[15] ——, "An upper bound to the capacity of discrete time Gaussian channel with feedback—II," *IEEE Trans. Inform. Theory*, vol. 40, pp. 588–593, Mar. 1994.
[16] ——, "An upper bound to the capacity of discrete time Gaussian channel with feedback—III," *Bull. Kyushu Inst. Tech., Pure and Applied Mathematics*, vol. 45, pp. 1–8, 1998.

The Capacity-Cost Function of Discrete Additive Noise Channels With and Without Feedback

Fady Alajaji, *Member, IEEE,* and Nicholas Whalen

***Abstract*—We consider modulo-q additive noise channels, where the noise process is a stationary irreducible and aperiodic Markov chain of order k. We begin by investigating the capacity-cost function $(C(\beta))$ of such additive-noise channels without feedback. We establish a tight upper bound to $(C(\beta))$ which holds for general (not necessarily Markovian) stationary q-ary noise processes. This bound constitutes the counterpart of the Wyner–Ziv lower bound to the rate-distortion function of stationary sources with memory. We also provide two simple lower bounds to $C(\beta)$ which along with the upper bound can be easily calculated using the Blahut algorithm for the computation of channel capacity. Numerical results indicate that these bounds form a tight envelope on $C(\beta)$.**

We next examine the effect of output feedback on the capacity-cost function of these channels and establish a lower bound to the capacity-cost function with feedback $(C_{\rm FB}(\beta))$. We show (both analytically and numerically) that for a particular feedback encoding strategy and a class of Markov noise sources, the lower bound to $C_{\rm FB}(\beta)$ is strictly greater than $C(\beta)$. This demonstrates that feedback can increase the capacity-cost function of discrete channels with memory.

***Index Terms*—Additive noise, capacity-cost function, channels with feedback, channels with memory, cost constraints.**

I. Introduction

In this work, we analyze the capacity of discrete (discrete-time finite-alphabet) channels with memory subject to an input cost constraint. More specifically, we consider modulo-q additive noise channels, where the noise process is a stationary irreducible and aperiodic Markov chain of order k.

Mod-q additive noise channels are symmetric; by this we mean that the block mutual information between input and output vectors of such channels is maximized for equiprobable input blocks (uniform independent and identically distributed (i.i.d.) input process). A closed-form solution exists for the capacity of such channels [1]. However, since additive channels with memory become nonsymmetric under input constraints, a closed-form expression for their capacity-cost function does not exist. This indicates the necessity to establish bounds to the channel capacity-cost function.

Insight into how and where to look for these bounds comes from Shannon, who first commented on the duality between a source and a channel [17]:

> There is a curious and provocative duality between the properties of a source with a distortion measure and those of a channel. This duality is enhanced if we consider channels in which there is a cost associated with the different input letters, and it is desired to find the capacity subject to the constraint that the expected cost not exceed a certain quantity [...]. The solution of this problem leads to a capacity cost function for the channel [...] this function is concave downward [$\cdots$]. In a somewhat dual way, evaluating the rate distortion function $R(D)$ for a source amounts, mathematically, to minimizing a mutual information [...] with a linear inequality constraint [...]. The solution leads to a function $R(D)$ which is convex downward [...]. This duality can be pursued further and is related to a duality between past and future and the notions of control and knowledge. Thus we may have knowledge of the past but cannot control it; we may control the future but not have knowledge of it.

Equipped with Shannon's illuminating observation, we begin by investigating the capacity-cost function $(C(\beta))$ of mod-q additive-noise channels without feedback. We derive a tight upper bound to $C(\beta)$ which holds for all discrete channels with stationary additive noise. This bound constitutes the counterpart of the Wyner–Ziv lower bound to the rate-distortion function of stationary sources. We also provide two simple lower bounds to $C(\beta)$ which along with the upper bound can be easily calculated using the Blahut algorithm for the computation of channel capacity. Numerical results indicate that these bounds form a tight envelope on $C(\beta)$.

We next study the capacity-cost function of the mod-q additive channels with feedback. We establish a lower bound to the capacity-cost function with feedback $(C_{\rm FB}(\beta))$ and introduce a feedback encoding strategy and a class of Markov noise sources for which the lower bound to $C_{\rm FB}(\beta)$ is strictly greater than $C(\beta)$. This is demonstrated both analytically and numerically.

The rest of this correspondence is organized as follows. In Section II, we define the capacity-cost function and present its properties. The analysis of the nonfeedback and feedback capacity-cost functions of additive noise channels is given in Sections III and IV, respectively. A summary is stated in Section V.

II. Preliminaries: The Capacity-Cost Function

Consider a discrete channel with finite-input alphabet $\mathcal{X}$, finite-output alphabet $\mathcal{Y}$, and n-fold transition probability

$$Q^{(n)}(y_1, y_2, \cdots, y_n \mid x_1, x_2, \cdots, x_n), \qquad x_i \in \mathcal{X},\ \ y_i \in \mathcal{Y},\ \ i = 1, 2, \cdots, n.$$

In general, the use of the channel is not free; we associate with each input letter x a nonnegative number $b(x)$, that we call the "cost" of x. The function $b(\cdot)$ is called the cost function. If we use the channel n consecutive times—i.e., we send an input vector $x^n = (x_1, x_2, \cdots, x_n)$—the cost associated with this input vector is "additive":

$$b(x^n) = \sum_{i=1}^{n} b(x_i).$$

For an input process $\{X_i\}_{i=1}^{\infty}$ with block input distribution

$$P_{X^n}(x^n) = \Pr(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$$

the *average cost* for sending X^n is defined by

$$E[b(X^n)] = \sum_{x^n} P_{X^n}(x^n) b(x^n) = \sum_{i=1}^{n} E[b(X_i)].$$

Manuscript received October 21, 1998; revised July 4, 1999. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant OGP0183645. The material in this correspondence was presented in part at the CISS, Princeton, NJ, March 1994.

The authors are with Mathematics and Engineering, Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada.

Communicated by S. Shamai, Associate Editor for Shannon Theory.
Publisher Item Identifier S 0018-9448(00)03099-6.

Definition 1: An n-dimensional input random vector $X^n = (X_1, X_2, \cdots, X_n)$ that satisfies

$$\frac{1}{n}E[b(X^n)] \le \beta$$

is called a β*-admissible* input vector. We denote the set of n-dimensional β*-admissible* input distributions by $\tau_n(\beta)$

$$\tau_n(\beta) = \left\{P_{X^n}(x^n) : \frac{1}{n}E[b(X^n)] \le \beta\right\}.$$

The capacity-cost function of discrete stationary[1] channels with memory is defined by [13]

$$C(\beta) = \sup_n C_n(\beta) = \lim_{n\to\infty} C_n(\beta) \tag{1}$$

where $C_n(\beta)$ is the n*th capacity-cost function* given by

$$C_n(\beta) \triangleq \max_{P_{X^n}(x^n)\in\tau_n(\beta)} \frac{1}{n}I(X^n;Y^n) \tag{2}$$

where $I(X^n;Y^n)$ is the block mutual information between the input vector X^n and the output vector Y^n.

The capacity-cost function $C(\beta)$ has an *operational* significance for channels satisfying certain regularity conditions (e.g., a stationary ergodic channel, a discrete channel with stationary and ergodic additive noise, or an information-stable channel [8], [18]). More specifically, $C(\beta)$ represents the supremum of all rates R for which there exist sequences of β-admissible block codes[2] with vanishing probability of error as n grows to infinity. In other words, $C(\beta)$ is the maximum amount of information that can be transmitted reliably over the channel, if the channel must be used in such a way that the average cost does not exceed β. If $b(x) = 0$ for every letter $x \in \mathcal{X}$, then $C(\beta)$ is just the channel *capacity* C as we know it.

$$C = \sup_n C_n = \lim_{n\to\infty} C_n, \tag{3}$$

where

$$C_n \triangleq \max_{P_{X^n}(x^n)} \frac{1}{n}I(X^n;Y^n). \tag{4}$$

In this work, we exclusively consider discrete channels with stationary ergodic mod-q additive noise. Since these channels have no input memory or anticipation [1], it follows from [8, Lemma 12.4.3] that, without loss of generality, we can restrict the maximizations in (2) and (4) over the set of n-dimensional distributions of stationary ergodic input processes. With this fact in mind, we next state the properties of $C(\beta)$ for the class of discrete channels with stationary ergodic additive noise. We first define, respectively, $\beta_{\min}$, $\beta_{\max}^{(n)}$, and $\beta_{\max}$ by

$$\beta_{\min} \triangleq \min_{x\in\mathcal{X}} b(x),$$

$$\beta_{\max}^{(n)} \triangleq \min\left\{\frac{1}{n}E\left[b(X^n)\right] : \frac{1}{n}I(X^n;Y^n) = C_n\right\}$$

and

$$\beta_{\max} \triangleq \min\left\{\lim_{n\to\infty}\frac{1}{n}E\left[b(X^n)\right] : \lim_{n\to\infty}\frac{1}{n}I(X^n;Y^n) = C\right\}.$$

[1]A discrete channel is said to be stationary if for every stationary input process, the joint input-output process is stationary.

