Feedback Capacity of the First-Order Moving Average Gaussian Channel

Young-Han Kim, *Student Member, IEEE*

***Abstract*—Despite numerous bounds and partial results, the feedback capacity of the stationary nonwhite Gaussian additive noise channel has been open, even for the simplest cases such as the first-order *autoregressive* Gaussian channel studied by Butman, Tiernan and Schalkwijk, Wolfowitz, Ozarow, and more recently, Yang, Kavčić, and Tatikonda. Here we consider another simple special case of the stationary first-order *moving average* additive Gaussian noise channel and find the feedback capacity in closed form. Specifically, the channel is given by $Y_i = X_i + Z_i,\ i = 1, 2, \ldots,$ where the input $\{X_i\}$ satisfies a power constraint and the noise $\{Z_i\}$ is a first-order moving average Gaussian process defined by $Z_i = \alpha U_{i-1} + U_i,\ |\alpha| \le 1,$ with white Gaussian innovations $U_i,\ i = 0, 1, \ldots.$**

We show that the feedback capacity of this channel is

$$C_{\mathrm{FB}} = -\log x_0$$

where x_0 is the unique positive root of the equation

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

and ρ 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 signaling 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 and subsequently refines the receiver's knowledge by processing the feedback noise signal through a linear stationary first-order autoregressive filter. The resulting error probability of the maximum likelihood decoding decays doubly exponentially in the duration of the communication. Refreshingly, 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 given by Butman.

***Index Terms*—Additive Gaussian noise channels, feedback capacity, first-order moving average, Gaussian feedback capacity, linear feedback, Schalkwijk–Kailath coding scheme.**

I. Introduction and Summary

CONSIDER the additive Gaussian noise channel with feedback as depicted in Fig. 1. The channel $Y_i = X_i + Z_i$, $i = 1, 2, \ldots,$ has additive Gaussian noise $Z_1, Z_2, \ldots,$ where $Z^n = (Z_1, \ldots, Z_n) \sim N_n(0, K_Z)$. We wish to communicate a message index $W \in \{1, 2, \ldots, 2^{nR}\}$ 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[1] $(X_1(W), X_2(W, Y_1), \ldots, X_n(W, Y^{n-1}))$ satisfying the expected power constraint

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

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

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

where the message W is independent of Z^n and is uniformly distributed over $\{1, 2, \ldots, 2^{nR}\}$. We call the sequence $\{C_{n,\mathrm{FB}}\}_{n=1}^{\infty}$ an *n-block feedback capacity* sequence if for every $\epsilon > 0$, there exists a sequence of $(2^{n(C_{n,\mathrm{FB}} - \epsilon)}, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$, and for every $\epsilon > 0$ and any sequence of codes with $2^{n(C_{n,\mathrm{FB}} + \epsilon)}$ codewords, $P_e^{(n)}$ is bounded away from zero for all n. We define the feedback capacity C_{FB} as

$$C_{\mathrm{FB}} := \lim_{n \to \infty} C_{n,\mathrm{FB}}$$

if the limit exists. This definition of feedback capacity as the supremum of achievable rates agrees with the usual operational definition for the capacity of memoryless channels without feedback [1].

In [2], Cover and Pombra characterized the n-block feedback capacity $C_{n,\mathrm{FB}}$ as

$$C_{n,\mathrm{FB}} = \max_{\mathrm{tr}(K_X) \le nP} \frac{1}{2n} \log \frac{\det(K_Y)}{\det(K_Z)}. \tag{1}$$

Here $K_X = K_X(n)$, $K_Y = K_Y(n)$ and $K_Z = K_Z(n)$ respectively denote the covariance matrices of X^n, Y^n and Z^n, and the maximization is over all X^n of the form $X^n = BZ^n + V^n$ with a strictly lower-triangular $n \times n$ matrix $B = B(n)$ and multivariate Gaussian V^n independent of Z^n such that $E\sum_{i=1}^{n} X_i^2 = \mathrm{tr}(K_X) \le nP$. Equivalently, we can rewrite (1) as

$$C_{n,\mathrm{FB}} = \max_{K_V, B} \frac{1}{2n} \log \frac{\det((B+I)K_Z(B+I)^T + K_V)}{\det(K_Z)} \tag{2}$$

Manuscript received November 11, 2004; revised July 12, 2005. This work was supported in part by the NSF under Grant CCR-0311633. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005.

The author is with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: yhk@stanford.edu).

Communicated by Y. Steinberg, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2006.876217

[1]More precisely, encoding functions $X_i : \{1, \ldots, 2^{nR}\} \times \mathbb{R}^{i-1} \to \mathbb{R},\ i = 1, 2, \ldots, n$.

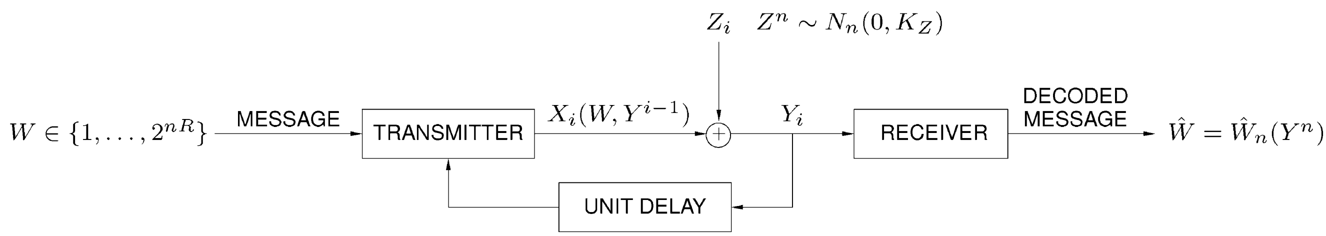

Fig. 1. Gaussian channel with feedback.

where the maximization is over all nonnegative definite $n \times n$ matrices $K_V = K_V(n)$ and strictly lower triangular $n \times n$ matrices $B = B(n)$ such that $\mathrm{tr}(BK_ZB^T + K_V) \le nP$.

When the noise process $\{Z_n\}$ is stationary, the n-block capacity is super-additive in the sense that

$$n\,C_{n,\mathrm{FB}} + m\,C_{m,\mathrm{FB}} \le (n+m)\,C_{n+m,\mathrm{FB}}$$

for all $n, m = 1, 2, \ldots$. Consequently, the feedback capacity C_{FB} is well-defined (see, for example, Pólya and Szegö [3]) as

$$\begin{aligned} C_{\mathrm{FB}} &= \lim_{n\to\infty} C_{n,\mathrm{FB}} \\ &= \lim_{n\to\infty} \max_{K_V(n), B(n)} \frac{1}{2n} \log \frac{\det((B+I)K_Z(B+I)^T + K_V)}{\det(K_Z)}. \end{aligned} \tag{3}$$

To obtain a closed-form expression for the feedback capacity C_{FB}, however, we need to go further than (3) since the above characterization does not give any hint on the sequence of the optimal $(B(n), K_V(n))_{n=1}^{\infty}$ achieving $C_{n,\mathrm{FB}}$ or, more importantly, 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 $\alpha,\ |\alpha| \le 1$, as

$$Z_i = \alpha U_{i-1} + U_i \tag{4}$$

where $\{U_i\}_{i=0}^{\infty}$ is a white Gaussian innovation process. Without loss of generality, we will assume that $U_i,\ i = 0, 1, \ldots$, has unit variance. There are alternative ways of defining Gaussian MA(1) processes, which we will review in Section II.

Note that the condition $|\alpha| \le 1$ is not restrictive. When $|\alpha| > 1$, it can be readily verified that the process $\{Z_i\}$ has 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 β is given by $\beta = 1/\alpha$, thus giving $|\beta| < 1$.

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

Theorem 1: For the additive Gaussian MA(1) noise channel $Y_i = X_i + Z_i,\ i = 1, 2, \ldots$, with the Gaussian MA(1) noise process $\{Z_i\}$ defined in (4), the feedback capacity C_{FB} under the power constraint $\sum_{i=1}^{n} EX_i^2 \le nP$ is given by

$$C_{\mathrm{FB}} = -\log x_0$$

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

$$P\,x^2 = (1 - x^2)(1 - |\alpha| x)^2. \tag{5}$$

As will be shown later in Sections III and IV, the feedback capacity C_{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

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

(See also the arguments given in [2, Sec. VIII] based on the stationarity of the noise process.)

The literature on Gaussian feedback channels is vast. We first mention some prior work closely related to our main discussion. In earlier work, Schalkwijk and Kailath [4], [5] (see also the discussion by Wolfowitz [6]) considered the feedback over the additive white Gaussian noise channel, and proposed a simple linear signaling scheme that achieves the feedback capacity. The coding scheme by Schalkwijk and Kailath can be summarized as follows: Let θ be one of 2^{nR} equally spaced real numbers on some interval, say, $[-1, 1]$. At time k, the receiver forms the maximum likelihood estimate $\hat{\theta}_k(Y_1, \ldots, Y_k)$ of θ. Using the feedback information, at time $k+1$, we send $X_{k+1} = \gamma_k(\theta - \hat{\theta}_k)$, where γ_k is a scaling factor properly chosen to meet the power constraint. After n transmissions, the receiver finds the value of θ among 2^{nR} alternatives that is closest to $\hat{\theta}_n$. This simple signaling scheme, without any coding, achieves the feedback capacity. As is shown by Shannon [7], feedback does not increase the capacity of memoryless channels. (See also Kadota *et al.* [8], [9] for continuous cases.) 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. In fact, there exists a feedback coding scheme such that the probability of decoding error decreases more rapidly than the exponential of any order [10]–[12]. Later, Schalkwijk extended his work to the center-of-gravity information feedback for higher dimensional signal spaces [13].

Butman [14] generalized the linear coding scheme of Schalkwijk and Kailath for white noise processes to autoregressive (AR) noise processes. For the first-order autoregressive (AR(1)) process $\{Z_i\}_{i=1}^{\infty}$ with regression parameter α, $|\alpha| < 1$, defined by

$$Z_i = \alpha Z_{i-1} + U_i \tag{6}$$

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

$$Px^2 = \frac{(1-x^2)}{(1+|\alpha|x)^2}. \tag{7}$$

This rate has been shown to be optimal among a certain class of linear feedback schemes by Wolfowitz [15] and Tiernan [16], and is strongly believed to be the capacity of the AR(1) feedback capacity. Tiernan and Schalkwijk [17] found an upper bound on the AR(1) feedback capacity, which meets Butman's lower bound for very low and very high signal-to-noise ratio. Butman [18] 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 [19], [20] 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 [21]; he observed that K_V in (2) is at most of rank k for a MA noise process with order k. Ordentlich 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 $(Y_1, \ldots, Y_{k-1})$. For the special case of MA(1) processes, this development, combined with the arguments given in [15], suggests that a linear signaling scheme similar to the Schalkwijk–Kailath scheme be optimal, which is proved by our Theorem 1.

A recent report by Yang, Kavčić, and Tatikonda [22] (see also Yang's thesis [23]) 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 coding 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

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

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 (5) and (7), respectively. Thus, in the special case $\beta = 0$, our Theorem 1 confirms the Yang–Kavčić–Tatikonda conjecture.

