

# Feedback Capacity of the First-Order Moving Average Gaussian Channel

Young-Han Kim\*

Information Systems Laboratory, Stanford University, Stanford, CA 94305, USA

Email: yhk@stanford.edu

**Abstract**—The feedback capacity of the stationary Gaussian additive noise channel has been open, except for the case where the noise is white. Here we obtain the closed-form feedback capacity of the first-order moving average additive Gaussian noise channel. Specifically, the channel is given by  $Y_i = X_i + Z_i$ ,  $i = 1, 2, \dots$ , where the input  $\{X_i\}$  satisfies average power constraint and the noise  $\{Z_i\}$  is a first-order moving average Gaussian process defined by  $Z_i = \alpha U_{i-1} + U_i$ ,  $|\alpha| \leq 1$ , with white Gaussian innovation  $\{U_i\}_{i=0}^\infty$ . We show that the feedback capacity of this channel is  $-\log x_0$ , where  $x_0$  is the unique positive root of the equation  $\rho x^2 = (1 - x^2)(1 - |\alpha|x)^2$ , and  $\rho$  is the ratio of the average input power per transmission to the variance of the noise innovation  $U_i$ . The optimal coding scheme parallels the simple linear signalling scheme by Schalkwijk and Kailath for the additive white Gaussian noise channel; the transmitter sends a real-valued information-bearing signal at the beginning of communication, then subsequently processes the feedback noise process through a simple linear stationary first-order autoregressive filter to help the receiver recover the information by maximum likelihood decoding. The resulting probability of error decays doubly exponentially in the duration of the communication. This feedback capacity of the first-order moving average Gaussian channel is very similar in form to the best known achievable rate for the first-order autoregressive Gaussian noise channel studied by Butman, Wolfowitz, and Tiernan, although the optimality of the latter is yet to be established.

## I. INTRODUCTION

Consider the additive Gaussian noise channel with feedback as depicted in Figure 1. The channel  $Y_i = X_i + Z_i$ ,

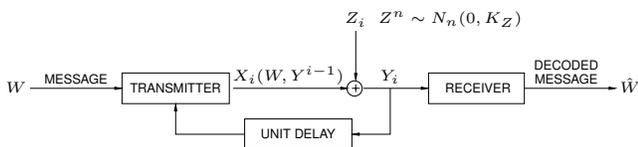


Fig. 1. Gaussian channel with feedback.

$i = 1, 2, \dots$ , has additive Gaussian noise  $Z_1, Z_2, \dots$ , where  $Z^n = (Z_1, \dots, Z_n) \sim N_n(0, K_Z)$ . We wish to communicate a message  $W \in \{1, 2, \dots, 2^{nR}\}$  reliably over the channel  $Y^n = X^n + Z^n$ . The channel output is causally fed back to the transmitter. We specify a  $(2^{nR}, n)$  code with the codewords  $(X_1(W), X_2(W, Y_1), \dots, X_n(W, Y^{n-1}))$  satisfying

the expected power constraint

$$E \frac{1}{n} \sum_{i=1}^n X_i^2(W, Y^{i-1}) \leq P, \quad (1)$$

and decoding function  $\hat{W}_n : \mathbb{R}^n \rightarrow \{1, 2, \dots, 2^{nR}\}$ . The probability of error  $P_e^{(n)}$  is defined by

$$P_e^{(n)} := \Pr\{\hat{W}_n(Y^n) \neq W\},$$

where the message  $W$  is independent of  $Z^n$  and is uniformly distributed over  $\{1, 2, \dots, 2^{nR}\}$ .

We define the  $n$ -block feedback capacity  $C_{n, \text{FB}}$  as

$$C_{n, \text{FB}} = \max \frac{1}{2n} \log \frac{\det((B + I)K_Z(B + I)^T + K_V)}{\det(K_Z)}, \quad (2)$$

where the maximization is over all  $n \times n$  nonnegative definite matrix  $K_V$  and  $n \times n$  strictly lower triangular matrix  $B$  such that  $\text{tr}(BK_ZB^T + K_V) \leq nP$ . In Cover and Pombra [1], it is shown that if for every  $\epsilon > 0$ , there exists a sequence of  $(2^{n(C_{n, \text{FB}} - \epsilon)}, n)$  codes with  $P_e^{(n)} \rightarrow 0$  as  $n \rightarrow \infty$ , and for every  $\epsilon > 0$  and any sequence of codes with  $2^{n(C_{n, \text{FB}} + \epsilon)}$  codewords,  $P_e^{(n)}$  is bounded away from zero for all  $n$ . When the noise process  $\{Z_n\}$  is stationary, the  $n$ -block capacity is super-additive in the sense that

$$n C_{n, \text{FB}} + m C_{m, \text{FB}} \leq (n + m) C_{n+m, \text{FB}}, \quad \text{for all } n, m.$$

