{f_{\underline{\alpha}_{t-1}}(y_t)}\,\mathrm{d}y_t. \tag{12}$$

Finally, combining (6), (7) and (12), we may express the directed mutual information rate as

$$\begin{aligned}\mathcal{I}(X\to Y) &\overset{\text{wp1}}{=} \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\left[\log\Lambda_t|y_1^{t-1}\right]-\frac{1}{2}\log\left(2\pi e\sigma^2\right) \\ &\overset{\text{wp1}}{=} \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\infty} f_{\underline{\alpha}_{t-1}}(y_t)\log\frac{\left(2\pi e\sigma^2\right)^{-\frac{1}{2}}}{f_{\underline{\alpha}_{t-1}}(y_t)}\,\mathrm{d}y_t. \end{aligned}\tag{13}$$

B. *Feedback strategy*

We implement the feedback as follows. Upon observing channel output y_1^{t-1} up to time $t-1$, we choose the transition probability mass function $P^{(t)}$ (consisting of individual transition probabilities $P_{ij}^{(t)}$ in (3)) such that the expectation $\mathrm{E}\left[\log\Lambda_t|y_1^{t-1}\right]$ in (12) is maximized, i.e., we choose

$$\begin{aligned}P_{\underline{\alpha}_{t-1}}^{(t)} &= \arg\max_{P^{(t)}}\ \mathrm{E}\left[\log\Lambda_t|y_1^{t-1}\right] \\ &= \arg\max_{P^{(t)}}\int_{-\infty}^{\infty} f_{\underline{\alpha}_{t-1}}(y_t)\log\frac{1}{f_{\underline{\alpha}_{t-1}}(y_t)}\,\mathrm{d}y_t. \end{aligned}\tag{14}$$

This amounts to choosing the parameters $0\le P_{ij}^{(t)}\le 1$ under the constraint $\sum_j P_{ij}^{(t)}=1$ such that the differential entropy of the Gaussian mixture in (12) is maximized. Since the Gaussian mixture pdf $f_{\underline{\alpha}_{t-1}}(y_t)$ in (14) is parameterized by $\underline{\alpha}_{t-1}$, see (10), the result of the optimization problem (14) is a transition probability mass function $P_{\underline{\alpha}_{t-1}}^{(t)}$ that depends on the value of the α-vector $\underline{\alpha}_{t-1}$ in (9). The optimal entropy rate of the output process Y_t is thus obtained by solving the optimization problem (14) at every stage of the forward recursion of the BCJR algorithm [11].

IV. Optimizing the Information Rate

The optimization problem posed in (14) is very simple in comparison to the optimization of time-invariant Markov transition probabilities in [3]. First, we observe that the one-dimensional integral in (14) is simple to evaluate numerically although a closed form solution is not known. Second, we note that the optimization can be solved via a simple gradient search procedure, where the gradient of the objective function equals

$$\frac{\partial\,\mathrm{E}\left[\log\Lambda_t|y_1^{t-1}\right]}{\partial P_{ij}^{(t)}} = \int_{-\infty}^{\infty}\alpha_{t-1}(i)f_{ij}(y_t)\log\frac{1}{f_{\underline{\alpha}_{t-1}}(y_t)}\,\mathrm{d}y_t. \tag{15}$$

The one-dimensional integral in (15) can also very easily be evaluated numerically.

The following result guarantees that the gradient search will converge to the global maximum of the objective function $\mathrm{E}\left[\log\Lambda_t|y_1^{t-1}\right]$ in (12).

Proposition 1: The objective function $\mathrm{E}\left[\log\Lambda_t|y_1^{t-1}\right]$ is concave over the convex set $\left\{P^{(t)}\right\}$ of transition probability mass functions.

Proof: Let $\dot{P}^{(t)}$ and $\bar{P}^{(t)}$ be two transition probability mass functions defined on the trellis $\mathcal{T}$. For any pair $(\dot{\mu},\bar{\mu})$ such that $0\le\dot{\mu}\le 1$, $0\le\bar{\mu}\le 1$ and $\dot{\mu}+\bar{\mu}=1$, we have that $P^{(t)}=\dot{\mu}\dot{P}^{(t)}+\bar{\mu}\bar{P}^{(t)}$ is also a valid transition probability mass function. Therefore the set $\left\{P^{(t)}\right\}$ of transition probability mass functions is convex. Next we note that the function $x\log\frac{1}{x}$ is concave over x, i.e.,

$$\dot{\mu}\left(\dot{x}\log\frac{1}{\dot{x}}\right)+\bar{\mu}\left(\bar{x}\log\frac{1}{\bar{x}}\right)\le(\dot{\mu}\dot{x}+\bar{\mu}\bar{x})\log\frac{1}{\dot{\mu}\dot{x}+\bar{\mu}\bar{x}}.$$

Substituting $\dot{x} = \sum_{i,j} \alpha_{t-1}(i) \cdot \dot{P}_{ij}^{(t)} \cdot f_{ij}(y_t)$ and $\bar{x} = \sum_{i,j} \alpha_{t-1}(i) \cdot \bar{P}_{ij}^{(t)} \cdot f_{ij}(y_t)$ proves the proposition. ■

We can now summarize the method for optimizing the information rate based on maximizing the objective function in (14).