To conclude this section, we review, in a rather incomplete manner, previous work on the Gaussian feedback channel in addition to aforementioned results, and then point out where the current work lies in the literature. The standard literature on the Gaussian feedback channel and associated simple feedback coding schemes traces back to a 1956 paper by Elias [24] and its sequels [25], [26]. Turin [27]–[29], Horstein [30], Khas'minskii [31], and Ferguson [32] studied a sequential binary signaling scheme over the Gaussian feedback channel with symbol-by-symbol decoding that achieves the feedback capacity with an error exponent better than the nonfeedback case. As mentioned above, Schalkwijk and Kailath [4], [5], [13] made a major breakthrough by showing that a simple linear feedback coding scheme achieves the feedback capacity with doubly exponentially decreasing probability of decoding error. This fascinating result has been extended in many directions. Omura [33] reformulated the feedback communication problem as a stochastic-control problem and applied this approach to multiplicative and additive noise channels with noiseless feedback and to additive noise channels with noisy feedback. Pinsker [10], Kramer [11], and Zigangirov [12] studied feedback coding schemes under which the probability of decoding error decays as the exponential of arbitrary high order. Wyner [34] and Kramer [11] studied the performance of the Schalkwijk–Kailath scheme under a peak energy constraint and reported the singly exponential behavior of the probability of decoding error under a peak energy constraint. The error exponent of the Gaussian feedback channel under the peak power constraint was later obtained by Schalkwijk and Barron [35]. Kashyap [36], Lavenberg [37], [38], and Kramer [11] looked at the case of noisy or intermittent feedback.

The question of transmitting a Gaussian source over a Gaussian feedback channel was studied by Kailath [39], Cruise [40], Schalkwijk and Bluestein [41], Ovseevich [42], and Ihara [43]. There are also many notable extensions of the Schalkwijk–Kailath scheme in the area of multiple user information theory. Using the Schalkwijk–Kailath scheme, Ozarow and Leung-Yan-Cheong [44] showed that feedback increases the capacity region of *stochastically* degraded broadcast channels, which is rather surprising since feedback does *not* increase the capacity region of *physically* degraded broadcast channels, as shown by El Gamal [45]. Ozarow [46] also established the feedback capacity region of two-user white Gaussian multiple access channel through a very innovative application of the Schalkwijk–Kailath coding scheme. The extension to a larger number of users was attempted by Kramer [47], where he also showed that feedback increases the capacity region of strong interference channels.

Following these results on the white Gaussian noise channel on hand, the next focus was on the feedback capacity of the colored Gaussian noise channel. Butman [14], [18] extended the Schalkwijk–Kailath coding scheme to autoregressive noise channels. Subsequently, Tiernan and Schalkwijk [17], [16], Wolfowitz [15], Ozarow [19], [20], Dembo [48], and Yang *et al.* [22] studied the feedback capacity of finite-order ARMA additive Gaussian noise channels and obtained many interesting upper and lower bounds. Using an asymptotic equipartition theorem for nonstationary nonergodic Gaussian noise processes, Cover and Pombra [2] obtained the n-block capacity (2) for the arbitrary colored Gaussian channel with or without feedback. (We can take $B = 0$ in (2) for the nonfeedback case.) Using matrix inequalities, they also showed that feedback does not increase the capacity much; namely, feedback at most doubles the capacity (a result obtained by Pinsker [49] and Ebert [50]), and feedback increases the capacity at most by half a bit.

The extensions and refinements of the result by Cover and Pombra abound. Dembo [48] showed that feedback does not increase the capacity at very low signal-to-noise ratio or very high signal-to-noise ratio. As mentioned above, Ordentlich [21] examined the properties of the optimal solution (K_V, B) in (2) and found the rank condition on the optimal K_V for finite-order MA noise processes. Chen and Yanagi [51]–[53] studied Cover's conjecture [54] that the feedback capacity is at most as large as the nonfeedback capacity with twice the power, and made several refinements on the upper bounds by Cover and Pombra. Thomas [55], Pombra and Cover [56], and Ordentlich [57] extended the factor-of-two bound result to the colored Gaussian multiple access channels with feedback. Recently Yang, Kavčić, and Tatikonda [22] revived the control-theoretic approach (cf. [33]) to the stationary ARMA(k) Gaussian feedback capacity problem. Although one-sentence summary would not do justice to their contribution, Yang *et al.* reformulated the feedback capacity problem as a stochastic control problem and used dynamic programming for the numerical computation of the n-block feedback capacity. In a series of papers [58]–[60], Ihara obtained coding theorems for continuous-time Gaussian channels with feedback and showed that the factor-of-two bound on the feedback capacity is tight by considering cleverly constructed nonstationary channels both in discrete time [61] and continuous time [59]. (See also [65, Examples 5.7.2 and 6.8.1].) In fact, besides the white Gaussian noise channel, Ihara's example has been the only nontrivial channel with known closed-form feedback capacity.

Hence Theorem 1 provides the first feedback capacity result on stationary colored Gaussian channels. Moreover, as will be discussed in Section IV, a simple linear signaling scheme similar to the Schalkwijk–Kailath scheme achieves the feedback capacity. This result links the Cover–Pombra formulation of the feedback capacity with the Schalkwijk–Kailath scheme and its generalizations to stationary colored channels, and strongly suggests the optimality[2] of the achievable rate for the AR(1) channel obtained by Butman [14] (cf. Proposition 1 in Section IV).

II. First-Order Moving Average Gaussian Processes

In this section, we digress a little to review a few characteristics of first-order moving average Gaussian processes. First, we give three alternative characterizations of Gaussian MA(1) processes. As defined in the previous section, the Gaussian MA(1) noise process $\{Z_i\}_{i=1}^{\infty}$ with parameter α can be characterized as

$$Z_i = \alpha U_{i-1} + U_i \tag{8}$$

where the innovations $U_0, U_1, \dots$ are i.i.d. $\sim N(0,1)$.

We reinterpret the above definition in (8) by regarding the noise process $\{Z_i\}$ as the output of the linear time-invariant filter with transfer function

$$H(z) = 1 + \alpha z^{-1} \tag{9}$$

which is driven by the white innovation process $\{U_i\}$. Thus we alternatively characterize the Gaussian MA(1) noise process $\{Z_i\}$ with parameter α and unit innovation through its power spectral density $S_Z(\omega)$ given by

$$S_Z(\omega) = |1 + \alpha e^{-j\omega}|^2 = 1 + \alpha^2 + 2\alpha\cos\omega. \tag{10}$$

We can further identify the power spectral density $S_Z(\omega)$ with the infinite Toeplitz covariance matrix of a Gaussian process. Thus, we can define $\{Z_i\}$ as $(Z_1, \dots, Z_n) \sim N_n(0, K_Z)$ for each finite horizon n, where K_Z is tridiagonal with

$$K_Z = \begin{bmatrix} 1+\alpha^2 & \alpha & 0 & \cdots & 0 \\ \alpha & 1+\alpha^2 & \alpha & \ddots & \vdots \\ 0 & \alpha & 1+\alpha^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \alpha \\ 0 & \cdots & 0 & \alpha & 1+\alpha^2 \end{bmatrix}$$

or equivalently

$$[K_Z]_{i,j} = \begin{cases} 1+\alpha^2, & |i-j| = 0 \\ \alpha, & |i-j| = 1 \\ 0, & |i-j| \ge 2. \end{cases}$$

Note that this covariance matrix K_Z is consistent with our initial definition of the MA(1) process given in (8). Thus all three definitions of the MA(1) process given above are equivalent. As we will see in the next section, the special structure of the MA(1) process, especially the tri-diagonality of the covariance matrix, makes the maximization in (2) easier than the generic case.

We will need the entropy rate of the MA(1) Gaussian process later in our discussion. As shown by Kolmogorov (see [1, Sec. 11.6]), the entropy rate of a stationary Gaussian process with power spectral density $S(\omega)$ can be expressed as

$$\frac{1}{4\pi}\int_{-\pi}^{\pi} \log(2\pi e S(\omega))\, d\omega.$$

We can calculate the above integral with the power spectral density $S_Z(\omega)$ in (10) by Jensen's formula[3] [65, Th. 15.18]

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log|e^{j\omega} - \alpha| d\omega = \begin{cases} 0, & |\alpha| \le 1 \\ \log|\alpha|, & |\alpha| > 1 \end{cases} \tag{11}$$

[2]At the time of this submission, extensions of these results are developed in a paper in preparation [63], [64], which confirm the optimality of the Schalkwijk–Kailath–Butman coding scheme for the AR(1) channel.

[3]The same J. L. W. V. Jensen, famous for his inequality on convex functions.

and obtain the entropy rate of the MA(1) Gaussian process (8) as

$$\begin{aligned}\frac{1}{4\pi}\int_{-\pi}^{\pi}\log\left(2\pi e S_Z(\omega)\right)\,d\omega &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\log\left(2\pi e|1+\alpha e^{-j\omega}|^2\right)\,d\omega \\ &= \begin{cases}\frac{1}{2}\log(2\pi e), & |\alpha|\le 1\\ \frac{1}{2}\log(2\pi e\alpha^2), & |\alpha|>1.\end{cases}\end{aligned} \tag{12}$$

(One can alternatively deal with the determinant of $K_Z(n)$ directly by a simple recursion. For example, we can show that $\det K_Z(n) = n+1$ for $|\alpha| = 1$.) For a more general discussion on the entropy rate of stationary Gaussian processes, refer to [62, Ch. 2].

We finish our digression by noting a certain reciprocal relationship between the Gaussian MA(1) process with parameter α and the Gaussian AR(1) process with parameter $-\alpha$. We can define the Gaussian AR(1) process $\{Z_i\}_{i=1}^{\infty}$ with parameter $-\alpha$, $|\alpha| < 1$, as

$$Z_i = -\alpha Z_{i-1} + U_i$$

where the innovations $U_1, U_2, \ldots$ are i.i.d. $\sim N(0,1)$ and $Z_0 \sim N(0, 1/(1-\alpha^2))$ is independent of $\{U_i\}_{i=1}^{\infty}$. Equivalently, we can define the above process as the output of the linear time-invariant filter with transfer function

$$G(z) = \frac{1}{1+\alpha z^{-1}} = \frac{1}{H(z)}$$

where $H(z)$ is the transfer function (9) of the MA(1) process with parameter α. This reciprocity is indeed reflected in the striking similarity between the fourth-order polynomial (5) for the capacity of the Gaussian MA(1) noise channel and the fourth-order polynomial (7) for the best known achievable rate of the Gaussian AR(1) noise channel.

III. Proof of Theorem 1

We will first transform the optimization problem

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

to a series of (asymptotically) equivalent forms. Then we solve the problem by imposing individual power constraints $(P_1,\ldots,P_n)$ on each input signal. Subsequently we optimize over $(P_1,\ldots,P_n)$ under the average power constraint

$$P_1 + \cdots + P_n \le nP.$$

Then using Lemma 2, we will prove that the uniform power allocation $P_1 = \cdots = P_n = P$ is asymptotically optimal. This leads 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:

$$\text{maximize} \quad \log\det((B+I)K_Z(B+I)^T + K_V) \tag{13}$$

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

$$H_Z = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0\\ \alpha & 1 & 0 & \cdots & 0\\ 0 & \alpha & 1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \alpha & 1\end{bmatrix}.$$

This matrix K_Z' is a covariance matrix of the Gaussian process $\{\tilde{Z}_i\}_{i=0}^{\infty}$ defined by

$$\begin{aligned}\tilde{Z}_1 &= U_1\\ \tilde{Z}_i &= U_i + \alpha U_{i-1}, \quad i = 2,3,\ldots\end{aligned}$$

where $\{U_i\}_{i=1}^{\infty}$ is the white Gaussian process with unit variance. It is easy to check that $K_Z \succeq K_Z'$ (i.e., $K_Z - K_Z'$ is nonnegative definite) and that the difference between K_Z and K_Z' is given by

$$[K_Z - K_Z']_{i,j} = \begin{cases}\alpha^2, & i=j=1\\ 0, & \text{otherwise.}\end{cases}$$

It is intuitively clear that there is no asymptotic difference in capacity between the channel with the original noise covariance K_Z and the channel with noise covariance K_Z'. We will prove this claim more rigorously in Appendix A. Throughout we will assume that the noise covariance matrix of the given channel is K_Z', which is equivalent to the statement that the time-zero noise innovation U_0 is revealed to both the transmitter and the receiver.