[2]A (nonfeedback) channel block code of length n over $\mathcal{X}$ is a subset

$$\mathcal{C} = \left\{c_{(1)}, c_{(2)}, \cdots, c_{(|\mathcal{C}|)}\right\}$$

of $\mathcal{X}^n$ where each $c_{(i)}$ is an n-tuple. The rate of the code is $R = \frac{1}{n}\log_2|\mathcal{C}|$. The code is β-admissible if $b(c_{(i)}^n) \le n\beta$ for $i = 1, 2, \cdots, |\mathcal{C}|$. If the encoder wants to transmit message W where W is uniform over $\{1, 2, \cdots, |\mathcal{C}|\}$, it sends the codeword $c_{(W)}$. At the channel output, the decoder receives Y^n and chooses as estimate of the message $\hat{W} = g(Y^n)$, where $g(\cdot)$ is a decoding rule. The (average) probability of decoding error is then $P_e^{(n)} = \Pr\{g(Y^n) \ne W\}$.

From the definition of $\beta_{\min}$ above, we can see that $\frac{1}{n}E[b(X^n)] \ge \beta_{\min}$; therefore, $C(\beta)$ is defined only for $\beta \ge \beta_{\min}$. Furthermore, by the additivity of the cost function and the stationarity of the capacity-achieving input distribution, we have that $\beta_{\max} = \lim_{n\to\infty}\beta_{\max}^{(n)}$.

Remark: For a discrete channel with mod-q additive noise $\{Z_n\}$ and linear cost constraints on the input—i.e. $b(x) = x$—we get that $\beta_{\min} = 0$, $\beta_{\max}^{(n)} = \beta_{\max} = (q-1)/2$, $C(\beta_{\min}) = 0$

$$C_n(\beta_{\max}^{(n)}) = C_n = \log_2 q - (1/n)H(Z^n)$$

and

$$C(\beta_{\max}) = C = \log_2 q - H(Z_\infty)$$

where $H(Z_\infty)$ is the noise entropy rate.

Lemma 1 [13]: The nth capacity-cost function $C_n(\beta)$ given by (2) is *concave* and *strictly increasing* in β for $\beta_{\min} \le \beta < \beta_{\max}^{(n)}$ and is equal to C_n for $\beta \ge \beta_{\max}^{(n)}$.

From the above lemma, the fact that the limit of a concave function is concave, and the definition of $\beta_{\max}$, we deduce the following result.

Lemma 2: The capacity-cost function $C(\beta)$ given by (1) is *concave* and *strictly increasing* in β for $\beta_{\min} < \beta < \beta_{\max}$, and is equal to C for $\beta \ge \beta_{\max}$.

III. Additive-Noise Channels Without Feedback

A. An Upper Bound to the Capacity-Cost Function

We consider a discrete channel with memory, with common input, noise, and output q-ary alphabet $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathcal{A}_q = \{0, 1, \cdots, q-1\}$ and described by $Y_n = X_n \oplus Z_n$, for $n = 1, 2, 3, \cdots$ where

- $\oplus$ represents the addition operation modulo q;
- the random variables X_n, Z_n, and Y_n are, respectively, the input, noise, and output of the channel;
- $\{X_n\} \perp \{Z_n\}$; i.e., the input and noise sequences are independent of each other;
- the noise process $\{Z_n\}_{n=1}^{\infty}$ is stationary.

We now turn to the analysis of the capacity-cost function $C(\beta)$ of this channel. Since the input achieving $C(\beta)$ is nonsymmetric for $\beta < \beta_{\max}$, the formula of $C(\beta)$ given by (1) will not have a closed-form expression. We will then try to derive an upper bound to $C(\beta)$.

In [21], Wyner and Ziv derived a lower bound to the rate-distortion function $(R(D))$ of stationary sources

$$R(D) \ge R_1(D) - \mu_1$$

where

- $R_1(D)$ is the rate-distortion function of an "associated" memoryless source with distribution equal to the marginal distribution $P^{(1)}(\cdot)$ of the stationary source;
- $\mu_1 \triangleq H(X_1) - H(X_\infty)$, is the amount of memory in the source. $H(X_1)$ is the entropy of the associated memoryless source with distribution $P^{(1)}(\cdot)$, and $H(X_\infty)$ is the entropy rate of the original stationary source.

This lower bound was later tightened by Berger [2]

$$R(D) \ge R_n(D) - \mu_n \ge R_1(D) - \mu_1 \tag{5}$$

where $R_n(D)$ is the nth rate-distortion function of the source, $R_1(D)$ is as defined above, and $\mu_n = (1/n)H(X^n) - H(X_\infty)$.

In light of the striking duality that exists between $R(D)$ and $C(\beta)$, we prove an equivalent upper bound to the capacity-cost function of a discrete additive-noise channel.

Theorem 1: Consider a discrete channel with additive stationary noise $\{Z_n\}$. Let $Q^{(n)}(\cdot)$ denote the n-fold probability distribution of the noise process. Then for $N = kn$ where $k, n = 1, 2, \cdots$, we have

$$C_N(\beta) \le C_n(\beta) + \Delta_{nN} \le C_1(\beta) + \Delta_{1N} \tag{6}$$

where

- $C_n(\beta)$ is the n-fold capacity-cost function of the channel as defined in (2);
- $C_1(\beta)$ is the capacity-cost function of the associated discrete memoryless channel (DMC) with i.i.d. additive noise process whose distribution is equal to the marginal distribution $Q^{(1)}(\cdot)$ of the stationary noise process;
- $$\Delta_{nN} \triangleq (1/n)H(Z^n) - (1/N)H(Z^N)$$
  with $Z^i = (Z_1, Z_2, \cdots, Z_i)$, $i = n$, or N, and
  $$\Delta_{1N} = H(Z_1) - (1/N)H(Z^N)$$
  where $H(Z_1)$ is the entropy of the i.i.d. noise process of the associated DMC.

Proof: The proof employs a dual generalization of Wyner and Ziv's proof of the lower bound to the rate-distortion function. We first need to use the expression

$$I(X^N; Y^N) \le \sum_{i=1}^{k} I\left(X_{(i)}^n; Y_{(i)}^n\right) + N\Delta_{nN} \tag{7}$$

where

$$\begin{aligned} X^N &= (X_{(1)}^n, X_{(2)}^n, \cdots, X_{(k)}^n) \\ Y^N &= (Y_{(1)}^n, Y_{(2)}^n, \cdots, Y_{(k)}^n) \\ X_{(i)}^n &= (X_{1,(i)}, X_{2,(i)}, \cdots, X_{n,(i)}) \end{aligned}$$

and

$$Y_{(i)}^n = (Y_{1,(i)}, Y_{2,(i)}, \cdots, Y_{n,(i)})$$

with $X_{l,(i)}$ and $Y_{l,(i)}$ denoting the lth component of the vectors $X_{(i)}^n$ and $Y_{(i)}^n$, respectively. Proving the above inequality goes as follows:

$$\begin{aligned} &\sum_{i=1}^{k} I\left(X_{(i)}^n; Y_{(i)}^n\right) + N\Delta_{nN} - I(X^N; Y^N) \\ &\quad = \sum_{i=1}^{k} \left[H\left(Y_{(i)}^n\right) - H\left(Y_{(i)}^n \mid X_{(i)}^n\right)\right] + \frac{N}{n} H(Z^n) \\ &\qquad - H(Z^N) - H(Y^N) + H(Y^N \mid X^N) \\ &\quad = \sum_{i=1}^{k} \left[H\left(Y_{(i)}^n\right) - H\left(Z_{(i)}^n\right)\right] + kH(Z^n) \\ &\qquad - H(Z^N) - H(Y^N) + H(Z^N) \\ &\quad = \sum_{i=1}^{k} H\left(Y_{(i)}^n\right) - H(Y^N) = \sum_{i=1}^{k} H\left(Y_{(i)}^n\right) \\ &\qquad - \sum_{i=1}^{k} H\left(Y_{(i)}^n \mid Y_{(i-1)}^n, Y_{(i-2)}^n, \cdots, Y_{(1)}^n\right) \\ &\quad \ge \sum_{i=1}^{k} H\left(Y_{(i)}^n\right) - \sum_{i=1}^{k} H\left(Y_{(i)}^n\right) = 0 \end{aligned}$$

where the third equality follows from the stationarity of the noise, and the last inequality follows from the fact that conditioning decreases entropy.