Then the feedback capacity  $C_{\text{FB}}$  is well-defined as

$$C_{\text{FB}} = \lim_{n \rightarrow \infty} C_{n, \text{FB}}.$$

To obtain a closed-form expression for the feedback capacity  $C_{\text{FB}}$ , however, we need to go further than (2) since the above characterization does not give us any hint on the sequence (in  $n$ ) of  $(B, K_V)$  maximizing  $C_{n, \text{FB}}$  or its limiting behavior.

In this paper, we study in detail the case where the additive Gaussian noise process  $\{Z_i\}_{i=1}^\infty$  is a moving average process of order one (MA(1)). We define the Gaussian MA(1) noise process  $\{Z_i\}_{i=1}^\infty$  with parameter  $\alpha$ ,  $|\alpha| \leq 1$ , as

$$Z_i = \alpha U_{i-1} + U_i, \quad (3)$$

where  $\{U_i\}_{i=0}^\infty$  is a white Gaussian innovation process. Without loss of generality, we will assume that  $U_i$ ,  $i = 0, 1, \dots$ , has unit variance.

Note that the condition  $|\alpha| \leq 1$  is not restrictive. When  $|\alpha| > 1$ , it can be readily verified that the process  $\{Z_i\}$  has

\*This work was partly supported by NSF Grant CCR-0311633.

the same distribution as the process  $\{\tilde{Z}_i\}$  defined by  $\tilde{Z}_i = \alpha(\beta U_{i-1} + U_i)$ , where the moving average parameter  $\beta$  is given by  $\beta = 1/\alpha$ , thus  $|\beta| < 1$ .

We state the main theorem, the proof of which will be given in Section II.

*Theorem 1:* For the additive Gaussian MA(1) noise channel  $Y_i = X_i + Z_i$ ,  $i = 1, 2, \dots$ , with the Gaussian MA(1) noise process  $\{Z_i\}$  defined in (3), the feedback capacity  $C_{\text{FB}}$  under the power constraint  $\sum_{i=1}^n EX_i^2 \leq nP$  is given by

$$C_{\text{FB}} = -\log x_0,$$

where  $x_0$  is the unique positive root of the fourth-order polynomial

$$P x^2 = (1 - x^2)(1 - |\alpha|x)^2. \quad (4)$$

As will be shown later in Sections II and III, the feedback capacity  $C_{\text{FB}}$  is achieved by an asymptotically stationary ergodic input process  $\{X_i\}$  satisfying  $EX_i^2 = P$  for all  $i$ . Thus by ergodic theorem, the feedback capacity does not diminish under a more restrictive power constraint  $\frac{1}{n} \sum_{i=1}^n X_i^2(W, Y^{i-1}) \leq P$ .

The literature on Gaussian feedback channels is vast. We mention some of prior works closely related to our main discussion. In earlier work, Schalkwijk and Kailath [2], [3] considered the feedback over the additive white Gaussian noise channel, and proposed a simple linear signalling scheme that achieves the feedback capacity. The coding scheme by Schalkwijk and Kailath can be summarized as follows: Let  $\theta$  be one of  $2^{nR}$  equally spaced real numbers on some interval, say,  $[0, 1]$ . At time  $k$ , the receiver forms the maximum likelihood estimate  $\hat{\theta}_k(Y_1, \dots, Y_k)$  of  $\theta$ . Using the feedback information, at time  $k + 1$ , we send  $X_{k+1} = \gamma_k(\theta - \hat{\theta}_k)$ , where  $\gamma_k$  is a scaling factor properly chosen to meet the power constraint. After  $n$  transmissions, the receiver finds the value of  $\theta$  among  $2^{nR}$  alternatives that is closest to  $\hat{\theta}_n$ . This simple signalling scheme, essentially without any coding, achieves the feedback capacity. As is shown by Shannon [4], feedback does not increase the capacity of memoryless channels. The benefit of feedback, however, does not consist of the simplicity of coding only. The probability of decoding error of the Schalkwijk–Kailath scheme decays doubly exponentially in the duration of communication, compared to the exponential decay for the nonfeedback scenario.