Now by identifying $K_V = F_V F_V^T$ for some lower-triangular F_V and identifying $F_Z = BH_Z$ for some strictly lower-triangular F_Z, we transform the optimization problem (13) into

$$\begin{aligned}&\text{maximize} && \log\det(F_VF_V^T + (F_Z+H_Z)(F_Z+H_Z)^T)\\ &\text{subject to} && \mathrm{tr}(F_VF_V^T + F_ZF_Z^T) \le nP\end{aligned} \tag{14}$$

with new variables (F_V, F_Z).

We shall use $2n$-dimensional row vectors f_i and h_i, $i = 1,\ldots,n$, 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, $i = 1,\ldots,n$, 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 identify Y_i with $f_i + h_i$. We also introduce new variables $(P_1,\ldots,P_n)$ representing the power constraint for each input f_i. Now the optimization problem in (13) becomes the following equivalent form:

$$\begin{aligned}&\text{maximize} && \log\det((F+H)(F+H)^T)\\ &\text{subject to} && \|f_i\|^2 \le P_i, \quad i=1,\ldots,n\\ &&& \sum_{i=1}^{n} P_i \le nP.\end{aligned} \tag{15}$$

Here $\|\cdot\|$ denotes the Euclidean norm of a $2n$-dimensional vector. Note that the variables $f_1,\ldots,f_n$ should satisfy $f_i \in \mathcal{V}_i,\ i=1,\ldots,n$, where

$$\mathcal{V}_i := \{(v_1,\ldots,v_{2n}) \in \mathbb{R}^{2n} : v_{i+1} = \cdots = v_n = 0 = v_{n+i} = \cdots = v_{2n}\}.$$

Step 2. Optimization Under the Individual Power Constraint for Each Signal: We solve the optimization problem (15) in $(f_1,\ldots,f_n)$ after fixing $(P_1,\ldots,P_n)$. This step is mostly algebraic, but we can easily give a geometric interpretation. We need some notation first.

We define an n-by-$2n$ matrix

$$S = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} := \begin{bmatrix} f_1+h_1 \\ \vdots \\ f_n+h_n \end{bmatrix} = F+H$$

and we define the n-by-$2n$ matrix E by

$$E = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} := [0_{n\times n}\ \ I_n]$$

where I_n is the identity. We also define an n-by-$2n$ matrix

$$G = \begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix} := \begin{bmatrix} h_1-e_1 \\ \vdots \\ h_n-e_n \end{bmatrix} = H-E.$$

We can identify the row vector e_i with the noise innovation U_i and the row vector g_i with $Z_i - U_i$.

We will use the notation F_k to denote the k-by-$2n$ submatrix of F which consists of the first k rows of F, that is,

$$F_k = \begin{bmatrix} f_1 \\ \vdots \\ f_k \end{bmatrix}.$$

We will use similar notation for the k-by-$2n$ submatrices of G, H, E, and S.

We now introduce a sequence of $2n$-by-$2n$ square matrices $\{\Pi_k\}_{k=1}^{n-1}$ as

$$\Pi_k = I - S_k^T\left(S_kS_k^T\right)^{-1}S_k.$$

Observe that S_k is of full rank and thus that $\left(S_kS_k^T\right)^{-1}$ always exists. We can view Π_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 $s_1,\ldots,s_k$ of S_k. (Or Π_k maps a generic random variable A to $A - E(A|Y^k)$.) It is easy to verify that $\Pi_k = \Pi_k^T = \Pi_k\Pi_k$ and $\Pi_kS_k^T = 0$.

Finally we define the intermediate objective functions of the maximization (15) as

$$J_k(P_1,\ldots,P_k) := \max_{\substack{f_1,\ldots,f_k \\ \|f_i\|^2 \le P_i}} \log\det\left(S_kS_k^T\right), \quad k=1,\ldots,n$$

so that

$$C_{n,\mathrm{FB}} = \max_{P_i:\sum P_i \le nP} \frac{1}{2n} J_n(P_1,\ldots,P_n).$$

We will show that if the $k-1$ rows $(f_1^*,\ldots,f_{k-1}^*)$ maximizes $J_{k-1}(P_1,\ldots,P_{k-1})$, then $(f_1^*,\ldots,f_{k-1}^*,f_k^*)$ maximizes $J_k(P_1,\ldots,P_k)$ for some f_k^* satisfying $f_k^* = f_k^*\Pi_{k-1}$. Thus the maximization for J_n can be solved in a greedy fashion by sequentially maximizing $J_1, J_2, \ldots, J_n$ through $f_1^*, f_2^*, \ldots, f_n^*$. Furthermore, we will obtain the recursive relationship

$$J_0 := 0 \tag{16}$$

$$J_1 = \log(1+P_1) \tag{17}$$

and

$$J_{k+1} = J_k + \log\left(1+\left(\sqrt{P_{k+1}} + |\alpha|\sqrt{1-\frac{1}{e^{J_k-J_{k-1}}}}\right)^2\right) \tag{18}$$

for $k=1,2,\ldots$.

We need the following result to proceed to the actual maximization.

Lemma 1: Suppose $P \ge 0$ and $1 \le k \le n-1$. Suppose S_k and Π_k defined as above. Let $\mathcal{V}$ be an arbitrary subspace of $\mathbb{R}^{2n}$ such that $\mathcal{V}$ is not contained in the span of $s_1,\ldots,s_k$. Then, for any $w \in \mathcal{V}$

$$\max_{v\in\mathcal{V}:\|v\|^2\le P} (v+w)\Pi_k(v+w)^T = (\sqrt{P} + \|w\Pi_k\|)^2.$$

Furthermore, if $w\Pi_k \ne 0$, the maximum is attained by

$$v^* = \sqrt{P}\frac{w\Pi_k}{\|w\Pi_k\|}. \tag{19}$$

Proof: When $w\Pi_k = 0$, that is, $w \in \mathrm{span}\{s_1,\ldots,s_k\}$, the maximum of $(v+w)\Pi_k(v+w)^T = v\Pi_kv^T$ is attained by any vector v, $\|v\|^2 = P$, orthogonal to $\mathrm{span}\{s_1,\ldots,s_k\}$, and we trivially have

$$\max_{v\in\mathcal{V}:\|v\|^2\le P} v\Pi_kv^T = P.$$

When $w\Pi_k \ne 0$, we have

$$\begin{aligned}(v+w)\Pi_k(v+w)^T &= \|(v+w\Pi_k)\Pi_k\|^2 \\ &\le \|v+w\Pi_k\|^2 \\ &\le (\sqrt{P}+\|w\Pi_k\|)^2\end{aligned}$$

where the first inequality follows from the fact that $I - \Pi_k$ is nonnegative definite. It is easy to check that we have equality if v is given by (19) (see Fig. 2). □

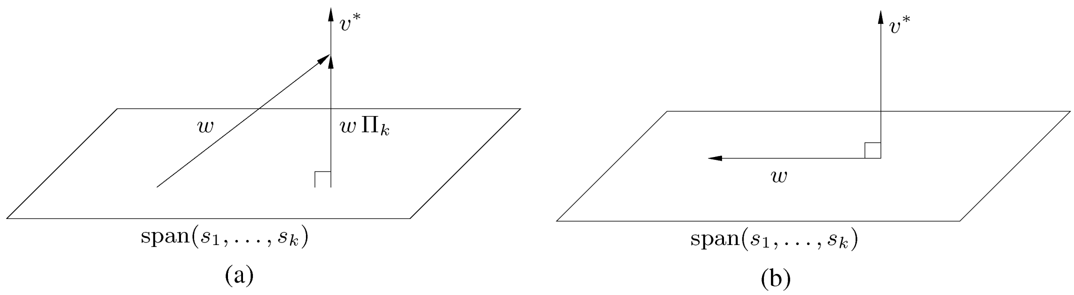

Fig. 2. Geometric interpretation of Lemma 1. (a) The case $w\Pi_k \neq 0$. (b) The case $w\Pi_k = 0$.

We observe that, for $k = 2,\ldots,n$

$$\begin{aligned}
&\det\left(S_k S_k^T\right)\\
&= \det\left(\begin{bmatrix} S_{k-1} \\ s_k \end{bmatrix}\begin{bmatrix} S_{k-1} \\ s_k \end{bmatrix}^T\right)\\
&= \det\begin{bmatrix} S_{k-1}S_{k-1}^T & S_{k-1}s_k^T \\ s_k S_{k-1}^T & s_k s_k^T \end{bmatrix}\\
&= \det(S_{k-1}S_{k-1}^T)\cdot s_k(I - S_{k-1}^T(S_{k-1}S_{k-1}^T)^{-1}S_{k-1})s_k^T\\
&= \det(S_{k-1}S_{k-1}^T)\cdot s_k\Pi_{k-1}s_k^T\\
&= \det(S_{k-1}S_{k-1}^T)\cdot (f_k+g_k+e_k)\Pi_{k-1}(f_k+g_k+e_k)^T\\
&= \det(S_{k-1}S_{k-1}^T)\cdot \left[1 + (f_k+g_k)\Pi_{k-1}(f_k+g_k)^T\right] \quad (20)
\end{aligned}$$

where the last equality follows since $e_k\Pi_{k-1} = e_k, e_k e_k^T = 1$, and $e_k g_k^T = e_k f_k^T = 0$. Now fix $f_1,\ldots,f_{k-1}$. Since $\mathcal{V}_k$ is not contained in $\mathrm{span}\{s_1,\ldots,s_{k-1}\}$ and $g_k \in \mathcal{V}_k$, we have from the above lemma and (20) that

$$\begin{aligned}
&\max_{f_k:\|f_k\|^2\le P_k} \det\left(S_k S_k^T\right)\\
&\qquad = \det(S_{k-1}S_{k-1}^T)\cdot(1 + (\sqrt{P_k} + \|g_k\Pi_{k-1}\|)^2). \quad (21)
\end{aligned}$$

If $\alpha \neq 0$, the maximum of is attained by

$$f_k^* = \sqrt{P_k}\frac{g_k\Pi_{k-1}}{\|g_k\Pi_{k-1}\|}. \quad (22)$$

In the special case $\alpha = 0$, that is, when the noise is white, we trivially have

$$\max_{f_k:\|f_k\|^2\le P_k} \det\left(S_k S_k^T\right) = \det\left(S_{k-1}S_{k-1}^T\right)\cdot(1+P_k)$$

which immediately implies that $J_k = J_{k-1} + \log(1+P_k) = \sum_{i=1}^k \log(1+P_i)$ which, in turn, combined with the concavity of the logarithm, implies that

$$C_{n,\mathrm{FB}} = C_{\mathrm{FB}} = \frac{1}{2}\log(1+P).$$

We continue our discussion throughout this step under the assumption $\alpha \neq 0$. Until this point we have not used the special structure of the MA(1) noise process. Now we rely heavily on this. We trivially have

$$J_1 = \max_{f_1}\log\left(s_1 s_1^T\right) = \log(1+P_1). \quad (23)$$

Following (21), we have, for $k = 2,\ldots,n$

$$\begin{aligned}
J_k = \max_{f_1,\ldots,f_{k-1}} \Big[&\log\det(S_{k-1}S_{k-1}^T)\\
&+ \log(1 + \left(\sqrt{P_k} + \|g_k\Pi_{k-1}\|\right)^2)\Big]. \quad (24)
\end{aligned}$$

We wish to show that both terms in (24) are individually maximized by the same optimizer

$$\begin{aligned}
(f_1^*,\ldots,f_{k-1}^*) &= \arg\max\left(\det(S_{k-1}S_{k-1}^T)\right)\\
&= \arg\max \|g_k\Pi_{k-1}\| \quad (25)
\end{aligned}$$

for $k = 2,\ldots,n$. Once we establish (25), the desired recursion formula (18) for J_k follows immediately from the definition of J_k and (24).