We now proceed to prove (6). Let $P_{X^N}(x^N) \in \tau_N(\beta)$ where $\tau_N(\beta)$ is described in Definition 1. For this input distribution, we denote $\beta_i \triangleq \frac{1}{n} E[b(X_{(i)}^n)]$ for $i = 1, 2, \cdots, k$; thus $\frac{1}{k}\sum_{i=1}^{k} \beta_i \le \beta$. By (7), we obtain with this $P_{X^n}(x^N)$ that

$$\frac{1}{N} I(X^N; Y^N) \le \frac{1}{N} \sum_{i=1}^{k} I\left(X_{(i)}^n; Y_{(i)}^n\right) + \Delta_{nN}.$$

But $\frac{1}{n} I(X_{(i)}^n; Y_{(i)}^n) \le C_n(\beta_i)$ for $i = 1, 2, \cdots, k$; thus

$$\frac{1}{N} I(X^N; Y^N) \le \frac{1}{k} \sum_{i=1}^{k} C_n(\beta_i) + \Delta_{nN}.$$

By concavity of $C_n(\cdot)$, we have

$$\frac{1}{k} \sum_{i=1}^{k} C_n(\beta_i) \le C_n\left(\frac{1}{k} \sum_{i=1}^{k} \beta_i\right)$$

and since $C_n(\cdot)$ is strictly increasing we have that

$$C_n\left(\frac{1}{k} \sum_{i=1}^{k} \beta_i\right) \le C_n(\beta).$$

Therefore,

$$\frac{1}{N} I(X^N; Y^N) \le C_n(\beta) + \Delta_{nN}$$

or

$$\max_{P^{(N)}(X^N) \in \tau_N(\beta)} \frac{1}{N} I(X^N; Y^N) = C_N(\beta) \le C_n(\beta) + \Delta_{nN}.$$

Thus the first inequality in (6) is proved. To prove the second inequality in (6), we need to show that $C_n(\beta) \le C_1(\beta) + \Delta_{1n}$ or $C_k(\beta) \le C_1(\beta) + \Delta_{1k}$. This is shown using the first inequality in (6) and letting $n = 1$. □

Using (6) and (1), we obtain the following tight upper bound on $C(\beta)$.

Corollary 1: Consider the channel described in Theorem 1. Then

$$C(\beta) \le C_n(\beta) + M_n \le C_1(\beta) + M_1 \tag{8}$$

where

- $C_n(\beta)$ and $C_1(\beta)$ are as defined in Theorem 1;
- $M_n \triangleq \Delta_{n\infty} = (1/n)H(Z^n) - H(Z_\infty)$, and $M_1 \triangleq \Delta_{1\infty} = H(Z_1) - H(Z_\infty)$ denotes the amount of memory in the noise process.

The bound given above is asymptotically tight with n since as $n \to \infty$, $M_n \to 0$.

Observation: We remark that the above bound given by (8) holds also for real-valued additive noise channels (i.e., for $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ being subsets of the real line) with the modulo addition operation replaced by regular addition and entropy replaced by differential entropy.

B. *Existing Lower Bounds to the Capacity-Cost Function*

Lower bounds on the capacity-cost function can be formed in two ways. First, the nth capacity-cost function provides a simple lower bound to $C(\beta)$: $C(\beta) = \sup_{n \ge 1} C_n(\beta) \ge C_n(\beta)$. Blahut's algorithm [3] is ideally suited for its computation using the $q^n \times q^n$ channel transition matrix Q, the probability of receiving Y^n given that X^n was transmitted.

Another lower bound exists for the capacity-cost function when the channel alphabet is binary. If we take the inputs to be i.i.d. and $q = 2$, we can apply Mrs. Gerber's Lemma [15] to obtain a lower bound

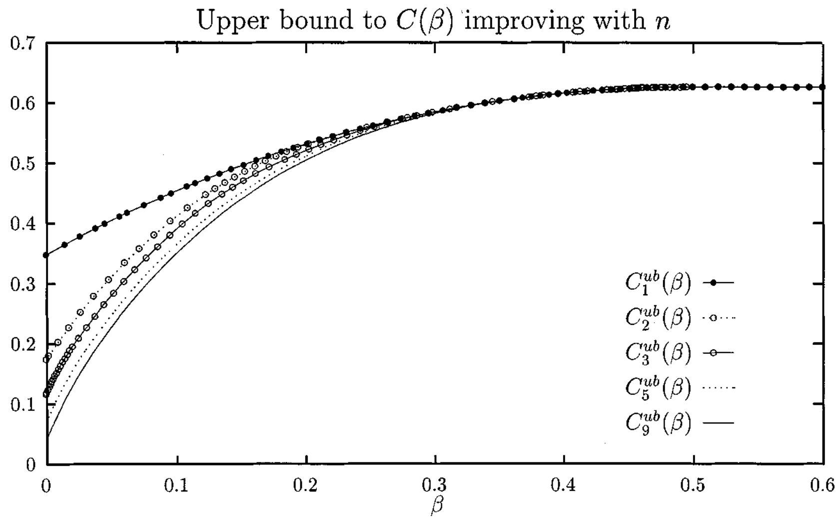

Fig. 1. $C_n^{\rm ub}(\beta)$ for $n = 1, 2, 3, 5$, and 9 for a binary channel with first-order Markov noise given by $\mathbf{\Pi}_2$. $\beta_{\max} = 1/2$.

on $C(\beta)$. More specifically, let $P(X_i = 1) \triangleq \alpha$ be the marginal distribution of an i.i.d. input process such that $E[b(X_i)] = \beta$, then

$$C(\beta) \geq h_b\left(\tilde{\alpha} * h_b^{-1}(\lambda)\right) - H(Z_\infty) \tag{9}$$

where $h_b(\cdot)$ is the binary entropy function, $a * b \triangleq a(1-b) + (1-a)b$, $\tilde{\alpha} \triangleq \min\{\alpha, 1-\alpha\}$, and $\lambda \triangleq \min\{H(Z_\infty), 1 - H(Z_\infty)\}$.

As we are dealing primarily with q-ary channels, we use the $C_n(\beta)$ lower bound in all cases except the binary case where we also apply Mrs. Gerber's bound. The $C_n(\beta)$ lower bound is in fact dual to the upper bound on $R(D)$ computed by Blahut in [3].

C. Numerical Results for Markov Noise Sources

We have thus far derived an upper bound on the capacity-cost function (Corollary 1)

$$C(\beta) \leq C_n(\beta) + M_n \triangleq C_n^{\rm ub}(\beta).$$

This bound becomes tight as $n \to \infty$. Furthermore, a simple lower bound to the capacity-cost function is given by

$$C_n^L(\beta) \triangleq C_n(\beta) \leq C(\beta).$$

We herein estimate $C(\beta)$ numerically for the case where the noise process is a stationary irreducible and aperiodic Markov chain of order k. This is accomplished by computing $C_n(\beta)$ and $C_n(\beta) + M_n$. $C_n(\beta)$ is calculated via Blahut's algorithm for the computation of the capacity-cost function [3], while M_n comes from a straightforward computation of the noise entropy rate

$$\begin{aligned} M_n &= \frac{1}{n} H(Z_1, \cdots, Z_n) - H(Z_{k+1} \,|\, Z_k, \cdots, Z_1) \\ &= \frac{1}{n}[H(Z_1, \cdots, Z_k) - kH(Z_{k+1} \,|\, Z_k, \cdots, Z_1)]. \end{aligned} \tag{10}$$

The above entropies are computed using the stationary distribution vector $\boldsymbol{\pi}$ which is obtained by solving $\boldsymbol{\pi}\mathbf{\Pi} = \boldsymbol{\pi}$, where $\mathbf{\Pi}$ is the state transition matrix for the kth-order Markov noise.