Butman [5] generalized the linear coding scheme of for white noise processes to autoregressive (AR) noise processes. For first-order autoregressive (AR(1)) processes  $\{Z_i\}_{i=1}^{\infty}$  with regression parameter  $\alpha$ ,  $|\alpha| < 1$ , defined by  $Z_i = \alpha Z_{i-1} + U_i$ , he obtained a lower bound on the feedback capacity as  $-\log x_0$ , where  $x_0$  is the unique positive root of the fourth-order polynomial

$$P x^2 = \frac{(1 - x^2)}{(1 + |\alpha|x)^2}. \quad (5)$$

This rate has been shown to be optimal among a certain class of linear feedback schemes by Wolfowitz [6] and Tiernan [7] and is strongly believed to be the capacity of the AR(1)

feedback capacity. Tiernan and Schalkwijk [8] found an upper bound of the AR(1) feedback capacity, which meets Butman's lower bound for very low and very high signal-to-noise ratio. Butman [9] also obtained capacity upper and lower bounds for AR processes with higher order.

For the case of moving average (MA) noise processes, there are far fewer results in the literature, although MA processes are usually more tractable than AR processes of the same order. Ozarow [10], [11] gave upper and lower bounds of the feedback capacity for AR(1) and MA(1) channels and showed that feedback strictly increases the capacity. Substantial progress was made by Ordentlich [12]; he observed that  $K_V$  in (2) is at most of rank  $k$  for a MA noise process with order  $k$ . He also showed that the optimal  $(K_V, B)$  necessarily has the property that the current input signal  $X_k$  is orthogonal to the past outputs  $(Y_1, \dots, Y_{k-1})$ . For the special case of MA(1) processes, this development, combined with the arguments given by Wolfowitz [6], suggests that a linear signalling scheme similar to the Schalkwijk–Kailath scheme be optimal.

Our Theorem 1 confirms the optimality of the modified Schalkwijk–Kailath scheme and provides the first proof of feedback capacity for any nonwhite stationary colored Gaussian channels. This development links the Cover–Pombra formulation of the feedback capacity with the Schalkwijk–Kailath scheme and its generalizations to stationary colored channels, and casts new hope on the optimality of the achievable rate for the AR(1) channel obtained by Butman [5].

A recent report by Yang, Kavčić, and Tatikonda [13] studies the feedback capacity of the general ARMA( $k$ ) case using the state-space model and offers a conjecture on the feedback capacity as a solution to an optimization problem that does not depend on the horizon  $n$ . For the special case  $k = 1$  with the noise process  $\{Z_i\}_{i=1}^{\infty}$  defined by

$$Z_i = \beta Z_{i-1} + \alpha U_{i-1} + U_i, \quad |\alpha|, |\beta| < 1$$

they conjecture that the Schalkwijk–Kailath–Butman scheme is optimal. The corresponding achievable rate can be written in a closed form as  $-\log x_0$ , where  $x_0$  is the unique positive root of the fourth-order polynomial

$$P x^2 = \frac{(1 - x^2)(1 - \sigma\alpha x)^2}{(1 + \sigma\beta x)^2}$$

and

$$\sigma = \begin{cases} 1, & \alpha + \beta \geq 0, \\ -1, & \alpha + \beta < 0. \end{cases}$$

By taking  $\beta = 0$  or  $\alpha = 0$ , we can easily recover (4) and (5), respectively. Thus, our Theorem 1 gives a partial confirmation to the Yang–Kavčić–Tatikonda conjecture.

## II. PROOF OF THEOREM 1

We give a sketch of the proof. Refer to [14] for details.

*Proof outline.*

We first transform the optimization problem given in (2) to a series of (asymptotically) equivalent forms. Then we

solve the problem by imposing individual power constraints  $(P_1, \dots, P_n)$  on each input signal. Subsequently we optimize over  $(P_1, \dots, P_n)$  under the average power constraint

$$P_1 + \dots + P_n \leq nP.$$

Using Lemma 1 below, we prove that the uniform power allocation  $P_1 = \dots = P_n = P$  is asymptotically optimal. This will lead to a closed-form solution given in Theorem 1.