We shall prove (25) by induction. First note that

$$\begin{aligned}
&g_1 = 0\\
&g_k = \alpha e_{k-1}, \qquad k = 2,3,\ldots \quad (26)
\end{aligned}$$

and

$$e_k s_k^T = 1, \qquad k = 1,2,\ldots. \quad (27)$$

Also recall that $s_k = f_k + g_k + e_k$ and

$$e_k\Pi_{k-1} = e_k. \quad (28)$$

For $k = 2$, we trivially have

$$\begin{aligned}
\|g_2\Pi_1\|^2 &= \alpha^2 e_1\left(I - \frac{s_1^T s_1}{s_1 s_1^T}\right)e_1^T\\
&= \alpha^2\left(1 - \frac{1}{s_1 s_1^T}\right)\\
&= \alpha^2\left(1 - \frac{1}{\det\left(s_1 s_1^T\right)}\right)
\end{aligned}$$

which establishes (25). Further, from (23) and (24), we can check that

$$\begin{aligned}
J_2 = \max_{f_1}\Bigg[&\log(s_1 s_1^T)\\
&+ \log\left(1 + \left(\sqrt{P_2} + |\alpha|\sqrt{1 - \frac{1}{s_1 s_1^T}}\right)^2\right)\Bigg]\\
= J_1 &+ \log\left(1 + \left(\sqrt{P_2} + |\alpha|\sqrt{1 - \frac{1}{e^{J_1}}}\right)^2\right).
\end{aligned}$$

Now suppose (25) holds for $k = 2, \ldots, m-1$. For $k \geq 3$, we observe that

$$\begin{aligned}
&\Pi_{k-1}\\
&= I - S_{k-1}^T (S_{k-1} S_{k-1}^T)^{-1} S_{k-1}\\
&= I - \begin{bmatrix} S_{k-2} \\ s_{k-1} \end{bmatrix}^T \begin{bmatrix} S_{k-2} S_{k-2}^T & S_{k-2} s_{k-1}^T \\ s_{k-1} S_{k-2}^T & s_{k-1} s_{k-1}^T \end{bmatrix}^{-1} \begin{bmatrix} S_{k-2} \\ s_{k-1} \end{bmatrix}\\
&= I - S_{k-2}^T \left(S_{k-2} S_{k-2}^T\right)^{-1} S_{k-2}\\
&\quad - \Pi_{k-2} s_{k-1}^T \left(s_{k-1} \Pi_{k-2} s_{k-1}^T\right)^{-1} s_{k-1} \Pi_{k-2}\\
&= \Pi_{k-2} (I - \Pi_{k-2} s_{k-1}^T \left(s_{k-1} \Pi_{k-2} s_{k-1}^T\right)^{-1} s_{k-1} \Pi_{k-2})\\
&\quad \times \Pi_{k-2}.
\end{aligned}$$

Now from (26), (27), and (28), we have

$$\begin{aligned}
&\|g_k \Pi_{k-1}\|^2\\
&= g_k \Pi_{k-1} g_k^T\\
&= g_k \Pi_{k-2} \left(I - \Pi_{k-2} s_{k-1}^T (s_{k-1} \Pi_{k-2} s_{k-1}^T)^{-1} s_{k-1} \Pi_{k-2}\right)\\
&\quad \times \Pi_{k-2} g_k^T\\
&= \alpha^2 e_{k-1} \left(I - \Pi_{k-2} s_{k-1}^T (s_{k-1}\right.\\
&\quad \left. \times \Pi_{k-2} s_{k-1}^T)^{-1} s_{k-1} \Pi_{k-2}\right) e_{k-1}^T\\
&= \alpha^2 \left(1 - \frac{1}{s_{k-1} \Pi_{k-2} s_{k-1}^T}\right)\\
&= \alpha^2 \left(1 - \frac{1}{1 + (f_{k-1} + g_{k-1}) \Pi_{k-2} (f_{k-1} + g_{k-1})^T}\right). \quad (29)
\end{aligned}$$

It follows from (20)–(22) that, for fixed $(f_1, \ldots, f_{m-2})$, both $\det(S_{m-1} S_{m-1}^T)$ and $\|g_m \Pi_{m-1}\|$ have the same maximizer

$$f_{m-1}^* = \sqrt{P_{m-1}} \frac{g_{m-1} \Pi_{m-2}}{\|g_{m-1} \Pi_{m-2}\|}.$$

Plugging this back to (29), for fixed $(f_1, \ldots, f_{m-2})$, we have

$$\begin{aligned}
&\max_{f_{m-1}} \|g_m \Pi_{m-1}\|^2\\
&\qquad = \alpha^2 \left(1 - \frac{1}{1 + (\sqrt{P_{m-1}} + \|g_{m-1} \Pi_{m-2}\|)^2}\right)
\end{aligned}$$

while

$$\begin{aligned}
&\max_{f_{m-1}} \det\left(S_{m-1} S_{m-1}^T\right)\\
&\qquad = \det(S_{m-2} S_{m-2}^T) \cdot (1 + (\sqrt{P_{m-1}} + \|g_{m-1} \Pi_{m-2}\|)^2).
\end{aligned}$$

But from the induction hypothesis, $\det(S_{m-2} S_{m-2}^T)$ and $\|g_{m-1} \Pi_{m-2}\|$ have the same maximizer $(f_1^*, \ldots, f_{m-2}^*)$. Thus $\det(S_{m-1} S_{m-1}^T)$ and $\|g_m \Pi_{m-1}\|$ have the same maximizer $(f_1^*, \ldots, f_{m-1}^*)$. Therefore, we have established (25) for $k = m$ and hence for all $k = 2, 3, \ldots$. From (24) and (25), we easily get the desired recursion formula as

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

for $k = 2, 3, \ldots$.

Step 3. Optimal Power Allocation Over Time: In the previous step, we solved the optimization problem (15) under a fixed power allocation $(P_1, \ldots, P_n)$. 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 $(P_1, \ldots, P_n)$ under the constraint $\sum_{i=1}^n P_i \leq nP$.

As we saw earlier, when $\alpha = 0$, we can use the concavity of the logarithm to show that, for all n

$$\begin{aligned}
C_{n,\mathrm{FB}} &= \frac{1}{2n} J_n(P_1, \ldots, P_n)\\
&= \max_{P_i : \sum_i P_i \leq nP} \frac{1}{2n} \sum_{i=1}^n \log(1 + P_i)\\
&= \frac{1}{2} \log(1 + P)
\end{aligned}$$

with $P_1^* = \cdots = P_n^* = P$. When $\alpha \neq 0$, it is not tractable to optimize $(P_1, \ldots, P_n)$ for J_n in (16)–(18) to get a closed-form solution of $C_{n,\mathrm{FB}}$ 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 $C_{\mathrm{FB}} = \lim_n C_{n,\mathrm{FB}}$.

Lemma 2: Let $\psi : [0, \infty) \times [0, \infty) \to [0, \infty)$ such that the following conditions hold:

i) $\psi(\xi, \zeta)$ is continuous, concave in (ξ, ζ), and strictly concave in ξ for all $\zeta > 0$;
ii) $\psi(\xi, \zeta)$ is increasing in ξ and ζ, respectively; and
iii) for each $\zeta > 0$, there is a unique solution $\xi^*(\zeta) > 0$ 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

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^n P_i \leq P.$$

Let $\{\xi_i\}_{i=0}^\infty$ be defined recursively as

$$\begin{aligned}
&\xi_0 = 0\\
&\xi_i = \psi(\xi_{i-1}, P_i), \quad i = 1, 2, \ldots.
\end{aligned}$$

Then

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

where $\xi^* = \xi^*(P)$ is the unique solution to $\xi = \psi(\xi, P)$. Furthermore, if $P_i \equiv P, \ i = 1, 2, \ldots$, then the corresponding ξ_i converges to ξ^*.

Proof: Fix $\epsilon > 0$. From the concavity and monotonicity of ψ, for n sufficiently large

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \xi_i &= \frac{1}{n} \sum_{i=1}^n \psi(\xi_{i-1}, P_i)\\
&\leq \psi\left(\frac{1}{n} \sum_{i=1}^n \xi_{i-1}, \frac{1}{n} \sum_{i=1}^n P_i\right)\\
&\leq \psi\left(\frac{1}{n} \sum_{i=1}^n \xi_{i-1}, P + \epsilon\right).
\end{aligned}$$

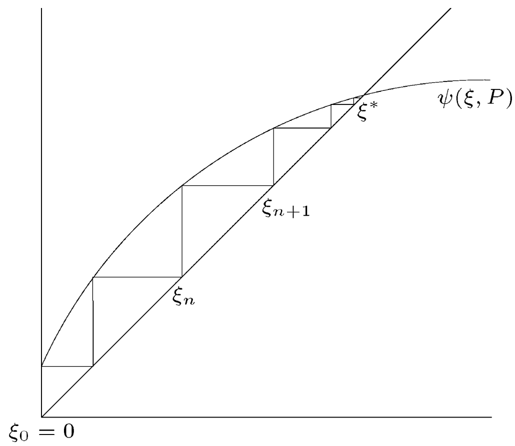

Fig. 3. Convergence to the unique point ξ^*.

Taking the limit on both sides and using the continuity of ψ, we have

$$\begin{aligned}\xi^{**} &:= \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \xi_i \\ &\le \limsup_{n\to\infty} \psi\left(\frac{1}{n}\sum_{i=1}^n \xi_{i-1}, P+\epsilon\right) \\ &= \psi(\xi^{**}, P+\epsilon).\end{aligned}$$

Since ϵ is arbitrary and ψ is continuous, we have $\xi^{**} \le \psi(\xi^{**}, P)$. But from uniqueness of ξ^* and strict concavity of ψ in ξ, we have

$$\xi \le \xi^* \text{ if and only if } \xi \le \psi(\xi, P). \tag{30}$$

Thus $\xi^{**} \le \xi^*$.