We will hereafter assume that the cost function $b(\cdot)$ is given by $b(x) = x$; i.e., we will impose a linear cost constraint on the channel input letters. In the numerical examples, we consider the cases $q = 2$ and $q = 3$.

$C_n(\beta)$ is, in fact, the capacity-cost function of a discrete memoryless channel whose input and output alphabets are the sets of words of length n and whose transition probabilities are given by the n-fold probability distributions of the process $\{Z_i\}$. Using the algorithm of [3, Theorem 10], we calculate $C_n(\beta)$ and M_n for different values of $\mathbf{\Pi}$, n, and alphabet size q. The results, computed to an accuracy of 10^{-6} bits are plotted in Figs. 1–4. We have used $n = 9$ for the binary channels (see Figs. 1, 2, and 4), and $n = 5$ for the ternary channels (see Fig. 3). Tighter results can be achieved for larger n. However, the tightness improves as $1/n$ since from (10) we have that $M_n = O(1/n)$, while the computation complexity increases exponentially in n and q.

In Figs. 1 and 2, the following channel transition and stationary probabilities were employed:

$$\mathbf{\Pi}_2 = \begin{bmatrix} 0.95 & 0.05 \\ 0.2 & 0.8 \end{bmatrix} \qquad \boldsymbol{\pi} = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix}.$$

The ternary example in Fig. 3 is computed using

$$\mathbf{\Pi}_3 = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.3 & 0.5 & 0.2 \\ 0.3 & 0.1 & 0.6 \end{bmatrix} \qquad \boldsymbol{\pi} = \begin{bmatrix} 0.6 \\ 0.2167 \\ 0.1833 \end{bmatrix}$$

while in Fig. 4, we form the envelope on the capacity-cost function of a channel with second-order binary noise described by

$$\mathbf{\Pi}_2^{(2)} = \begin{bmatrix} 0.95 & 0.05 & 0 & 0 \\ 0 & 0 & 0.25 & 0.75 \\ 0.90 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \end{bmatrix} \qquad \boldsymbol{\pi} = \begin{bmatrix} 0.655 \\ 0.036 \\ 0.036 \\ 0.273 \end{bmatrix}.$$

Note that $C_n^L(0) = 0 = C(0)$ while $C_n^{\rm ub}(\beta_{\max}) = C(\beta_{\max})$. This shows that for the extreme values of β, at least one of the bounds will achieve the "true" capacity-cost function. Furthermore, the difference between the two bounds, which is equal to M_n, vanishes as $(1/n)$.

In Figs. 2 and 4, we also include Mrs. Gerber's lower bound given in (9). While it is obviously very weak for low-cost (β), as the per-letter cost approaches $\beta_{\max}$, this bound outperforms Blahut's bound for computable block lengths n. The reason for the good performance of Mrs. Gerber's bound for values of β close to $\beta_{\max}$ comes from the fact that the capacity of binary Markov channels, which is given by

$$C = 1 - H(Z_\infty)$$

is achieved at $\beta_{\max}$ by an i.i.d. uniformly distributed input process. This makes Mrs. Gerber's bound tight at $\beta_{\max}$ (the right-hand side of (9) with $\alpha = (1/2)$ is equal to C).

IV. Additive-Noise Channels with Feedback

We have already addressed the estimation of the capacity-cost function $C(\beta)$ for q-ary additive noise channels without feedback by establishing upper and lower bounds to $C(\beta)$. We herein investigate the

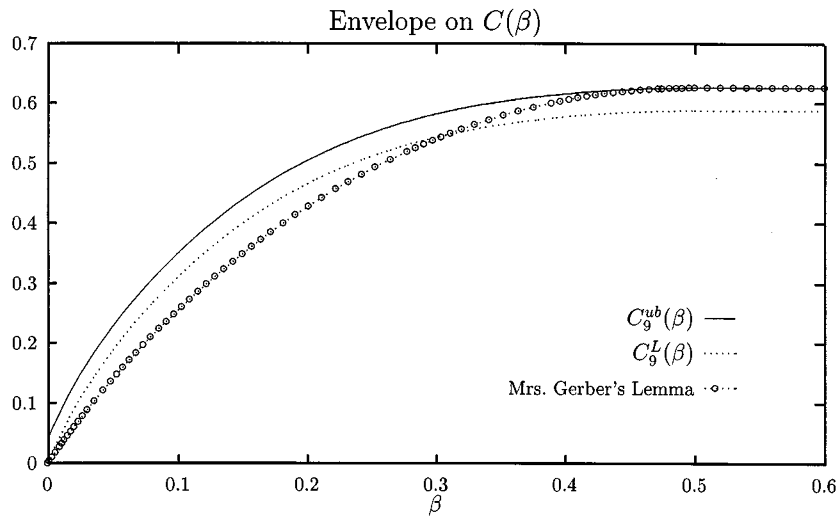

Fig. 2. Comparison of $C_9^{\mathrm{ub}}(\beta)$ with $C_9^L(\beta)$ and Mrs. Gerber's lower bound for a binary channel with first-order Markov noise given by $\mathbf{\Pi}_2$. $\beta_{\max} = 1/2$.

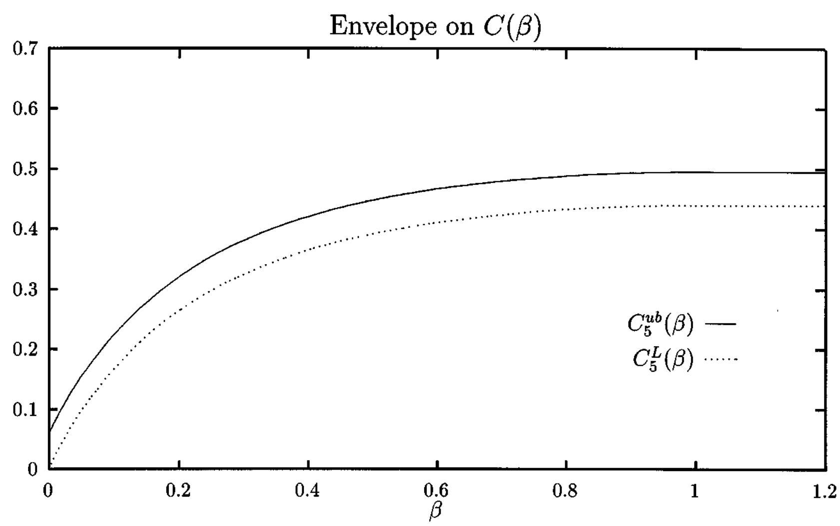

Fig. 3. Comparison of $C_5^{\mathrm{ub}}(\beta)$ with $C_5^L(\beta)$ for a ternary channel with first-order Markov noise given by $\mathbf{\Pi}_3$. $\beta_{\max} = 1$.

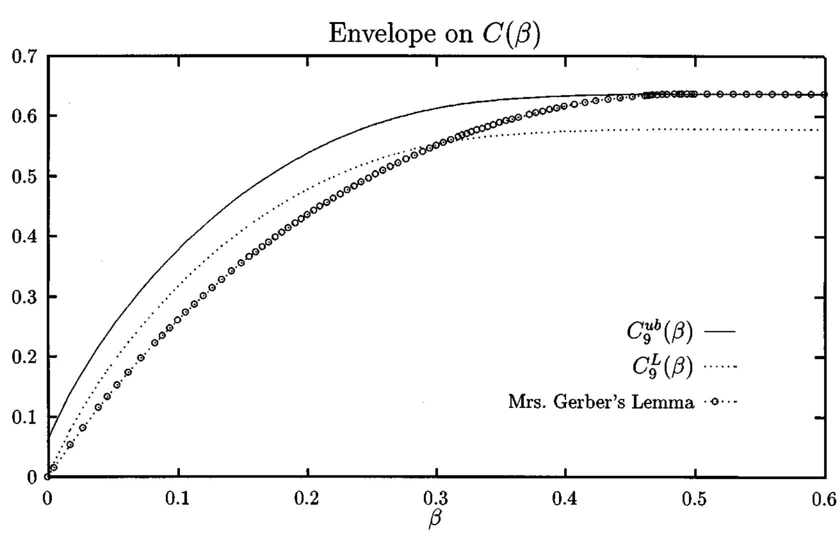

Fig. 4. Comparison of $C_9^{\mathrm{ub}}(\beta)$ with $C_9^L(\beta)$ and Mrs. Gerber's lower bound for a binary channel with second-order Markov noise given by $\mathbf{\Pi}_2^{(2)}$. $\beta_{\max} = .5$.

effect of output feedback on the capacity-cost function of such channels for the case where the noise process consists of a stationary irreducible and aperiodic (hence ergodic) Markov chain of order k. By "output feedback" we mean that there exists a "return channel" from the receiver to the transmitter; we assume this return channel is noiseless, delayless, and has a large capacity. The receiver uses the return channel to inform the transmitter what letters were actually received; these letters are received at the transmitter before the next letter is transmitted and, therefore, can be used in choosing the next transmitted letter.