*Step 1. Transformations into equivalent optimization problems.*

Recall that we wish to solve the optimization problem:

$$\text{maximize } \log \det((B + I)K_Z(B + I)^T + K_V) \quad (6)$$

over all nonnegative definite  $K_V$  and strictly lower triangular  $B$  satisfying  $\text{tr}(BK_ZB^T + K_V) \leq nP$ . We approximate the covariance matrix  $K_Z$  of the given MA(1) noise process with parameter  $\alpha$  by another covariance matrix  $K'_Z$ . Define  $K'_Z = H_Z H_Z^T$  where the lower-triangular Toeplitz matrix  $H_Z$  is given by

$$(H_Z)_{ij} = \begin{cases} 1, & i = j, \\ \alpha, & i = j + 1, \\ 0, & \text{otherwise.} \end{cases}$$

The matrix  $K'_Z$  is a covariance matrix of the noise process, conditioned on the zeroth innovation  $U_0$ . It is not hard to verify that there is no asymptotic difference in capacity between the channel with the original noise covariance  $K_Z$  and the channel with  $K'_Z$ . Throughout we will assume that the noise covariance matrix of the given channel is  $K'_Z$ .

Now by identifying  $K_V$  with  $F_V F_V^T$  for some lower-triangular  $F_V$  and identifying  $F_Z$  with  $BH_Z$  for some strictly lower-triangular  $F_Z$ , we transform the optimization problem (6) with new variables  $(F_V, F_Z)$  as

$$\begin{aligned} &\text{maximize} && \log \det(F_V F_V^T + (F_Z + H_Z)(F_Z + H_Z)^T) \\ &\text{subject to} && \text{tr}(F_V F_V^T + F_Z F_Z^T) \leq nP. \end{aligned}$$

We shall use  $2n$ -dimensional row vectors  $f_i$  and  $h_i$ , to denote the  $i$ -th row of  $F := [F_V \ F_Z]$  and  $H := [0_{n \times n} \ H_Z]$ , respectively. There is an obvious identification between the time- $i$  input signal  $X_i$  and the vector  $f_i$ , for we can regard  $f_i$  as a point in the Hilbert space with the innovations of  $V^n$  and  $Z^n$  as a basis. We can similarly identify  $Z_i$  with  $h_i$  and  $Y_i$  with  $f_i + h_i$ . We also introduce new variables  $(P_1, \dots, P_n)$  representing the power constraint for each input  $f_i$ . Now the above optimization problem becomes the following equivalent form:

$$\begin{aligned} &\text{maximize} && \log \det((F + H)(F + H)^T) \\ &\text{subject to} && \|f_i\|^2 \leq P_i, \quad i = 1, \dots, n, \\ &&& \sum_{i=1}^n P_i \leq nP. \end{aligned} \quad (7)$$

Here  $\|\cdot\|$  denotes the Euclidean norm of a  $2n$ -dimensional vector. Note that the variables  $(f_1, \dots, f_n)$  should satisfy the relevant triangularity conditions inherited from  $(F_V, F_Z)$ .

*Step 2. Optimization under the individual power constraint for each signal.*

We solve the optimization problem (7) in  $(f_1, \dots, f_n)$  after fixing  $(P_1, \dots, P_n)$ . We need some notation first.

We define an  $n$ -by- $2n$  matrix  $S := F + H$ , and denote by  $S_k$  the  $k$ -by- $2n$  submatrix of  $S$  consisting of the first  $k$  rows of  $S$ . We also define a sequence of  $2n$ -by- $2n$  matrices  $\{\Pi_k\}_{k=1}^{n-1}$  as  $\Pi_k = I - S_k^T (S_k S_k^T)^{-1} S_k$ . Note that  $S_k$  is of full rank and thus that  $(S_k S_k^T)^{-1}$  always exists. We can view  $\Pi_k$  as a map of a  $2n$ -dimensional row vector (acting from the right) to its component orthogonal to the subspace spanned by the rows  $s_1, \dots, s_k$  of  $S_k$ . (Or  $\Pi_k$  maps a generic random variable  $A$  to  $A - E(A|Y^k)$ .) We observe that, for  $k = 2, \dots, n$ ,

$$\det(S_k S_k^T) = \det(S_{k-1} S_{k-1}^T) \cdot s_k \Pi_{k-1} s_k^T. \quad (8)$$

Finally we define the intermediate objective functions of the maximization (7) as

$$J_k(P_1, \dots, P_k) := \max_{\substack{f_1, \dots, f_k \\ \|f_i\|^2 \leq P_i}} \log \det(S_k S_k^T), \quad k = 1, \dots, n,$$

so that  $C_{n,\text{FB}} = \max_{P_i: \sum P_i \leq nP} \frac{1}{2n} J_n(P_1, \dots, P_n)$ .