It remains to show that we can actually attain ξ^* by choosing $P_i \equiv P$, $i = 1, 2, \ldots$. Let $\xi_i = \psi(\xi_{i-1}, P)$, $i = 1, 2, \ldots$. From the monotonicity of $\psi(\cdot, P)$ and (30), we have

$$\xi_{i-1} \le \xi_i = \psi(\xi_{i-1}, P) \le \xi^* = \psi(\xi^*, P) \quad \text{for all } i.$$

Thus the sequence $\{\xi_i\}$ has a limit, which we denote as ξ^{**}. But from the continuity of $\psi(\cdot, P)$, we must have

$$\begin{aligned}\xi^{**} = \lim_{n\to\infty} \xi_n &= \lim_{n\to\infty} \psi(\xi_n, P) \\ &= \psi(\lim_{n\to\infty} \xi_n, P) = \psi(\xi^{**}, P).\end{aligned}$$

Thus $\xi^{**} = \xi^*$ (see Fig. 3). □

We continue our main discussion. 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. For concavity, we rely on the simple composition rule for concave functions [66, Sec. 3.2.4] without messy calculus. Let $\psi_1(\xi) = \frac{1}{2}\log(1+\xi)$, $\psi_2(\xi, \zeta) = (\sqrt{\xi} + \sqrt{\zeta})^2$, and $\psi_3(\xi) = |\alpha|^2(1 - \exp(-2\xi))$. Then $\psi(\xi, \zeta) = \psi_1(\psi_2(\psi_3(\xi), \zeta))$. Now that ψ_1 is strictly concave and strictly increasing, ψ_2 is concave (strictly concave in ξ alone for each $\zeta > 0$) and elementwise strictly increasing, and ψ_3 is strictly concave, we can conclude that ψ is concave in (ξ, ζ) and strictly concave in ξ for all $\zeta > 0$. Since for any $\zeta > 0$, $\psi(0, \zeta) > 0$ and $\psi(\xi, \zeta) \to c(\zeta) < \infty$ as ξ tends to infinity, the uniqueness of the root of $\xi = \psi(\xi, \zeta)$ is trivial from the continuity of ψ.

For an arbitrary infinite sequence $\{P_i\}_{i=1}^\infty$ satisfying

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n P_i \le nP \tag{31}$$

we define

$$\begin{aligned}&\xi_0 = 0 \\ &\xi_i = \psi(\xi_{i-1}, P_i), \quad i = 1, 2, \ldots.\end{aligned}$$

Note that

$$\begin{aligned}\xi_1 &= \frac{1}{2}J_1(P_1) \\ \xi_i &= \frac{1}{2}\big(J_i(P_1, \ldots, P_i) - J_{i-1}(P_1, \ldots, P_{i-1})\big), \quad i \ge 2.\end{aligned}$$

Now from Lemma 2, we have

$$\limsup_{n\to\infty} \frac{1}{2n} J_n(P_1, \ldots, P_n) = \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \xi_i \le \xi^*$$

where ξ^* is the unique solution to

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

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

$$\begin{aligned}\sup \limsup_{n\to\infty} \frac{1}{2n} J_n(P_1, \ldots, P_n) &= \lim_{n\to\infty} \frac{1}{2n} J_n(P, \ldots, P) \\ &= \xi^*\end{aligned}$$

where the supremum (in fact, maximum) is taken over all infinite sequences $\{P_i\}$ satisfying the asymptotic average power constraint (31).

Finally, we prove that $C_{\mathrm{FB}} = \xi^*$. More specifically, we will show that

$$\begin{aligned}C_{\mathrm{FB}} &= \lim_{n\to\infty} C_{n,\mathrm{FB}} \\ &= \lim_{n\to\infty} \max_{P_i : \sum_i P_i \le nP} \frac{1}{2n} J_n(P_1, \ldots, P_n) \qquad (32) \\ &= \sup_{\{P_i\}_{i=1}^\infty} \limsup_{n\to\infty} \frac{1}{2n} J_n(P_1, \ldots, P_n) \qquad (33) \\ &= \xi^*.\end{aligned}$$

The only subtlety here is how to justify the interchange of the order of limit and supremum in (32) and (33). It is easy to verify that

$$\lim_{n\to\infty} \max_{P_i:\sum_i P_i\le nP} \frac{1}{2n} J_n(P_1,\ldots,P_n) \ge \sup_{\{P_i\}_{i=1}^\infty} \limsup_{n\to\infty} \frac{1}{2n} J_n(P_1,\ldots,P_n)$$

for it is always advantageous to choose for each n a finite sequence $(P_1,\ldots,P_n)$ with $\sum_{i=1}^n P_i \le nP$ rather than fixing a single infinite sequence $\{P_i\}$ with $P_i = P$ for all i. (Recall that the supremum on the right side is achieved by the uniform power allocation.)

To prove the other direction of inequality, we fix $\epsilon > 0$, and choose n and $(P_1^*,\ldots,P_n^*)$ such that

$$\sum_{i=1}^n P_i^* \le nP$$

and

$$\frac{1}{2n} J_n(P_1^*,\ldots,P_n^*) \ge C_{\rm FB} - \epsilon. \tag{34}$$

Now we construct an infinite sequence $\{P_i\}_{i=1}^\infty$ by concatenating $(P_1^*,\ldots,P_n^*)$ repeatedly, that is, $P_{kn+i} = P_i^*$ for all $i = 1,\ldots,n$, and $k = 0,1,\ldots$. Obviously, this choice of $\{P_i\}$ satisfies the power constraint (31). As before, let $\xi_i = \psi(\xi_{i-1}, P_i),\ i = 1,2,\ldots$. By induction, it is easy to see that

$$\xi_i \le \xi_{kn+i}, \quad i = 1,2,\ldots,n \tag{35}$$

for all $k = 0,1,\ldots$. For $k = 0$, (35) holds trivially. Suppose (35) holds for $k = 0,\ldots,m-1$. Then from the monotonicity of $\psi(\xi,\zeta)$ in ξ, we have

$$\begin{aligned}\xi_i &= \psi(\xi_{i-1}, P_i)\\ &= \psi(\xi_{i-1}, P_i^*)\\ &\le \psi(\xi_{mn+i-1}, P_i^*)\\ &= \psi(\xi_{mn+i-1}, P_{mn+i})\\ &= \xi_{mn+i}\end{aligned}$$

for all $i = 1,\ldots,n$. Thus, (35) holds for all k. Therefore

$$\begin{aligned}\frac{1}{2kn} J_{kn}(P_1,\ldots,P_{kn}) &= \frac{1}{kn}\sum_{i=1}^{kn} \xi_i\\ &\ge \frac{1}{kn}\left(k\cdot\sum_{i=1}^n \xi_i\right)\\ &= \frac{1}{2n} J_n(P_1,\ldots,P_n)\end{aligned}$$

which, combined with (34), implies that

$$\limsup_{n\to\infty} \frac{1}{2n} J_n(P_1,\ldots,P_n) \ge C_{\rm FB} - \epsilon, \quad k = 1,2,\ldots$$

which, in turn, implies that

$$\sup_{\{P_i\}_{i=1}^\infty} \limsup_{n\to\infty} \frac{1}{2n} J_n(P_1,\ldots,P_n) \ge C_{\rm FB} - \epsilon.$$

Since ϵ is arbitrary, we have the desired inequality. Thus $C_{\rm FB} = \xi^*$.

We conclude this section by characterizing the capacity $C_{\rm FB} = \xi^*$ in an alternative form. Recall that ξ^* is the unique solution to

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

Let $x_0 = \exp(-\xi^*)$, or equivalently, $\xi^* = -\log x_0$. It is easy to verify that $0 < x_0 \le 1$ is the unique positive solution to

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

or equivalently

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

This establishes the feedback capacity $C_{\rm FB}$ of the additive Gaussian noise channel with the noise covariance K_Z', which is, in turn, the feedback capacity of the first-order moving average additive Gaussian noise channel with parameter α, as is argued at the end of Step 1 and proved in Appendix A. This completes the proof of Theorem 1.

IV. 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 $(K_V^*(n), B^*(n))$ for the n-block can be found sequentially, built on the optimal $(K_V^*(n-1), B^*(n-1))$ 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 [21]. Second, the current input signal X_k is orthogonal to the past output signals $(Y_1,\ldots,Y_{k-1})$. In the notation of Section III, 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 (possibly nonstationary nonergodic) noise covariance matrix K_Z [67], [21]. It should be pointed out that the recursion formula (16)–(18) can be also derived from the orthogonality property and the optimality of rank-one K_V.

We explore the possibility of extending the current proof technique to a more general class of noise processes. The immediate 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:

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

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

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

It is not difficult to see that this rate is strictly less than the feedback capacity of the interleaved MA(1) channel unless $\alpha = 0$. A similar argument can prove that Butman's conjecture on the $\mathrm{AR}(k)$ capacity [18, 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, it is known [15], [16] that the optimal achievable rate is given by $-\log x_0$, where x_0 is the unique positive root of the fourth order polynomial

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

There is, however, a major difficulty in establishing the above two conditions by the two-stage optimization strategy we used in the previous section, namely, first maximizing $(f_1, \ldots, f_n)$ and then $(P_1, \ldots, P_n)$. For certain values of individual signal power constraints $(P_1, \ldots, P_n)$, the optimal $(f_1, \ldots, f_n)$ does not satisfy the sequentiality, resulting in K_V with rank higher than one. Hence, a greedy maximization of $\log \det(S_k S_k^T)$ does not establish the recursion formula for the AR(1) n-block capacity that corresponds to our (16)–(18):

$$J_0 := 0 \tag{36}$$

$$J_1 = \log(1 + P_1) \tag{37}$$

$$J_{k+1} = J_k + \log\left(1 + \left(\sqrt{P_{k+1}} + |\alpha|\sqrt{P_k} e^{-\frac{J_k - J_{k-1}}{2}}\right)^2\right) \tag{38}$$

for $k = 1, 2, \ldots$. (See [15], [16], and [18] for the derivation of the above recursion formula.) Even under the assumption that the optimal K_V for the AR(1) channel has rank one, it has been unclear whether the uniform power allocation over time is asymptotically optimal.

Nonetheless, using a technique similar to the one deployed in Lemma 2, we can prove the optimality of the uniform power allocation, resolving a question raised by Butman [14], [18] and Tiernan [16] among others.

Proposition 1: For the additive Gaussian AR(1) noise channel $Y_i = X_i + Z_i$, $i = 1, 2, \ldots$, with Gaussian AR(1) noise process $\{Z_i\}$ defined in (6), let R_n denote the best n-block achievable rate of the Schalkwijk–Kailath–Butman coding scheme under the power constraint P. Equivalently

$$\begin{aligned} R_n &= \max_{P_i : \sum_i P_i \le nP} \frac{1}{2n} J_n(P_1, \ldots, P_n) \\ &= \max_{K_V, B} \frac{1}{2n} \log \frac{\det((B+I)K_Z(B+I)^T + K_V)}{\det(K_Z)} \end{aligned}$$

where the maximization is over all nonnegative definite $n \times n$ matrices $K_V = K_V(n)$ of *rank one* and strictly lower triangular $n \times n$ matrices $B = B(n)$ such that $\mathrm{tr}(BK_Z B^T + K_V) \le nP$. Then

$$\lim_{n\to\infty} R_n = \lim_{n\to\infty} \frac{1}{2n} J_n(P, \ldots, P) = -\log x_0$$

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

$$Px^2 = \frac{(1 - x^2)}{(1 + |\alpha|x)^2}. \tag{39}$$

Since the proof is a little technical in nature, we defer it to Appendix B.

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. Before we proceed, we point out that the optimal input process $\{X_i\}$ we obtained in the previous section is asymptotically stationary. This observation is not hard to prove through the well-developed theory on the asymptotic behavior of recursive estimators [68, Ch. 14].