In previous related work, Shannon first proved in [16] that feedback does not increase the capacity of discrete memoryless channels (DMC's). He also conjectured that feedback can increase the capacity of channels with memory.

In a seemingly different work [11], [12], Jelinek investigated the capacity of finite-state indecomposable channels with side information at the transmitter. In particular, he showed that the capacity of state-computable finite-state Markovian indecomposable channels with (modulo) additive noise, where the noise is a deterministic function $\psi(\cdot)$ of the state of a Markov source, is not increased with the availability of side information at the transmitter [12, Theorem 8]. This result has an interesting connection with the problem of feedback for channels with additive Markov noise. Specifically, if the $\psi(\cdot)$ function is the identity function, then the current state corresponds to the previous noise sample and the problem of side information at the transmitter reduces to the problem of feedback for a channel with additive Markov noise. Thus [12, Theorem 8] implies that feedback does not increase the capacity of channels with (modulo) additive ergodic Markov noise. In [1], Alajaji demonstrated that for channels with (modulo) additive noise, where the noise is an *arbitrary* (not necessarily stationary nor ergodic) process, feedback does not increase capacity. Recently, Erez and Zamir established simple expressions for the capacity of discrete (modulo) additive-noise channels with causal and noncausal side information at the transmitter [7].

For continuous-amplitude channels, Pinsker [14] and Ebert [6] showed that feedback at most doubles the capacity of nonwhite additive Gaussian noise channels. No increase is possible for additive white Gaussian noise channels. In [4], Cover and Pombra also proved that feedback increases the capacity of nonwhite Gaussian noise channels by at most half a bit. Ihara and Yanagi [9], [10], [22] provided general (sufficient and necessary) conditions on the noise and average power of nonwhite Gaussian channels under which the capacity is increased by feedback. Finally, in [19], Viswanathan obtained a computable expression for the capacity of finite-state Markov channels with perfect channel state information at the receiver and delayed feedback. He also applied his result to derive the capacity of finite-state additive Gaussian as well as log-normal shadow fading channels.

A. A Lower Bound to the Feedback Capacity-Cost Function

A feedback channel block code with block length n and rate R consists of the following.

- An index set $\{1, 2, \cdots, 2^{nR}\}$ on the messages W.
- A sequence of encoding functions $f_i: \{1, 2, \cdots, 2^{nR}\} \times \mathcal{Y}^{i-1} \rightarrow \mathcal{X}$ for $i = 1, 2, \cdots, n$.
- A decoding function, $g: \mathcal{Y}^n \rightarrow \{1, 2, \cdots, 2^{nR}\}$, which is a deterministic rule assigning an estimate $\hat{W}$ to each output vector.

To convey message $W \in \{1, 2, \cdots, 2^{nR}\}$, the user sends the codeword $X^n = (X_1, X_2, \cdots, X_n)$, where $X_i = f_i(W, Y_1, Y_2, \cdots, Y_{i-1})$ for $i = 1, 2, \cdots, n$. The decoder receives $Y^n = (Y_1, Y_2, \cdots, Y_n)$ and guesses the original message to be $g(Y^n)$. A decoder error occurs if $g(Y^n) \neq W$. We assume that W is uniformly distributed over $\{1, 2, \cdots, 2^{nR}\}$. The probability of decoding error is then given by

$$P_e^{(n)} = \Pr\{g(Y^n) \neq W\}.$$

Since we are studying the capacity-cost function, we require an average cost constraint on the channel input code letters X_i's. We say that a feedback rate R is *achievable* if there exists a sequence of β-admissible (as defined in Footnote 2) feedback codes with block length n and rate R such that $\lim_{n\rightarrow\infty} P_e^{(n)} = 0$. The supremum of all achievable feedback code rates is then the channel capacity-cost function with feedback, denoted by $C_{\rm FB}(\beta)$.

Since the channel is additive, we have $Y_i = X_i \oplus Z_i$ where $\{Z_i\}$ is a q-ary stationary irreducible and aperiodic Markov chain of order k. We assume that W and $\{Z_i\}$ are independent of each other. Note, however, that because of the feedback, X^n and Z^n are no longer independent; X_i may depend recursively on Z^{i-1}.

With output feedback, the encoder is informed at time instant i about all the previously received output symbols $Y_1, Y_2, \cdots, Y_{i-1}$; and thus knows all the previous channel noise samples $Z_1, Z_2, \cdots, Z_{i-1}$, $i = 1, 2, \cdots$. Note also that for a finite memory system of order k, the feedback of terms more than k time steps old provides no new information. Therefore, we can express the feedback function in terms of the input components and the noise state as

$$X_i = f_i(W, Z_{i-k}, \cdots, Z_{i-1}). \tag{11}$$

In general, the feedback rule $f_i(W, Z_{i-k}, \cdots, Z_{i-1})$ is time-varying. In this work, we obtain a lower bound to $C_{\rm FB}(\beta)$ by only focusing on time invariant feedback strategies.

For this $\bmod q$ channel with feedback, we define $C^{\rm lb}(\beta)$ using a *fixed* encoding rule f^* as

$$C^{\rm lb}(\beta) = \sup_n C_n^{\rm lb}(\beta) = \lim_{n\rightarrow\infty} C_n^{\rm lb}(\beta) \tag{12}$$

where

$$C_n^{\rm lb}(\beta) = \max_{P_{V^n}(v^n)\in\tilde{\tau}_n(\beta)} \frac{1}{n} I(V^n; Y^n) \tag{13}$$

where $Y_i = X_i \oplus Z_i$, $X_i = f^*(V_i, Z_{i-k}, \cdots, Z_{i-1})$ for $i = 1, \cdots, n$, V^n is a q-ary n-tuple independent of Z^n, and

$$\tilde{\tau}_n(\beta) \triangleq \left\{P_{V^n}(v^n) : \frac{1}{n} E[b(X^n)] \leq \beta\right\}. \tag{14}$$

Observe that the cost constraint is imposed on the feedback vector X^n rather than V^n. We next state without proving the following result. Its proof, which employs the usual random coding argument and the asymptotic equipartition property, is given in [20].

Theorem 2 (Achievability of $C^{\rm lb}(\beta)$: $C_{\rm FB}(\beta) \geq C^{\rm lb}(\beta)$): Consider a q-ary kth-order additive Markov noise channel defined above with a fixed time-invariant feedback function f^*. If $C_n^{\rm lb}(\beta)$ is as defined in (12), then there exists a sequence of β-admissible feedback codes of block length n and rate R such that $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$ for all rates $R < C^{\rm lb}(\beta)$.

B. Nonlinear Feedback for Which $C_{\rm FB}(\beta) > C(\beta)$

We next introduce a simple nonlinear feedback scheme and a class of noise processes for which feedback increases $C(\beta)$. For a channel with q-ary kth-order additive Markov noise, let $V^n(W) = (V_1, \cdots, V_n)$ be a q-ary n-tuple representing message $W \in \{1, 2, \cdots, 2^{nR}\}$. Then, to transmit W, the encoder sends $X^n(W) = (X_1, X_2, \cdots, X_n)$, where

$X_i = f^*(V_i, Y_1, Y_2, \cdots, Y_{i-1})$ for $i = 1, 2, \cdots, n$ and the time-invariant feedback encoding function $f^*(\cdot)$ is given as follows:

$$X_i = V_i \quad \text{if } i \le k; X_i = f^*(V_i, S_i) \triangleq \begin{cases} V_i, & S_i \ne \tilde{s} \\ 0, & S_i = \tilde{s} \end{cases} \quad \text{if } i > k \tag{15}$$

where

$$S_i \triangleq (Z_{i-k}, Z_{i-k+1}, \cdots, Z_{i-1})$$

denotes the state of the noise process at time i, and $\tilde{s}$ is some preselected state. Note that in the above coding scheme, V^n is nothing but the "nonfeedback" codeword; that is, if the channel is without feedback, then $X^n(W) = V^n$.