From (8), we have

$$\begin{aligned} J_n &= \max_{f_1^T} \left( \log \det(S_{n-1} S_{n-1}^T) + \log(s_n \Pi_{n-1} s_n^T) \right) \\ &= \max_{f_1^{n-1}} \left( \log \det(S_{n-1} S_{n-1}^T) + \max_{f_n} \log(s_n \Pi_{n-1} s_n^T) \right). \end{aligned} \quad (9)$$

For fixed  $(f_1, \dots, f_{n-1})$  we can easily maximize the second term of (9) by some  $f_n^*$  satisfying  $f_n^* = f_n^* \Pi_{n-1}$ . Furthermore, thanks to the special structure of the noise covariance matrix  $K'_Z$ , we can show that both terms of (9) are maximized by the same  $(f_1^*, \dots, f_{n-1}^*)$ . In other words, if  $(f_1^*, \dots, f_{n-1}^*)$  maximizes  $J_{n-1}(P_1, \dots, P_{n-1})$ , then  $(f_1^*, \dots, f_{n-1}^*, f_n^*)$  maximizes  $J_n(P_1, \dots, P_n)$  for some  $f_n^*$ . Thus the maximization for  $J_n$  can be solved in a greedy fashion by sequentially maximizing  $J_1, J_2, \dots, J_n$  through  $f_1^*, f_2^*, \dots, f_n^*$ , whence we obtain the recursive relationship

$$J_k = J_{k-1} + \log \left( 1 + \left( \sqrt{P_k} + |\alpha| \sqrt{1 - \frac{1}{e^{J_{k-1} - J_{k-2}}}} \right)^2 \right), \quad (10)$$

for  $k \geq 1$ , with  $J_0 = J_{-1} = 0$ .

*Step 3. Optimal power allocation over time.*

In the previous step, we solved the optimization problem (7) under a fixed power allocation  $(P_1, \dots, P_n)$ . Thanks to the special structure of the MA(1) noise process, this brute force optimization was tractable via backward dynamic programming. Here we optimize the power allocation  $(P_1, \dots, P_n)$  under the constraint  $\sum_{i=1}^n P_i \leq nP$ ,

When  $\alpha = 0$ , that is, when the noise is white, we can use the concavity of the logarithm to show that, for all  $n$ ,

$$\begin{aligned} C_{n,\text{FB}} &= \max_{P_i: \sum_i P_i \leq nP} \frac{1}{2n} \sum_{i=1}^n \log(1 + P_i) \\ &= \frac{1}{2} \log(1 + P), \end{aligned}$$

with  $P_1^* = \dots = P_n^* = P$ . When  $\alpha \neq 0$ , it is not tractable to optimize  $(P_1, \dots, P_n)$  for  $J_n$  in (10) and get a closed-form solution of  $C_{n,\text{FB}}$  for finite  $n$ . The following lemma, however, enables us to figure out the asymptotically optimal power allocation and to obtain a closed-form solution for  $C_{\text{FB}} = \lim_n C_{n,\text{FB}}$ .

*Lemma 1:* Let  $\psi : [0, \infty) \times [0, \infty) \rightarrow [0, \infty)$  such that the following conditions hold:

- (i)  $\psi(\xi, \zeta)$  is continuous, concave in  $(\xi, \zeta)$ , and strictly concave in  $\zeta$ ;
- (ii)  $\psi(\xi, \zeta)$  is increasing in  $\xi$  and  $\zeta$ , respectively; and
- (iii) for any  $\zeta > 0$ , there is a unique solution  $\xi^* = \xi^*(\zeta) > 0$  to the equation  $\xi = \psi(\xi, \zeta)$ .