At the beginning, we send[4]

$$X_1 \sim N(0, P).$$

For subsequent transmissions, we transmit the filtered version of the noise innovation process up to the time $k - 1$:

$$X_k = \beta X_{k-1} + \sigma U_{k-1}, \quad k = 2, 3, \ldots. \tag{40}$$

In other words, we use a first-order regressive filter with transfer function given by

$$\frac{\sigma z^{-1}}{1 - \beta z^{-1}}. \tag{41}$$

Here $\beta = -\mathrm{sgn}(\alpha) x_0$ with x_0 being the same unique positive root of the fourth-order polynomial (5) in Theorem 1. The scaling factor σ is chosen to satisfy the power constraint as

$$\sigma = \mathrm{sgn}(\alpha)\sqrt{P(1 - \beta^2)}$$

[4]Technically, we generate 2^{nR} $X_1(W)$ code functions i.i.d. according to $N(0, P)$ for some $R < C_{\mathrm{FB}}$, and transmit one of them.

where

$$\operatorname{sgn}(\zeta) = \begin{cases} 1, & \zeta \geq 0 \\ -1, & \zeta < 0. \end{cases}$$

This input process and the MA(1) noise process

$$Z_k = \alpha U_{k-1} + U_k, \quad k = 1, 2, \ldots$$

yield the output process given by

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

for $k = 2, 3, \ldots$, which is asymptotically stationary with power spectral density

$$\begin{aligned} S_Y(\omega) &= \left| 1 + \alpha e^{-j\omega} + \frac{\sigma e^{-j\omega}}{1 - \beta e^{-j\omega}} \right|^2 \\ &= \left| \frac{1 + (\alpha - \beta + \sigma) e^{-j\omega} - \alpha\beta e^{-j2\omega}}{(1 - \beta e^{-j\omega})} \right|^2 \\ &= \left| \frac{(1 + \alpha\beta^2 e^{-j\omega})(1 - \beta^{-1} e^{-j\omega})}{(1 - \beta e^{-j\omega})} \right|^2 \\ &= \beta^{-2} |1 + \alpha\beta^2 e^{-j\omega}|^2. \end{aligned} \tag{42}$$

The "asymptotic stationarity" here should not bother us since the output process $\{Y_k\}_{k=2}^{\infty}$ is stationary for $k \geq 2$ and $h(Y_1 | Y_2, \ldots, Y_n)$ is uniformly bounded in n; hence the entropy rate of the process $\{Y_k\}_{k=1}^{\infty}$ is determined by $(Y_2, Y_3, \ldots)$. Thus from (12) in Section II, the entropy rate of the output process $\{Y_k\}$ is given by

$$\begin{aligned} \frac{1}{4\pi} \int_{-\pi}^{\pi} \log\left(2\pi e S_Y(\omega)\right) d\omega &= \frac{1}{2} \log(2\pi e \beta^{-2}) \\ &= \frac{1}{2} \log\left(2\pi e x_0^{-2}\right). \end{aligned}$$

Hence we attain the feedback capacity C_{FB}. Furthermore, it can be shown that the mean-square error of X_1 given the observations $Y_1, \ldots, Y_n$ decays exponentially with rate $\beta^{-2} = 2^{2C_{\mathrm{FB}}}$. In other words,

$$\begin{aligned} \operatorname{Var}(X_1 | Y_1, \ldots, Y_n) &= E(X_1 - E(X_1 | Y_1, \ldots, Y_n))^2 \\ &\doteq P 2^{-2nC_{\mathrm{FB}}}. \end{aligned} \tag{43}$$

Note that the optimal filter (41) has an interesting feature. In the light of (42), we can think of the output process $\{Y_k\}$ as the filtered version of the noise innovation process $\{U_k\}$ through the monic filter

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

As the entropy rate formula (12), or more fundamentally, Jensen's formula (11) shows, the entropy rate of $\{Y_k\}$ is totally determined by zeros of the filter outside the unit circle, which, for our case, is β^{-1}. Hence, we can interpret the feedback capacity problem as the problem of relocating the zero of the original noise filter $1 + \alpha z^{-1}$ to the outside of the unit circle and making the modulus of that zero as large as possible by adding a strictly causal filter $H(z)$ using the power $(2\pi)^{-1} \int |H(e^{-j\omega})|^2 d\omega = P$. Here we have shown that the optimal filter is given by (41). Under this interpretation, the initial input X_1 is merely a perturbation which guarantees that the output process is not causally invertible from the innovation process and hence that the entropy rate is fully determined by the spectral density of the stationary part. (Without X_1, the entropy rate of $\{Y_k\}$ is exactly same as the entropy rate of $\{Z_k\}$.)

From a classical viewpoint, we can interpret the signal X_k as the adjustment of the receiver's estimate of the message-bearing signal X_1 after observing $(Y_1, \ldots, Y_{k-1})$. We can further check that following signaling schemes are equivalent (and thus optimal) up to scaling:

$$\begin{aligned} X_k &\propto X_1 - \hat{X}_1(Y^{k-1}) \\ &\propto X_j - \hat{X}_j(Y^{k-1}) \qquad \text{for } j < k \\ &\propto U_{k-1} - \hat{U}_{k-1}(Y^{k-1}) \\ &\propto \hat{Z}_k(Y^{k-1}, X^{k-1}) - \hat{Z}_k(Y^{k-1}). \end{aligned}$$

The connection to the Schalkwijk–Kailath coding scheme is now apparent. Recall that there is a simple linear relationship [68, Sec. 3.4], [69, Sec. 4.5] 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 θ. Thus we can easily transform the above coding scheme based on the asymptotic equipartition property [2] to a variant of the Schalkwijk–Kailath 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, $\theta \in \Theta := \{-\sqrt{P}, -\sqrt{P} + \Delta, \ldots, \sqrt{P} - \Delta, \sqrt{P}\}$, where $\Delta = \frac{2\sqrt{P}}{2^{nR} - 1}$. Subsequent transmissions follow (40). The receiver forms the maximum likelihood estimate $\hat{\theta}_n(Y_1, \ldots, Y_n)$ and finds the nearest signal point to $\hat{\theta}_n$ in Θ.

The analysis of the error for this coding scheme follows Schalkwijk [5] and Butman [14]. From (43) 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 θ, Gaussian with mean θ and variance exponentially decaying with rate $\beta^{-2} = 2^{2nC_{\mathrm{FB}}}$. Thus, the nearest neighbor decoding error, ignoring lower order terms, is given by

$$\begin{aligned} P_e^{(n)} &= E_\theta \left[\Pr\left(|\hat{\theta}_n - \theta| \geq \frac{\Delta}{2} \,\middle|\, \theta \right) \right] \\ &\doteq \operatorname{erfc}\left(\sqrt{\frac{3}{2\sigma_\theta^2}} 2^{n(C_{\mathrm{FB}} - R)} \right) \end{aligned}$$

where

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2)\, dt$$

and σ_θ^2 is the variance of input signal θ chosen uniformly over Θ. As long as $R < C_{\rm FB}$, 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 $C_{\rm FB}$ is achievable for any additive non-Gaussian noise channel with feedback.

APPENDIX A
ASYMPTOTIC EQUIVALENCE OF K_Z AND K'_Z FOR FEEDBACK CAPACITY

Recall that $Z^n \sim N_n(0, K_Z)$ and $\tilde{Z}^n \sim N_n(0, K'_Z)$. To stress the dependence of the capacity on the power constraint and the noise covariance, we use the notation $C_{n,\rm FB}(K, P)$ for n-block feedback capacity of the channel with n-by-n noise covariance matrix K under the power constraint $E\sum_{i=1}^n X_i^2 \le nP$. With a little abuse of notation, we similarly use $C_{\rm FB}(K, P)$ for feedback capacity of the channel with infinite noise covariance matrix naturally extended from K under the power constraint P.

Suppose (B^*, K_V^*) achieves

$$\begin{aligned} &C_{n,\rm FB}(K_Z, P)\\ &\quad = \max \frac{1}{2n} \log \frac{\det((B+I)K_Z(B+I)^T + K_V)}{\det(K_Z)} \end{aligned}$$

and (B^{**}, K_V^{**}) achieves $C_{n,\rm FB}(K'_Z, P)$. Since $K'_Z \preceq K_Z$, we have

$$\operatorname{tr}\left(B^* K'_Z (B^*)^T + K_V^*\right) \le \operatorname{tr}\left(B^* K_Z (B^*)^T + K_V^*\right) \le nP$$

which shows that (B^*, K_V^*) is a feasible (not necessarily optimal) solution to $C_{n,\rm FB}(K'_Z, P)$. On the other hand, we have

$$(B^*+I)K'_Z(B^*+I)^T \preceq (B^*+I)K_Z(B^*+I)^T \tag{44}$$

so that

$$\begin{aligned} &C_{n,\rm FB}(K_Z, P)\\ &\quad = I(V^n; V^n + (B^*+I)Z^n)|_{V^n\sim N(0,K_V^*)}\\ &\quad \le I(V^n; V^n + (B^*+I)\tilde{Z}^n)|_{V^n\sim N(0,K_V^*)} &(45)\\ &\quad \le I(V^n; V^n + (B^{**}+I)\tilde{Z}^n)|_{V^n\sim N(0,K_V^{**})} &(46)\\ &\quad = C_{n,\rm FB}(K'_Z, P) \end{aligned}$$

where (45) follows from (44), divisibility of the Gaussian distribution, and the data processing inequality [1, Sec. 2.8]; and (46) follows because (B^{**}, K_V^{**}) achieves the feedback capacity $C_{n,\rm FB}(K'_Z, P)$ and (B^*, K_V^*) is a feasible solution to the maximization problem for $C_{n,\rm FB}(K'_Z, P)$. By letting n tend to infinity, we obtain

$$\lim_{n\to\infty} C_{n,\rm FB}(K_Z, P) \le \liminf_{n\to\infty} C_{n,\rm FB}(K'_Z, P).$$

For the other direction of inequality, we first consider the case $|\alpha| < 1$. Fix n and define the conditional covariance matrix $K_Z^{(m)}$, $m = 0, 1, \ldots$, of Z^n conditioned on m past values as

$$\begin{aligned} K_Z^{(0)} &:= K_Z\\ K_Z^{(m)} &:= \operatorname{Cov}(Z^n | Z_0, \ldots, Z_{-m+1}), \quad m = 1, 2, \ldots. \end{aligned}$$

It is easy to see that under this notation, the (elementwise) limit of covariance matrices $K_Z^{(m)}$ exists and

$$\lim_{m\to\infty} K_Z^{(m)} = K'_Z.$$

By sending a length-m training sequence over the channel with the noise covariance matrix K_Z, i.e., by transmitting $X_{-m+1} = \cdots = X_0 = 0$ and then estimating the noise process at the receiver using $Z_0, \ldots, Z_{-m+1}$, we can achieve the rate $nC_{n,\rm FB}(K_Z^{(m)})$ over $n+m$ transmissions. Hence, we have

$$nC_{n,\rm FB}(K_Z^{(m)}, P) \le (n+m)C_{n+m,\rm FB}(K_Z, P)$$

for all P. By carefully increasing both n and m, we will derive the desired inequality.