Under linear or power cost constraints this feedback strategy asks the transmitter to monitor the noise state S. If the encoder detects a particular *bad* state $\tilde{s}$ (i.e., one whose transition probabilities are nearly uniform) at step i, then the transmitter is instructed to send the least expensive word regardless of the current message symbol V_i. In our examples the least expensive letter has $b(0) = 0$.

Let us now apply this feedback strategy to a q-ary channel with a particular additive Markov noise of order k.

Lemma 3: Consider a q-ary channel with stationary irreducible and aperiodic additive Markov noise of order k with the feedback rule given in (15) relating X_i, V_i, and S_i. If for a particular noise state $\tilde{s}$ the conditional probabilities of the current noise sample are uniformly distributed; i.e.,

$$P_{Z_i \mid S_i}(z_i \mid \tilde{s}) = \frac{1}{q}, \qquad \forall z_i \in \mathcal{A}_q$$

then the conditional probabilities of y^n given v^n are *equal* for both the feedback and nonfeedback channels

$$P_{Y^n \mid V^n}(y^n \mid v^n) = P^{\text{FB}}_{Y^n \mid V^n}(y^n \mid v^n), \qquad \text{for all } y^n, v^n \in \mathcal{A}_q^n.$$

Proof: The transition probabilities for the nonfeedback channel are given by

$$\begin{aligned} P_{Y^n \mid V^n}(y^n \mid v^n) &= P_{Z^n}(z^n = y^n \ominus v^n) \\ &= P_{Z^k}(y_1 \ominus v_1, y_2 \ominus v_2, \cdots, y_k \ominus v_k) \\ &\quad \times \prod_{i=k+1}^{n} P_{Z_i \mid S_i}(y_i \ominus v_i \mid s_i) \end{aligned} \tag{16}$$

where $s_i = (z_{i-k}, z_{i-k+1}, \cdots, z_{i-1})$ is the state of the Markov chain at step i for a given input–output pair (v^n, y^n). Using the same notation but with a superscript to denote feedback, the transition probabilities of the feedback channel are given by

$$\begin{aligned} P^{\text{FB}}_{Y^n \mid V^n}(y^n \mid v^n) &= P_{Z^k}(y_1 \ominus v_1, y_2 \ominus v_2, \cdots, y_k \ominus v_k) \\ &\quad \times \prod_{i=k+1}^{n} P^*_{Z_i \mid S_i}(y_i \ominus f^*(v_i, s_i) \mid s_i) \end{aligned} \tag{17}$$

where

$$P^*_{Z_i \mid S_i}(y_i \ominus f^*(v_i, s_i) \mid s_i) = \begin{cases} P_{Z_i \mid S_i}(y_i \ominus v_i \mid s_i), & \text{if } s_i \ne \tilde{s} \\ P_{Z_i \mid S_i}(y_i \mid s_i), & \text{if } s_i = \tilde{s}. \end{cases}$$

Notice that (16) and (17) are identical except possibly when noise state $\tilde{s}$ occurs. But $P_{Z \mid S}(z \mid \tilde{s}) = (1/q)$ for all $z \in \{0, 1, \cdots, q-1\}$; this implies that

$$P_{Z_i \mid S_i}(y_i \mid \tilde{s}) = P_{Z_i \mid S_i}(y_i \ominus v_i \mid \tilde{s}) = \frac{1}{q}. \tag{18}$$

Therefore,

$$P_{Y^n \mid V^n}(y^n \mid v^n) = P^{\text{FB}}_{Y^n \mid V^n}(y^n \mid v^n)$$

for the feedback encoding scheme in (15) if the conditional probabilities of Z_i given $S_i = \tilde{s}$ are uniform. □

Lemma 3 implies that since the nonfeedback channel is symmetric, the feedback channel is also symmetric. From [5, Theorem 8.2.1] we can infer that a uniform distribution on the input blocks V^n induces a uniform distribution on the output blocks Y^n.

We have so far shown that for a particular type of Markov noise sources, our feedback rule has no effect on the channel conditional distribution. It does, however, affect the *cost* of individual input blocks. The following lemma compares the expected cost of nonfeedback channel inputs with the cost of feedback channel inputs encoded using our strategy.

Lemma 4: Consider the nonfeedback and feedback channels described above, with the feedback strategy given in (15), and $P_{Z \mid S}(z \mid \tilde{s}) = \frac{1}{q}$ for all z. Let $P^*_{V^n}(v^n)$ be a stationary input distribution that achieves $C_n(\beta)$ for $\beta > \beta_{\min}$. Then

$$C_n^{\text{lb}}\left(\beta_n^{\text{lb}}\right) \ge C_n(\beta)$$

where β_n^{lb} is the expected per letter cost under $P^*_{V^n}(v^n)$ and the feedback encoding strategy, and is given by

$$\beta_n^{\text{lb}} = \left[1 - \frac{n-k}{n} P_S(\tilde{s})\right] \beta.$$

Proof: For the nonfeedback channel

$$\beta = \frac{1}{n} \sum_{v^n} P^*_{V^n}(v^n) b(v^n) = \sum_{v} P^*_{V_1}(v) b(v)$$

since $P^*_{V^n}(v^n)$ is a stationary input distribution that achieves the nonfeedback capacity-cost function $C_n(\beta)$. For the feedback channel we charge costs to the channel input letters after applying the feedback rule f^*. Thus

$$\begin{aligned} \beta_n^{\text{lb}} &= \frac{1}{n} E\left[b(X^n)\right] = \frac{1}{n} \sum_{i=1}^{n} E\left[b(X_i)\right] \\ &= \frac{1}{n} \sum_{i=1}^{k} \sum_{v_i} P_{V_i}(v_i) b(V_i) \\ &\quad + \frac{1}{n} \sum_{i=k+1}^{n} E[b(f^*(V_i, S_i))] \\ &= \frac{k}{n}\beta + \frac{n-k}{n} \sum_{v} \sum_{s} P_S(s) P_V(v) b(f^*(v, s)) \\ &= \frac{k}{n}\beta + \frac{n-k}{n} \left[\sum_{v} P_S(\tilde{s}) P_V(v) b(0) \right. \\ &\qquad \left. + \sum_{s \ne \tilde{s}} \sum_{v} P_S(s) P_V(v) b(v) \right] \\ &= \frac{k}{n}\beta + \frac{n-k}{n} \sum_{s \ne \tilde{s}} P_S(s) \beta \\ &= \left[1 - \frac{n-k}{n} P_S(\tilde{s})\right] \beta \end{aligned} \tag{19}$$

where $s_i = (z_{i-k}, \cdots, z_{i-1})$. Note that, since we are dealing with stationary irreducible and aperiodic Markov noise processes, $P_S(\tilde{s}) > 0$

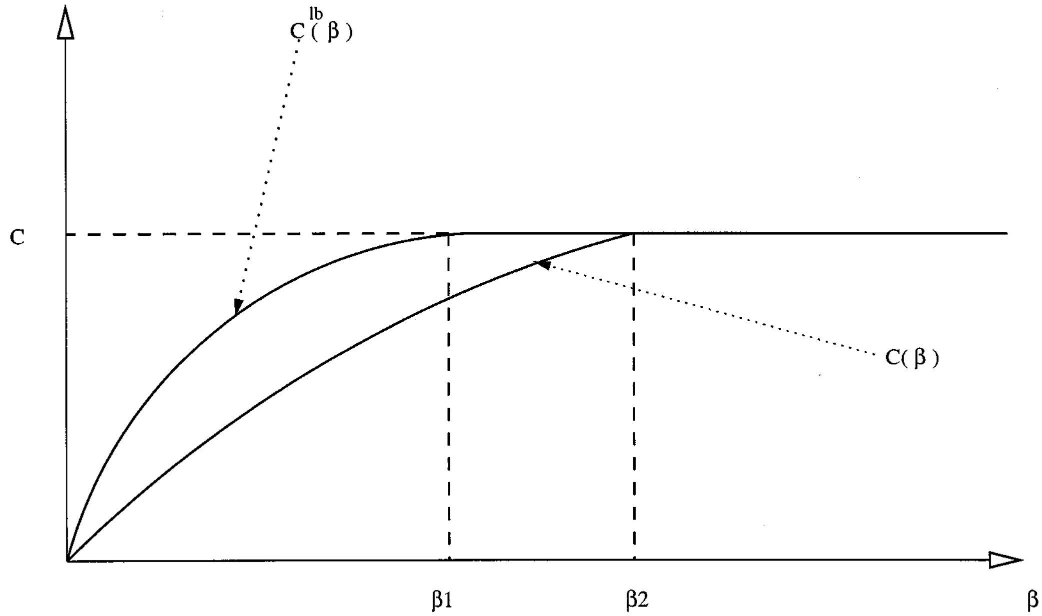

Fig. 5. $C^{\rm lb}(\beta)$ versus $C(\beta)$; $C = C_{\rm FB} = \log q - H(Z_\infty)$; $\beta 1 = \tilde{\beta}^{\rm lb}$; and $\beta 2 = \beta_{\max}$.