For some fixed  $P > 0$ , let  $\{P_i\}_{i=1}^\infty$  be any infinite sequence of nonnegative numbers satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n P_i \leq P. \quad (11)$$

Let  $\{\xi_i\}_{i=0}^\infty$  be defined recursively as  $\xi_0 := 0$  and  $\xi_i = \psi(\xi_{i-1}, P_i)$ ,  $i = 1, 2, \dots$ . Then

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i \leq \xi^*.$$

Furthermore, if  $P_i \equiv P$ ,  $i = 1, 2, \dots$ , then the corresponding  $\xi_i$  converges to  $\xi^*$ .

To apply the above lemma to our problem, we define

$$\psi(\xi, \zeta) := \frac{1}{2} \log \left( 1 + \left( \sqrt{\zeta} + |\alpha| \sqrt{1 - \frac{1}{e^{2\xi}}} \right)^2 \right).$$

The conditions (i) – (iii) of Lemma 2 can be easily checked. Now from (10) and Lemma 2, we can show that

$$\limsup_{n \rightarrow \infty} \frac{1}{2n} J_n(P_1, \dots, P_n) = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i \leq \xi^*,$$

where  $\xi^*$  is the unique solution to

$$\xi = \psi(\xi, P) = \frac{1}{2} \log \left( 1 + \left( \sqrt{P} + |\alpha| \sqrt{1 - \frac{1}{e^{2\xi}}} \right)^2 \right). \quad (12)$$

Since our choice of  $\{P_i\}$  is arbitrary, we conclude that

$$\sup \limsup_{n \rightarrow \infty} \frac{1}{2n} J_n(P_1, \dots, P_n) = \xi^*,$$

where the supremum (in fact, maximum) is over all infinite sequences  $\{P_i\}$  satisfying the asymptotic average power constraint (11).

Finally, we show that  $\xi^*$  is indeed the feedback capacity by observing

$$C_{\text{FB}} = \lim_{n \rightarrow \infty} \max_{P_i: \sum_i P_i \leq nP} \frac{1}{2n} J_n(P_1, \dots, P_n) \quad (13)$$

$$= \sup_{\{P_i\}_{i=1}^\infty} \limsup_{n \rightarrow \infty} \frac{1}{2n} J_n(P_1, \dots, P_n) = \xi^*. \quad (14)$$

(One can justify the interchange of the order of limit and supremum in (13) and (14) using standard techniques in analysis.)

We conclude this section by characterizing the capacity  $C_{\text{FB}} = \xi^*$  in an alternative form. Let  $x_0 = \exp(-\xi^*)$ , or equivalently,  $\xi^* = -\log x_0$ . From (12), it is easy to verify that  $0 < x_0 \leq 1$  is the unique positive solution to

$$\frac{1}{x^2} = 1 + \left( \sqrt{P} + |\alpha| \sqrt{1 - x^2} \right)^2,$$

or equivalently,

$$P x^2 = (1 - x^2)(1 - |\alpha|x)^2.$$

This establishes the feedback capacity  $C_{\text{FB}}$  of the first-order moving average additive Gaussian noise channel with parameter  $\alpha$ .

### III. DISCUSSION

The derived asymptotically optimal feedback input signal sequence, or equivalently, the sequence of matrices  $(K_V^{(n)}, B^{(n)})$  has two prominent properties. First, the optimal  $(K_V, B)$  for the  $n$ -block can be found sequentially, built on the optimal  $(K_V, B)$  for the  $(n-1)$ -block. Although this property may sound quite natural, it is not true in general for other channel models. Later in this section, we will see an MA(2) channel counterexample. As a corollary to this sequentiality property, the optimal  $K_V$  has rank one, which agrees with the previous result by Ordentlich [12]. Secondly, the current input signal  $X_k$  is orthogonal to the past output signals  $(Y_1, \dots, Y_{k-1})$ . In the notation of Section II, we have  $f_k S_{k-1}^T = 0$ . This orthogonality property is indeed a necessary condition for the optimal  $(K_V, B)$  for any noise covariance matrix  $K_Z$ . (See [12], [15].)