Consider using (B^{**}, K_V^{**}), which is optimal for the channel with noise covariance matrix K'_Z, for the channel with noise covariance $K_Z^{(m)}$. Since $K_Z^{(m)} \succeq K'_Z$, the resulting power usage can be greater than nP. However, we have

$$\begin{aligned} &\operatorname{tr}\left(K_V^{**} + B^{**} K_Z^{(m)} (B^{**})^T\right)\\ &\quad = \operatorname{tr}\Big(K_V^{**} + B^{**} K'_Z (B^{**})^T\\ &\qquad + B^{**}(K_Z^{(m)} - K'_Z)(B^{**})^T\Big)\\ &\quad \le nP + \operatorname{tr}\left(B^{**}(K_Z^{(m)} - K'_Z)(B^{**})^T\right). \end{aligned}$$

Now we observe that $K_Z^{(m)}$ and K'_Z differ only at the $(1,1)$ entry. Furthermore, the convergence of $K_Z^{(m)}(1,1) = \operatorname{Var}(Z_1|Z_0, \ldots, Z_{m-1})$ to $K'_Z(1,1) = \operatorname{Var}(Z_1|Z_0, Z_{-1}, \ldots)$ is exponentially fast in m (uniformly in n). Hence, we can bound the amount of additional power usage as

$$\begin{aligned} &\operatorname{tr}\left(B^{**}(K_Z^{(m)} - K'_Z)(B^{**})^T\right)\\ &\quad \le n^2 \max_{1\le i,j\le n} (B_{ij}^{**})^2 \max_{1\le i,j\le n}\left(K_Z^{(m)} - K'_Z(m)\right)\\ &\quad \le cn^3 e^{-\gamma m} =: n\epsilon_{n,m} \end{aligned}$$

where c and γ are constants independent of n and m. Combining above observations, we have the following chain of inequalities for all n and m:

$$\begin{aligned}
&(n+m)C_{n+m,\mathrm{FB}}(K_Z, P+\epsilon_{n,m}) \\
&\geq nC_{n,\mathrm{FB}}(K_Z^{(m)}, P+\epsilon_{n,m}) \\
&\geq \frac{1}{2}\log \frac{\det\left(K_V^{**}+(I+B^{**})K_Z^{(m)}(I+B^{**})^T\right)}{\det K_Z^{(m)}} \\
&\geq \frac{1}{2}\log \frac{\det\left(K_V^{**}+(I+B^{**})K'_Z(I+B^{**})^T\right)}{\det K_Z^{(m)}} \\
&= nC_{n,\mathrm{FB}}(K'_Z, P) + \frac{1}{2}\log\frac{\det K'_Z}{\det K_Z{}^{(m)}}.
\end{aligned} \tag{47}$$

Finally we let n and m grow to infinity such that

$$\frac{m}{n} \to 0 \quad \text{and} \quad n^2 e^{-m} \to 0.$$

The inequality (47) certainly implies that

$$\lim_{n\to\infty} C_{n,\mathrm{FB}}(K_Z, P+\epsilon) \geq \limsup_{n\to\infty} C_{n,\mathrm{FB}}(K'_Z, P)$$

for every $\epsilon > 0$. The desired inequality follows from the continuity[5] of the $C_{\mathrm{FB}}(K_Z, P)$ in P.

For the case $|\alpha| = 1$, we can perturb the noise process using a negligible amount of power and proceeds similarly as above. Indeed, we can perturb the original covariance matrices K'_Z and K_Z into the perturbed covariance matrices $K'_Z(\epsilon)$ and $K_Z(\epsilon)$ that correspond to the MA(1) process with parameter $\alpha(1-\epsilon)$, so that for appropriately chosen $\delta_k(\epsilon), k = 1,2,3$, with $\delta_k \to 0$ as $\epsilon \to 0$, we have

$$\begin{aligned}
\limsup_{n\to\infty} & C_{n,\mathrm{FB}}(K'_Z, P) \\
&\leq C_{\mathrm{FB}}(K'_Z(\epsilon), P+\delta_1(\epsilon)) && (48)\\
&= C_{\mathrm{FB}}(K_Z(\epsilon), P+\delta_1(\epsilon)) && (49)\\
&\leq C_{\mathrm{FB}}\left((1+\delta_2(\epsilon))^{-1}K_Z, P+\delta_3(\epsilon)\right) && (50)\\
&= C_{\mathrm{FB}}\left(K_Z, (1+\delta_2(\epsilon))(P+\delta_3(\epsilon))\right)
\end{aligned}$$

where (48) follows because we can transform the channel $K'_Z(\epsilon)$ into K'_Z using very small power, (49) follows from the result for $|\alpha| < 1$ we obtained above, and (50) follows since we can perturb the channel $(1+\delta_2(\epsilon))^{-1}K_Z$ into $K_Z(\epsilon)$ by adding some extra white noise. Since $C_{\mathrm{FB}}(K_Z, P)$ is continuous in P, we get

$$\limsup_{n\to\infty} C_{n,\mathrm{FB}}(K'_Z, P) \leq C_{\mathrm{FB}}(K_Z, P).$$

This completes the proof of the asymptotic equivalence of K_Z and K'_Z.

[5]The continuity of $C_{\mathrm{FB}}(K_Z,\cdot)$ follows from the concavity of $C_{\mathrm{FB}}(K_Z,\cdot)$, which, in turn, follows from the concavity of $C_{n,\mathrm{FB}}(K_Z,\cdot)$ [70, Th. 1]; recall that $C_{\mathrm{FB}}(K_Z,\cdot)$ is the pointwise limit of $C_{n,\mathrm{FB}}(K_Z,\cdot)$.

Appendix B
Proof of Proposition 1

Define

$$\phi(\xi,\zeta_1,\zeta_2) = \frac{1}{2}\log(1+(\sqrt{\zeta_1}+|\alpha|\sqrt{\zeta_2}e^{-\xi})^2)$$

for $\xi,\zeta_1,\zeta_2 \geq 0$, and

$$\psi(\xi,\zeta) = \phi(\xi,\zeta,\zeta), \quad \xi,\zeta \geq 0.$$

It is easy to check the following:

i) $\phi(\xi,\zeta_1,\zeta_2)$ is increasing and concave in (ζ_1,ζ_2);

ii) for each $\zeta_1,\zeta_2 \geq 0, \phi(\xi,\zeta_1,\zeta_2)$ is a decreasing contraction of ξ in the sense that

$$\phi(\xi_1,\zeta_1,\zeta_2) - \phi(\xi_2,\zeta_1,\zeta_2) \leq \xi_2 - \xi_1$$

for all ξ_1 and ξ_2; and consequently,

iii) for each $\zeta > 0$, there is a unique solution $\xi^*(\zeta)$ to the equation $\xi = \psi(\xi,\zeta)$ such that $\psi(\xi,\zeta) > \xi$ for all $\xi < \xi^*(\zeta)$ and $\psi(\xi) < \xi$ for all $\xi > \xi^*(\zeta)$.

For an arbitrary infinite sequence $\{P_i\}_{i=0}^\infty$ with $P_0 = 0$ and

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n P_i \leq P \tag{51}$$

we define

$$\begin{aligned}
\xi_0 &= 0\\
\xi_i &= \phi(\xi_{i-1}, P_i, P_{i-1}) \quad i = 1,2,\ldots.
\end{aligned}$$

Then we can rewrite the recursion formula (36)–(38) as

$$\begin{aligned}
\xi_1 &= \frac{1}{2}J_1(P_1)\\
\xi_i &= \frac{1}{2}\left(J_i(P_1,\ldots,P_i) - J_{i-1}(P_1,\ldots,P_{i-1})\right)
\end{aligned}$$

for $i = 2,3,\ldots$, so that

$$\frac{1}{2n}J_n(P_1,\ldots,P_n) = \frac{1}{n}\sum_{i=1}^n \xi_i.$$

Now we show that

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

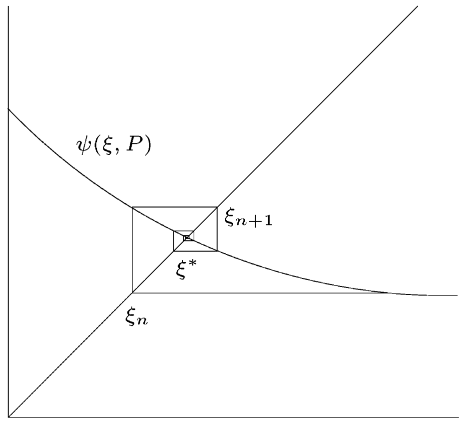

Fig. 4. Convergence to the unique point ξ^*.

where $\xi^* = \xi^*(P)$ is the unique solution to the equation $\xi = \psi(\xi, P)$. Indeed,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\xi_i &= \frac{1}{n}\sum_{i=1}^{n}\phi(\xi_{i-1}, P_i, P_{i-1})\\ &= \frac{1}{n}\sum_{i=1}^{n}[\phi(\xi_{i-1}, P_i, P_{i-1}) - \phi(\xi^{**}, P_i, P_{i-1})\\ &\qquad + \phi(\xi^{**}, P_i, P_{i-1})]\\ &\le \frac{1}{n}\sum_{i=1}^{n}(\xi^{**} - \xi_{i-1} + \phi(\xi^{**}, P_i, P_{i-1}))\\ &\le \frac{1}{n}\sum_{i=1}^{n}(\xi^{**} - \xi_{i-1})\\ &\qquad + \phi\left(\xi^{**}, \frac{1}{n}\sum_{i=1}^{n}P_i, \frac{1}{n}\sum_{i=1}^{n}P_{i-1}\right)\end{aligned}$$

where the first inequality follows from the aforementioned property ii) and the second inequality follows from the property i) and Jensen's inequality. By taking limits on both sides, we get from continuity of $\phi(\xi, \zeta_1, \zeta_2)$ in (ζ_1, ζ_2)

$$\xi^{**} \le \phi(\xi^{**}, P, P) = \psi(\xi^{**}, P)$$

which, from the property iii), implies that $\xi^{**} \le \xi^*$. We can also check that letting $P_i \equiv P$ for all $i = 1, 2, \ldots$ attains $\xi^{**} = \xi^*$ from the property ii) and the principle of contraction mappings [71, Sec. 14]. (See Fig. 4 above and the detailed analysis in [15, Sec. 5].) Thus, we conclude that the supremum of $\limsup_{n\to\infty}(2n)^{-1}J_n(P_1, \ldots, P_n)$ over all infinite power sequences $\{P_i\}$ satisfying the power constraint (51) is achieved by the uniform power allocation. From simple change of variable $x_0 = \exp(-\xi^*)$, we have $\xi^* = -\log x_0$ where $0 < x_0 \le 1$ is the unique positive solution to (39).

As in the MA(1) case before, it remains to justify the interchange of the order of limit and supremum in

$$\begin{aligned}\lim_{n\to\infty} R_n &= \lim_{n\to\infty} \max_{P_i: \sum_i P_i \le nP} \frac{1}{2n}J_n(P_1, \ldots, P_n)\\ &= \sup_{\{P_i\}_{i=1}^{\infty}} \limsup_{n\to\infty} \frac{1}{2n}J_n(P_1, \ldots, P_n)\\ &= \lim_{n\to\infty}\frac{1}{2n}J_n(P, \ldots, P)\\ &= \xi^*.\end{aligned}$$

Obviously we have

$$\begin{aligned}\lim_{n\to\infty} \max_{P_i: \sum_i P_i \le nP} &\frac{1}{2n}J_n(P_1, \ldots, P_n)\\ &\ge \lim_{n\to\infty}\frac{1}{2n}J_n(P, \ldots, P).\end{aligned}$$

For the other direction of inequality, first fix n and then take $(P_1^*, \ldots, P_{n-1}^*)$ that achieves R_{n-1}. We construct the infinite sequence $\{P_i\}_{i=1}^{\infty}$ by concatenating $(P_1^*, \ldots, P_{n-1}^*, 0)$ repeatedly, that is, $P_{kn+i} = P_i^*,\ 1 \le i \le n-1,\ k = 0, 1, \ldots$, and $P_{kn} = 0$ for all $k = 1, \ldots$. Now we can easily verify that

$$\begin{aligned}J_{kn}(P_1, \ldots, P_{kn}) &= kJ_{n-1}(P_1^*, \ldots, P_{n-1}^*)\\ &= 2k(n-1)R_{n-1}.\end{aligned}$$

(Taking $P_{kn} = 0$ resets the dependence on the past.) By taking limits on both sides, we get

$$\begin{aligned}\lim_{n\to\infty} R_{n-1} &= \lim_{n\to\infty}\frac{n}{n-1}\frac{1}{2kn}J_{kn}(P_1, \ldots, P_{kn})\\ &\le \limsup_{n\to\infty}\frac{1}{2n}J_n(P_1, \ldots, P_n)\\ &\le \lim_{n\to\infty}\frac{1}{2n}J_n(P, \ldots, P).\end{aligned}$$

This completes the proof of Proposition 1.