Algorithm 1 FOR OPTIMIZING FEEDBACK-DEPENDENT MARKOV PROCESS TRANSITION PROBABILITIES

Initialization: 1) $\alpha_0(1)=1$, and $\alpha_0(k)=0$ for $1<k\leq M$.
2) Set the starting state to $s_0 = 1$.
3) Pick n to be a large positive integer

Repeat for $1 \leq t \leq n$

Step 1: For the conditional pdf defined in (10), solve the optimization problem in (14) via a gradient search technique, to find $P_{ij}^{(t)}$ for all $(i,j)\in\mathcal{T}$.

Step 2: Given $S_{t-1} = s_{t-1}$, create a realization s_t of S_t according to the conditional transition probabilities $P_{ij}^{(t)}$ found in Step 1.

Step 3: Create a noisy channel output y_t according to the pdf $f_{Y_t|S_{t-1},S_t}(y_t|S_{t-1} = s_{t-1}, S_t = s_t)$.

Step 4: Perform one forward-recursion step of the BCJR (Baum-Welch) algorithm [11] to obtain $\alpha_t(i)$ for all $1 \leq i \leq M$.

Step 5: Compute λ_t using (11).

end

At the end of the execution of Algorithm 1, using (6) and (7), we estimate the directed information rate as

$$\hat{\mathcal{I}}\left(X \to Y\right) = \frac{1}{n}\sum_{t=1}^{n} \log \lambda_t - \frac{1}{2}\log\left(2\pi e \sigma^2\right). \tag{16}$$

The estimate in (16) converges with probability 1 to the directed mutual information rate $\mathcal{I}\left(X \to Y\right)$, which is a lower bound on the feedback capacity, established by the following theorem (which is an adaptation of the feedback capacity theorem due to Tatikonda in [12], p. 93)

Theorem 1: All rates below the rate computed by Algorithm 1 are achievable with a feedback code.

Proof: Since the input sequence X_1^n and the state sequence S_1^n uniquely determine each other, in the following we interchangeably use the symbol X for the symbol S. By Theorem 4.4.1 in [12], the feedback capacity of a channel is determined by the supremum of the directed mutual information rate $I\left(X \to Y\right)$, where the supremum is taken over all possible conditional channel input distributions $\Pr\left(X_t|X_1^{t-1} = x_1^{t-1}, Y_1^{t-1} = y_1^{t-1}\right)$. By construction, the choice of the feedback-dependent transition probability mass functions $P_{ij}^{(t)}$ in (14) indeed creates one such input distribution. The only obstacle preventing us from applying Tatikonda's Theorem 4.4.1 in [12] is that the theorem applies to cases where both the input and output alphabets are finite. We take care of this difficulty by constructing a sequence of quantizers of the output sequence, for which the number of quantization levels grows very large while the quantization error approaches zero. Thus, the rate computed by Algorithm 1 is the supremum of all rates achievable by quantizing the output to a finite number of quantization levels. ■

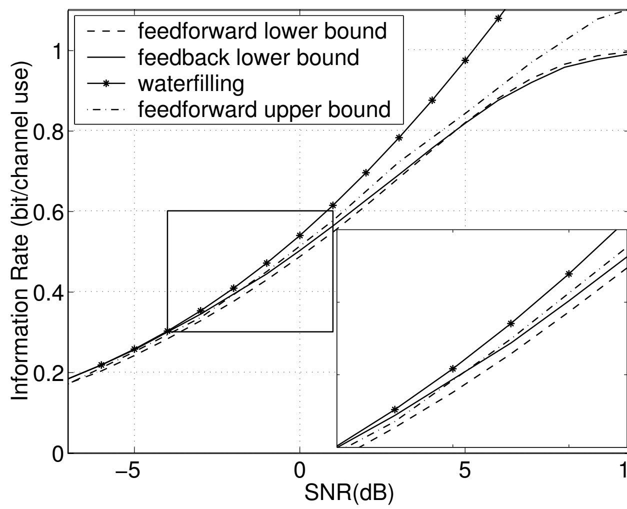


Fig. 2. Capacity bounds for the dicode $(1-D)$ channel. At low SNRs, the feedback capacity bound beats the lower bound on the feedforward capacity calculated by optimizing the time-invariant transition probabilities [3]. At low SNRs, the feedback capacity bound also surpasses the tightest known upper bound on the feedforward capacity [4] and it meets the waterfilling channel capacity for non-constrained channel inputs. At high SNRs, the feedback capacity bound is not tight as it is below the feedforward lower bound [3].

V. Optimization Results

We show the feedback capacity bounds obtained by executing Algorithm 1. Our numerical evaluations have shown that increasing the Markov memory of the input does not increase the feedback capacity bound. Therefore, we show bounds obtained using only inputs whose Markov memory matches the channel memory.