We explore the possibility of extending the current proof technique to a more general class of noise processes. The answer is negative. We comment on two simple cases: MA(2) and AR(1). Consider the following MA(2) noise process which is essentially two interleaved MA(1) processes:

$$Z_i = U_i + \alpha U_{i-2}, \quad i = 1, 2, \dots$$

It is easy to see that this channel has the same capacity as the MA(1) channel with parameter  $\alpha$ , which can be attained by signalling separately for each interleaved MA(1) channel. This suggests that the sequentiality property does not hold for this example. Indeed, we sequentially optimize the  $n$ -block capacity to obtain the rate  $-\log x_0$ , where  $x_0$  is the unique positive root of the sixth order polynomial

$$P x^2 = (1 - x^2)(1 - |\alpha|x^2)^2.$$

It is not hard to see that this rate is strictly less than the MA(1) feedback capacity unless  $\alpha = 0$ . Note that a similar argument can prove that Butman's conjecture on the AR( $k$ ) capacity [9, Abstract] is not true in general, for  $k > 1$ .

In contrast to MA(1) channels, we are missing two basic ingredients for AR(1) channels — the optimality of rank-one

$K_V$  and the asymptotic optimality of the uniform power allocation. Under these two conditions, both of which are yet to be justified, it is known [6], [7] that the optimal achievable rate is given by (5). Despite the striking similarity, or reciprocity, between the polynomial (5) and the polynomial (4) for the MA(1) feedback, there is a major difficulty in establishing the above two conditions by the two-stage optimization strategy we used in the previous section, namely, first maximizing  $(f_1, \dots, f_n)$  and then  $(P_1, \dots, P_n)$ . For certain values of individual signal power constraints  $(P_1, \dots, P_n)$ , the optimal  $(f_1, \dots, f_n)$  does not satisfy the sequentiality, resulting in  $K_V$  with rank higher than one. Hence we cannot obtain the recursion formula for the  $n$ -block capacity [6, Section 5] that corresponds to (10) through a greedy maximization of  $J_n(P_1, \dots, P_n)$ .

Finally we show that the feedback capacity of the MA(1) channel can be achieved by using a simple stationary filter of the noise innovation process. At the beginning of the communication, we send  $X_1 \sim N(0, P)$ . For subsequent transmissions, we use a first-order regressive filter and transmit the filtered version of the noise innovation process as

$$X_k = \beta X_{k-1} + \sigma U_{k-1}, \quad k \geq 2. \quad (15)$$

Here  $\beta = -\text{sgn}(\alpha) x_0$  with  $x_0$  being the same unique positive root of the fourth-order polynomial (4) in Theorem 1, and the scaling factor  $\sigma$  is chosen to satisfy the power constraint as  $\sigma = \text{sgn}(\alpha) \sqrt{P(1 - \beta^2)}$ . This input process and the MA(1) noise process (3) yield the output process  $\{Y_i\}_{i=1}^\infty$  given by

$$\begin{aligned} Y_1 &= X_1 + \alpha U_0 + U_1, \\ Y_k &= \beta Y_{k-1} - \alpha \beta U_{k-2} + (\alpha - \beta + \sigma) U_{k-1}, \quad k \geq 2, \end{aligned}$$

which is asymptotically stationary with power spectral density  $S_Y(\omega) = \beta^{-2} |1 + \alpha \beta^2 e^{-j\omega}|^2$  for  $\{Y_k\}, k \geq 2$ . It is easy to check that the entropy rate of the output process  $\{Y_k\}$  is given by  $\frac{1}{2} \log(2\pi e \beta^{-2})$ . Hence we achieve the feedback capacity  $C_{\text{FB}}$ . Furthermore, it can be shown that the mean-square error of  $X_1$  given the observations  $Y_1, \dots, Y_n$  decays exponentially with rate  $\beta^{-2} = 2^{2C_{\text{FB}}}$ . In other words,

$$\text{var}(X_1 | Y_1, \dots, Y_n) \doteq P 2^{-2nC_{\text{FB}}}. \quad (16)$$

We can interpret the signal  $X_k$  as the adjustment of the receiver's estimate of the message bearing signal  $X_1$  after observing  $(Y_1, \dots, Y_{k-1})$ . The connection to the Schalkwijk-Kailath coding scheme is now apparent. Recall that there is a simple linear relationship between the minimum mean square error estimate (in other words, the minimum variance biased estimate) for the Gaussian input  $X_1$  and the maximum likelihood estimate (or equivalently, the minimum variance unbiased estimate) for an arbitrary real input  $\theta$ . Thus we can easily transform the above coding scheme based on the asymptotic equipartition property [1] to the Schalkwijk-like linear coding scheme based on the maximum likelihood nearest neighborhood decoding of uniformly spaced  $2^{nR}$  points. More specifically, we send as  $X_1$  one of  $2^{nR}$  possible signals, say,  $\theta \in \Theta := \{-\sqrt{P}, -\sqrt{P} + \Delta, -\sqrt{P} + 2\Delta, \dots, \sqrt{P} -$