ACKNOWLEDGMENT

The author is pleased to express his gratitude to Tom Cover for his invaluable insights and guidance throughout this work, to Styrmir Sigurjónsson and Erik Ordentlich for enlightening discussions, and to Sina Zahedi for his numerical optimization program, which was especially helpful in the initial phase of this study. He also thanks one of the anonymous reviewers for his/her careful reading of the paper, which resulted in many improvements of the manuscript.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[2] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. IT-35, no. 1, pp. 37–43, Jan. 1989.

[3] G. Pólya and G. Szegő, *Problems and Theorems in Analysis*. New York: Springer-Verlag, 1998.
[4] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 172–182, Apr. 1966.
[5] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback–II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 183–189, Apr. 1966.
[6] J. Wolfowitz, "Note on the Gaussian channel with feedback and a power constraint," *Inf. Contr.*, vol. 12, pp. 71–78, 1968.
[7] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. IT-2, no. 3, pp. 8–19, Sep. 1956.
[8] T. T. Kadota, M. Zakai, and J. Ziv, "Capacity of a continuous memoryless channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-17, pp. 372–378, 1971.
[9] ——, "Mutual information of the white Gaussian channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. IT-17, pp. 368–371, 1971.
[10] M. S. Pinsker, "The probability of error in block transmission in a memoryless Gaussian channel with feedback," *Probl. Inf. Transm.*, vol. 4, no. 4, pp. 3–19, 1968.
[11] A. J. Kramer, "Improving communication reliability by use of an intermittent feedback channel," *IEEE Trans. Inf. Theory*, vol. IT-15, pp. 52–60, Jan. 1969.
[12] K. S. Zigangirov, "Upper bounds for the probability of error for channels with feedback," in *Probl. Inf. Transm. (Dubna, 1969)*, 1970, vol. 6, no. 2, pp. 87–92.
[13] J. P. M. Schalkwijk, "Center-of-gravity information feedback," *IEEE Trans. Inf. Theory*, vol. IT-14, pp. 324–331, 1968.
[14] S. Butman, "A general formulation of linear feedback communication systems with solutions," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 3, pp. 392–400, May 1969.
[15] J. Wolfowitz, "Signalling over a Gaussian channel with feedback and autoregressive note," *J. Appl. Probab.*, vol. 12, no. 4, pp. 713–723, 1975.
[16] J. C. Tiernan, "Analysis of the optimum linear system for the autoregressive forward channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. IT-22, pp. 359–363, May 1976.
[17] 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. Inf. Theory*, vol. IT-20, pp. 311–316, 1974.
[18] S. Butman, "Linear feedback rate bounds for regressive channels," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 3, pp. 363–366, May 1976.
[19] L. H. Ozarow, "Random coding for additive Gaussian channels with feedback," *IEEE Trans. Inf. Theory*, vol. IT-36, no. 1, pp. 17–22, 1990.
[20] ——, "Upper bounds on the capacity of Gaussian channels with feedback," *IEEE Trans. Inf. Theory*, vol. IT-36, no. 1, pp. 156–161, 1990.
[21] E. Ordentlich, "A class of optimal coding schemes for moving average additive Gaussian noise channels with feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, Trondheim, Norway, Jun./Jul. 1994, p. 467.
[22] S. Yang, A. Kavčic, and S. Tatikonda, "Linear Gaussian channels: Feedback capacity under power constraints," in *Proc. IEEE Int. Symp. Inf. Theory*, Chicago, IL, Jun./Jul. 2004, p. 72.
[23] S. Yang, "The Capacity of Communication Channels With Memory," Ph.D. dissertation, Harvard University, Cambridge, MA, Jun. 2004.
[24] P. Elias, "Channel Capacity Without Coding," MIT Res. Lab. Electronics, Cambridge, MA, Oct. 1956, Quarterly Progr. Rep.
[25] P. E. Green, Jr., "Feedback communication systems," in *Lectures on Communication System Theory* (with an appendix by P. Elias, "Channel capacity without coding," pp. 363–368), E. Baghdady, Ed. New York: McGraw-Hill, 1961, pp. 345–368.
[26] P. Elias, "Networks of Gaussian channels with applications to feedback systems," *IEEE Trans. Inf. Theory*, vol. IT-13, pp. 493–501, 1967.
[27] G. L. Turin, "Signal design for sequential detection systems with feedback," *IEEE Trans. Inf. Theory*, vol. IT-11, pp. 401–408, 1965.
[28] ——, "Comparison of sequential and nonsequential detection systems with uncertainty feedback," *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 5–8, 1966.
[29] ——, "More on uncertainty feedback: The bandlimited case," *IEEE Trans. Inf. Theory*, vol. IT-14, pp. 321–324, 1968.
[30] M. Horstein, "On the design of signals for sequential and nonsequential detection systems with feedback," *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 448–455, 1966.
[31] R. Z. Khas'minskii, "Sequential signal transmission in a Gaussian channel with feedback," *Probl. Inf. Transm.*, vol. 3, no. 2, pp. 37–44, 1967.
[32] M. J. Ferguson, "Optimal signal design for sequential signaling over a channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-14, pp. 331–340, 1968.
[33] J. K. Omura, "Optimum linear transmission of analog data for channels with feedback," *IEEE Trans. Inf. Theory*, vol. IT-14, pp. 38–43, 1968.
[34] A. D. Wyner, "On the Schalkwijk–Kailath coding scheme with a peak energy constraint," *IEEE Trans. Inf. Theory*, vol. IT-14, pp. 129–134, Jan. 1968.
[35] J. P. M. Schalkwijk and M. E. Barron, "Sequential signalling under a peak power constraint," *IEEE Trans. Inf. Theory*, vol. IT-17, pp. 278–282, May 1971.
[36] R. L. Kashyap, "Feedback coding schemes for an additive noise channel with a noisy feedback link," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 3, pp. 1355–1387, 1968.
[37] S. S. Lavenberg, "Feedback communication using orthogonal signals," *IEEE Trans. Inf. Theory*, vol. IT-15, pp. 478–483, 1969.
[38] ——, "Repetitive signaling using a noisy feedback channel," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 3, pp. 269–278, May 1971.
[39] T. Kailath, "An application of Shannon's rate-distortion theory to analog communication over feedback channels," in *Proc. Princeton Symp. Syst. Sci.*, Princeton, NJ, Mar. 1967.
[40] T. J. Cruise, "Achievement of rate-distortion bound over additive white noise channel utilizing a noiseless feedback channel," *Proc. IEEE*, vol. 55, no. 4, pp. 583–584, 1967.
[41] J. P. M. Schalkwijk and L. I. Bluestein, "Transmission of analog waveforms through channels with feedback," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 4, pp. 617–619, 1967.
[42] I. A. Ovseevič, "Optimal transmission of Gaussian data over a channel with white Gaussian noise in the presence of an inverse connection," *Probl. Inf. Transm.*, vol. 6, no. 3, pp. 3–14, 1970.
[43] S. Ihara, "Optimal coding in white Gaussian channel with feedback," in *Proceedings of the Second Japan-USSR Symposium on Probability Theory, ser. Lecture Notes in Math*. Berlin, Germany: Springer-Verlag, 1973, vol. 330, pp. 120–123.
[44] L. H. Ozarow and S. K. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 667–671, 1984.
[45] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 379–381, 1978.
[46] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 623–629, 1984.
[47] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1423–1438, 2002.
[48] A. Dembo, "On Gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. IT-35, no. 5, pp. 1072–1076, Sep. 1989.
[49] M. S. Pinsker, *Talk Delivered at the Soviet Inf. Theory Meeting*. 1969, no abstract published.
[50] P. M. Ebert, "The capacity of the Gaussian channel with feedback," *Bell Syst. Tech. J.*, vol. 49, pp. 1705–1712, 1970.
[51] K. Yanagi, "An upper bound to the capacity of discrete time Gaussian channel with feedback–II," *IEEE Trans. Inf. Theory*, vol. IT-40, pp. 588–593, Mar. 1994.
[52] H. W. Chen and K. Yanagi, "Refinements of the half-bit and factor-of-two bounds for capacity in Gaussian channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-45, no. 1, pp. 319–325, Jan. 1999.
[53] H. W. Chen and K. Yanagi, "Upper bounds on the capacity of discrete-time blockwise white Gaussian channels with feedback," *IEEE Trans. Inf. Theory*, vol. IT-46, no. 3, pp. 1125–1131, May 2000.
[54] T. M. Cover, "Conjecture: Feedback doesn't help much," in *Open Problems in Communication and Computation*, T. M. Cover and B. Gopinath, Eds. New York: Springer-Verlag, 1987, pp. 70–71.
[55] J. A. Thomas, "Feedback can at most double Gaussian multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 5, pp. 711–716, Sep. 1987.
[56] S. Pombra and T. M. Cover, "Non white Gaussian multiple access channels with feedback," *IEEE Trans. Inf. Theory*, vol. IT-40, no. 3, pp. 885–892, May 1994.
[57] E. Ordentlich, "On the factor-of-two bound for Gaussian multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-42, no. 6, pp. 2231–2235, Nov. 1996.
[58] S. Ihara, "On the capacity of the continuous time Gaussian channel with feedback," *J. Multivariate Anal.*, vol. 10, no. 3, pp. 319–331, 1980.
[59] ——, "Capacity of mismatched Gaussian channels with and without feedback," *Probab. Theory Rel. Fields*, vol. 84, pp. 453–471, 1990.
[60] ——, "Coding theorems for a continuous-time Gaussian channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-40, pp. 2041–2045, 1994.

[61] ——, "Capacity of discrete time Gaussian channel with and without feedback–I," *Mem. Fac. Sci. Kochi Univ. Ser. A Math.*, vol. 9, pp. 21–36, 1988.
[62] ——*, Information Theory for Continuous Systems*. River Edge, NJ: World Scientific, 1993.
[63] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory*, arXiv e-print: cs.IT/060291, submitted for publication.
[64] ——, "Gaussian Feedback Capacity," Ph.D. dissertation, Stanford University, Stanford, CA, Jun. 2006.
[65] W. Rudin*, Real and Complex Analysis*. New York: McGraw-Hill, 1987.
[66] S. Boyd and L. Vandenberghe*, Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
[67] S. Ihara, "On the capacity of the discrete time Gaussian channel with feedback," in *Transactions of the Eighth Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes*. Dordrecht, Germany: Reidel, 1979, vol. C, pp. 175–186.
[68] T. Kailath, A. H. Sayed, and B. Hassibi*, Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
[69] D. G. Luenberger*, Optimization by Vector Space Methods*. New York: Wiley, 1969.
[70] K. Yanagi, H. W. Chen, and J. W. Yu, "Operator inequality and its application to capacity of Gaussian channel," *Taiwanese J. Math.*, vol. 4, no. 3, pp. 407–416, Sep. 2000.
[71] A. N. Kolmogorov and S. V. Fomin*, Elements of the Theory of Functions and Functional Analysis*. Rochester, NY: Graylock Press, 1957.