In Figures 2 and 3 we show the feedback capacity lower bounds for the dicode $(1-D)$ channel and the $1-D/2$ channel, respectively. We chose these two channels for easy comparison to the Vontober-Arnold upper bound on the feedforward capacity [4]. An interesting feature of the feedback capacity lower bound is that at low SNRs the bound seems to be tight. In particular, for the dicode $(1-D)$ channel at low SNRs, the feedback capacity lower bound just meets the waterfilling capacity (which is the feedforward channel capacity when the input sequence is not constrained [5], [9]), see Figure 2, while for the $1-D/2$ channel at low SNRs the bound even surpasses the waterfilling capacity, see Figure 3. Furthermore, for the $1-D/2$ channel, the computed lower bound on the feedback capacity is higher than the tightest known upper bound for the feedforward

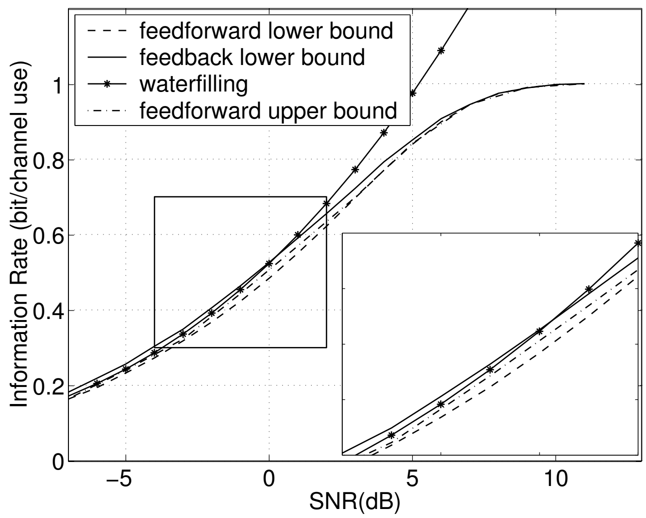


Fig. 3. Capacity bounds for the $1 - D/2$ partial response channel. For this channel the feedback capacity lower bound surpasses the Vontobel-Arnold upper bound on the feedforward capacity [4] for all SNRs. At low SNRs, the feedback capacity bound even surpasses the waterfilling feedforward channel capacity for non-constrained channel inputs.

capacity [4], see Figure 3. This demonstrates that (at least for the $1 - D/2$ channel) the rates achievable with feedback codes are higher than the rates achievable without feedback.

Given that for some channels (e.g, the dicode channel, Figure 2) the bound is not tight at high SNRs, it would be interesting to see whether it can be tightened with a different feedback optimization approach. In this paper we chose an optimization method that maximizes each term in (7) individually. This, however, is not a guarantee that the sum in (7) will be maximized because the terms are causally dependent. Constructing a feedback strategy to jointly maximize all terms in (7) remains an open problem.

VI. Conclusion

We presented a technique for lower bounding the feedback capacity of finite-state machine channels. The method is based on minimizing the term-by-term differential entropy of the channel output, given the observations of previous channel outputs and assuming that the input process has a finite Markov memory. The optimized transition probabilities are dependent on the α-coefficients computed using the BCJR algorithm [11]. There is no guarantee that this method achieves the feedback capacity of the channel (in fact, for some channels the bound is loose at high SNRs). However, the bound computed in this manner seems to be tight at low SNRs and for some channels tight for all SNRs. It was also observed that for some channels at low SNRs the bound surpasses the waterfilling feedforward capacity.

References

[1] D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," in *Proceedings IEEE International Conference on Communications 2001*, (Helsinki, Finland), June 2001.

[2] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proceedings IEEE Global communications Conference 2001*, (San Antonio, Texas), November 2001.

[3] A. Kavčić, "On the capacity of Markov sources over noisy channels," in *Proceedings IEEE Global Communications Conference 2001*, (San Antonio, Texas), November 2001. available at http://hrl.harvard.edu/~kavcic/publications.html.

[4] P. Vontobel and D. M. Arnold, "An upper bound on the capacity of channels with memory and contraint input," in presented at *IEEE Information Theory Workshop*, (Cairns, Australia), September 2001.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons, 1991.

[6] S. Butman, "A general formulation of linear feedback communications systems with solutions," *IEEE Trans. Inform. Theory*, vol. 15, pp. 392–400, May 1969.

[7] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inform. Theory*, vol. 35, pp. 37–43, January 1989.

[8] L. H. Ozarow, "Upper bounds on the capacity of Gaussian channels with feedback," *IEEE Trans. Inform. Theory*, vol. 36, pp. 156–161, January 1990.

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

[10] A. R. Barron, "The strong ergodit theorem for densities: Generalized Shannon-McMillan-Breiman theorem," *Ann. Probab.*, vol. 13, pp. 1292–1303, November 1985.

[11] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Sept. 1974.

[12] S. C. Tatikonda, *Control Under Communications Constraints*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, September 2000. available to dwnload at http://www.cs.berkeley.edu/~tatikond/thesis.html.