$2\Delta, \sqrt{P} - \Delta, \sqrt{P}\}$ , where  $\Delta = 2\sqrt{P}/(2^{nR} - 1)$ . Subsequent transmissions follow (15). The receiver forms the maximum likelihood estimate  $\hat{\theta}_n(Y_1, \dots, Y_n)$  and finds the nearest signal point to  $\hat{\theta}_n$  in  $\Theta$ . The analysis of the error for this coding scheme follows Schalkwijk [3] and Butman [5]. From (16) and the standard result on the relationship between the minimum variance unbiased and biased estimation errors, the maximum likelihood estimation error  $\hat{\theta}_n - \theta$  is, conditioned on  $\theta$ , Gaussian with mean  $\theta$  and variance exponentially decaying with rate  $\beta^{-2} = 2^{2nC_{\text{FB}}}$ . Thus, the nearest neighbor decoding error, ignoring lower order terms, is given by

$$E_\theta \left[ \Pr \left( |\hat{\theta}_n - \theta| \geq \frac{\Delta}{2} \mid \theta \right) \right] \doteq \text{erfc} \left( \sqrt{\frac{3}{2\sigma_\theta^2}} 2^{n(C_{\text{FB}} - R)} \right),$$

where  $\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt$  and  $\sigma_\theta^2$  is the variance of input signal  $\theta$  chosen uniformly over  $\Theta$ . As far as  $R < C_{\text{FB}}$ , the decoding error decays doubly exponentially in  $n$ . Note that this coding scheme uses only the second moments of the noise process. This implies that the rate  $C_{\text{FB}}$  is achievable for the additive noise channel with any non-Gaussian noise process with the same covariance structure.

## REFERENCES

- [1] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 37–43, January 1989.
- [2] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback—I: No bandwidth constraint," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 172–182, April 1966.
- [3] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback—II: Band-limited signals," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 183–189, April 1966.
- [4] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, September 1956.
- [5] S. A. Butman, "A general formulation of linear feedback communication systems with solutions," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 392–400, May 1969.
- [6] J. Wolfowitz, "Signalling over a Gaussian channel with feedback and autoregressive noise," *J. Appl. Probab.*, vol. 12, pp. 713–723, 1975.
- [7] J. C. Tiernan, "Analysis of the optimum linear system for the autoregressive forward channel with noiseless feedback," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 359–363, May 1976.
- [8] J. C. Tiernan and J. P. M. Schalkwijk, "An upper bound to the capacity of the band-limited Gaussian autoregressive channel with noiseless feedback," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 311–316, May 1974.
- [9] S. A. Butman, "Linear feedback rate bounds for regressive channels," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 363–366, May 1976.
- [10] L. H. Ozarow, "Random coding for additive Gaussian channels with feedback," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 17–22, January 1990.
- [11] L. H. Ozarow, "Upper bounds on the capacity of Gaussian channels with feedback," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 156–161, January 1990.
- [12] E. Ordentlich, "A class of optimal coding schemes for moving average additive Gaussian noise channels with feedback," in *Proc. IEEE Int. Symp. Inform. Theory*, p. 467, June 1994.
- [13] S. Yang, A. Kavcic, and S. Tatikonda, "Linear Gaussian channels: feedback capacity under power constraints," in *Proc. IEEE Int. Symp. Inform. Theory*, p. 72, June 2004.
- [14] Y.-H. Kim, "Feedback capacity of the first-order moving average Gaussian channel," submitted to *IEEE Trans. Inform. Theory*, November 2004. (Preprint available at [www.arxiv.org/abs/cs.IT/0411036](http://www.arxiv.org/abs/cs.IT/0411036).)
- [15] S. Ihara, "On the capacity of the discrete time Gaussian channel with feedback," in *Trans. Eighth Prague Conf. Inform. Theory, Czech. Acad. Sci.* 1979.