and thus $\beta_n^{\rm lb} < \beta$. Now, by Lemma 3, the channel transition probabilities are identical for the feedback and nonfeedback channels. Using $P^*_{V^n}(v^n)$ as a particular input distribution, we obtain that

$$\begin{aligned} C_n^{\rm lb}\left(\beta_n^{\rm lb}\right) &\geq \frac{1}{n}\sum_{v^n,y^n} P^*_{V^n}(v^n)P^{\rm FB}_{Y^n\,|\,V^n}(y^n\,|\,v^n) \\ &\quad\times \log\frac{P^{\rm FB}_{Y^n\,|\,V^n}(y^n\,|\,v^n)}{\sum_{v^n} P^*_{V^n}(v^n)P^{\rm FB}_{Y^n\,|\,V^n}(y^n\,|\,v^n)} \\ &= \frac{1}{n}\sum_{v^n,y^n} P^*_{V^n}(v^n)P_{Y^n\,|\,V^n}(y^n\,|\,v^n) \\ &\quad\times \log\frac{P_{Y^n\,|\,V^n}(y^n\,|\,v^n)}{\sum\limits_{v^n} P^*_{V^n}(v^n)P_{Y^n\,|\,V^n}(y^n\,|\,v^n)} \\ &= C_n(\beta). \end{aligned}$$
□

Theorem 3: Consider the q-ary nonfeedback and feedback channels with stationary irreducible and aperiodic k'th order additive Markov noise and feedback rule described above. Let $P_{Z_i\,|\,S_i}(z_i\,|\,\tilde{s}) = (1/q)$ for all $z_i \in \mathcal{A}_q$. Then for $0 < \beta < \beta_{\max}$

$$C_{\rm FB}(\beta) > C(\beta).$$

Proof: From Lemma 4, we have that

$$C_n^{\rm lb}\left(\beta_n^{\rm lb}\right) \geq C_n(\beta) \tag{20}$$

where

$$\beta_n^{\rm lb} = \left[1 - \frac{n-k}{n}P_S(\tilde{s})\right]\beta.$$

Therefore, taking the limit as $n \to \infty$ in (20), and using the fact that the limit of a concave function is concave and thus continuous, yield

$$C^{\rm lb}(\beta^{\rm lb}) \geq C(\beta) \tag{21}$$

where

$$\beta^{\rm lb} = \lim_{n\to\infty} \beta_n^{\rm lb} = [1 - P_S(\tilde{s})]\beta.$$

Since $C(\beta)$ is strictly increasing in β and $\beta > \beta^{\rm lb}$, we obtain that

$$C^{\rm lb}(\beta^{\rm lb}) > C(\beta^{\rm lb})$$

which implies that

$$C_{\rm FB}(\beta^{\rm lb}) > C(\beta^{\rm lb}) \quad \text{for } 0 < \beta^{\rm lb} < \beta_{\max}.$$
□

Observation: We already know from [1] that for additive noise channels an i.i.d. uniform input achieves the capacity without feedback (C), and that the capacity with feedback $(C_{\rm FB})$ is equal to C. Thus

$$C_{\rm FB} = C = \log q - H(Z_\infty) \tag{22}$$

for channels with stationary ergodic noise. From (21), (22), and the fact that $C^{\rm lb}(\cdot)$ is a lower bound to $C_{\rm FB}(\cdot)$ we remark that

$$\begin{aligned} C_{\rm FB}(\beta) &= C^{\rm lb}(\beta) = C_{\rm FB} = C \\ &= \log q - H(Z_\infty), \qquad \forall \beta \geq \tilde{\beta}^{\rm lb} \end{aligned}$$

where $\tilde{\beta}^{\rm lb} = [1 - P_S(\tilde{s})]\beta_{\max}$. We summarize the results of this observation and Theorem 3 by illustrating them in Fig. 5.

C. Numerical Examples

We have thus far demonstrated analytically that for a class of Markov noise sources and a specific feedback scheme, feedback can increase the capacity-cost function. This was achieved by showing that the lower bound to the capacity-cost function with feedback $(C^{\rm lb}(\beta))$ is strictly greater than the nonfeedback capacity-cost function $(C(\beta))$. We herein illustrate this result numerically by comparing, for a given block length n, $C^{\rm lb}(\beta)$ with the upper bound to $C(\beta)$ given (in Section III) by

$$C(\beta) \leq C_n^{\rm ub}(\beta) \triangleq C_n(\beta) + M_n.$$

Since

$$C^{\rm lb}(\beta) = \sup_n C_n^{\rm lb}(\beta)$$

it suffices to show that $C_n^{\rm lb}(\beta)$ is strictly greater than $C_n^{\rm ub}(\beta)$. As in Section III, we perform this numerical investigation using Blahut's algorithm for the computation of the capacity-cost function [3].

We use the binary examples of Section III, but with different transition probabilities. Examples for channels with ternary or quaternary alphabets can be obtained from [20]. In some instances we use a

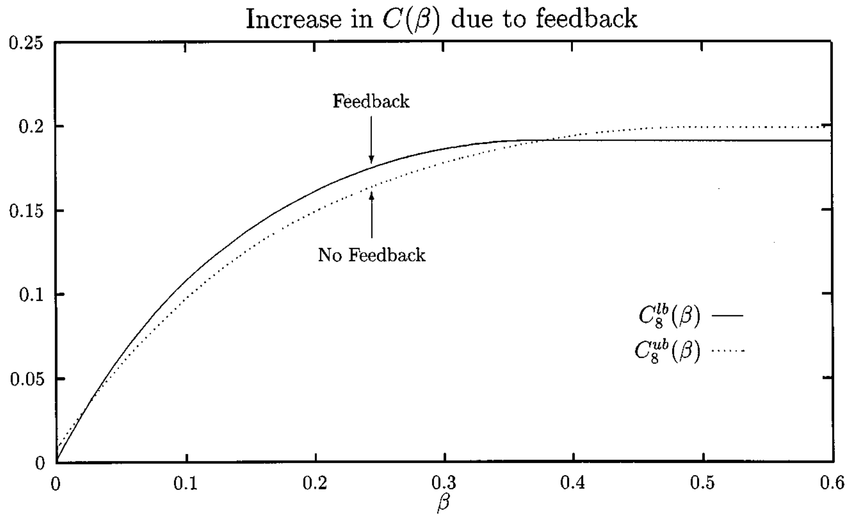

Fig. 6. Feedback increase: $C_8^{\rm lb}(\beta) > C_8^{\rm ub}(\beta)$ for $\beta < \tilde{\beta}_8^{\rm lb} = 0.375$, for a binary channel with first-order Markov noise given by $\mathbf{\Pi}_{2,\rm FB}$.

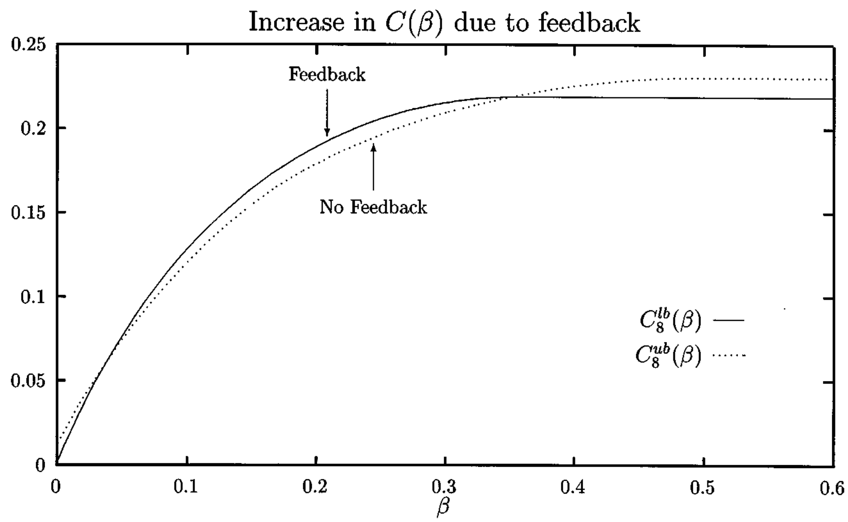

Fig. 7. Feedback increase: $C_8^{\rm lb}(\beta) > C_8^{\rm ub}(\beta)$ for $\beta < \tilde{\beta}_8^{\rm lb} = 0.375$, for a binary channel with first-order Markov noise given by $\mathbf{\Pi}_{2,\rm FB}^{(*)}$.

channel with a uniformly poor state $\tilde{s}$, and in others we use a nearly uniformly poor state. In both instances, we observe an increase in the capacity-cost function with feedback. The results, computed to an accuracy of 10^{-6}, are displayed in Figs. 6–8 for different channel parameters.

In Fig. 6, we use a binary channel with a first-order Markov noise described by

$$\mathbf{\Pi}_{2,\rm FB} = \begin{bmatrix} 0.8 & 0.2 \\ 0.5 & 0.5 \end{bmatrix}.$$

For the feedback scheme, we use $\tilde{s} = 1$. In Fig. 7, we employ a binary channel with first-order Markov noise defined by

$$\mathbf{\Pi}_{2,\rm FB}^{*} = \begin{bmatrix} 0.82 & 0.18 \\ 0.45 & 0.55 \end{bmatrix}$$

with $\tilde{s} = 1$. Fig. 7 shows that we obtain a numerical increase in $C_8^{\rm lb}(\beta)$ over $C_8^{\rm ub}(\beta)$ even if the state $\tilde{s}$ is not uniformly corrupting.

Finally, the example in Fig. 8 employs a second-order binary noise process with

$$\mathbf{\Pi}_{2,\rm FB}^{(2)} = \begin{bmatrix} 0.80 & 0.20 & 0 & 0 \\ 0 & 0 & 0.50 & 0.50 \\ 0.78 & 0.22 & 0 & 0 \\ 0 & 0 & 0.50 & 0.50 \end{bmatrix}.$$

In this example, the feedback rule is applied as follows:

$$X_i = f^*(V_i, S_i) \triangleq \begin{cases} V_i, & S_i \in \{(00),(10)\} \\ 0, & S_i \in \{(01),(11)\}. \end{cases}$$

All figures clearly indicate that feedback increases the capacity-cost function. Note that as the block length n increases, the increase due to feedback becomes larger since $C_n^{\rm ub}(\beta)$ decreases with n while $C_n^{\rm lb}(\beta)$ increases.

V. Summary

In this work, we investigated the capacity-cost function $C(\beta)$ of q-ary channels with additive Markov noise. We introduced average cost constraints on the input sequences of the additive channels, rendering them nonsymmetric. We proved a tight upper bound to $C(\beta)$; the bound

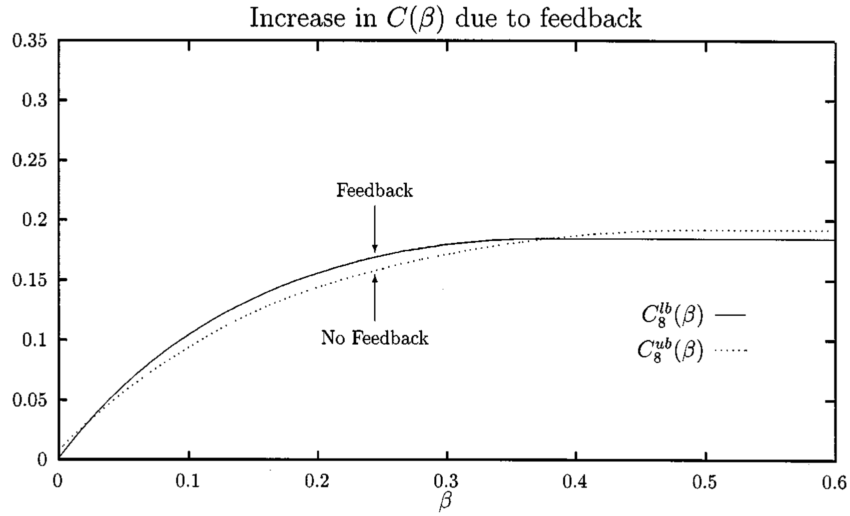

Fig. 8. Feedback increase: $C_8^{\rm lb}(\beta) > C_8^{\rm ub}(\beta)$ for a binary channel with second-order Markov noise given by $\mathbf{\Pi}_{2,\rm FB}^{(2)}$.

turns out to constitute the counterpart of the Wyner–Ziv lower bound to the rate-distortion function $R(D)$. This illustrates the striking duality that exists between $R(D)$ and $C(\beta)$, as luminously remarked by Shannon. Using this bound along with two other lower bounds to $C(\beta)$, we illustrated the computation of $C(\beta)$ via Blahut's algorithm for the calculation of channel capacity.

We then examined the effect of output feedback on the capacity-cost function of these channels. We demonstrated, both analytically and numerically, that for a particular feedback-encoding strategy and a class of Markov noise sources, feedback can increase the capacity-cost function. Future studies may include the investigation of the effect of feedback on the reliability function of discrete channels with memory.

References

[1] F. Alajaji, "Feedback does not increase the capacity of discrete channels with additive noise," *IEEE Trans. Inform. Theory*, vol. 41, pp. 546–549, Mar. 1995.

[2] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[3] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, July 1972.

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

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

[6] P. M. Ebert, "The capacity of the Gaussian channel with feedback," *Syst. Tech. J.*, vol. 49, pp. 1705–1712, Oct. 1970.

[7] U. Erez and R. Zamir, "Noise prediction for channel coding with side information at the transmitter," *IEEE Trans. Inform. Theory*, Nov. 1998, submitted for publication.

[8] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.

[9] S. Ihara, "Capacity of discrete time Gaussian feedback channel with and without feedback, I," *Memoirs of the Faculty of Science Kochi University, Series A Mathematics*, vol. 9, pp. 21–36, Mar. 1988.

[10] S. Ihara and K. Yanagi, "Capacity of discrete time Gaussian feedback channel with and without feedback, II," *Japan J. Appl. Math.*, vol. 6, pp. 245–258, June 1989.

[11] F. Jelinek, "Indecomposable channels with side information at the transmitter," *Inform. Contr.*, vol. 8, pp. 36–55, 1965.

[12] ——, "Determination of capacity achieving input probabilities for a class of finite state channels with side information," *Inform. Contr.*, vol. 9, pp. 101–129, 1966.

[13] R. J. McEliece, *The Theory of Information and Coding: A Mathematical Framework for Communication*. Cambridge, U.K.: Cambridge Univ. Press, 1984.

[14] M. S. Pinsker, presented at the The Soviet Information Theory Meeting, 1969.

[15] S. Shamai (Shitz) and A. D. Wyner, "A binary analog to the entropy-power inequality," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1428–1430, Nov. 1990.

[16] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, Sept. 1956.

[17] ——, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, pt. 4, pp. 142–163, Mar. 1959.

[18] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.

[19] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inform. Theory*, vol. 45, pp. 761–771, Mar. 1999 .

[20] N. Whalen, "The Capacity-Cost Function of Channels with Additive Markov Noise," M.Sc. Thesis, Queen's Univ., Kingston, Ont., Canada, June 1998, [Online] Available: http:\\markov.mast.queensu.ca.

[21] A. Wyner and J. Ziv, "Bounds on the rate-distortion function for stationary sources with memory," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 508–513, Sept. 1971.

[22] K. Yanagi, "Necessary and sufficient condition for capacity of the discrete time Gaussian channel to be increased by feedback," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1788–1791, Nov. 1992.