Achieving the Stationary Feedback Capacity for Gaussian Channels

Jialing Liu and Nicola Elia [1]
Dept. Electrical and Computer Engineering, Iowa State University
{liujl,nelia}@iastate.edu

Abstract- In this paper, we study a Gaussian channel with memory and with noiseless feedback, for which we present a coding scheme to achieve the *stationary feedback capacity* (the maximum information rate over all *stationary* input distributions, conjectured to be the asymptotic feedback capacity). The coding scheme essentially implements the celebrated Kalman filter algorithm; is equivalent to an estimation system over the same channel without feedback; and reveals that the achievable information rate of the feedback communication system can be alternatively given by the decay rate of the Cramer-Rao bound of the associated estimation system. Thus, combined with the control theoretic characterizations of feedback communication (proposed by Elia), this implies that the fundamental limitations in feedback communication, estimation, and control coincide. In addition, the proposed coding scheme simplifies the coding complexity and shortens the coding delay, and its construction amounts to solving a finite-dimensional optimization problem. We also provide a further simplification to the optimal input distribution developed by Yang, Kavcic, and Tatikonda.

I. Introduction

Communication systems in which the transmitter have access to noiseless feedback of channel outputs have been widely studied; see [1]–[13] and references therein for the study of Gaussian channels with feedback. [1], [2] proposed ingenious feedback codes for additive white Gaussian noise (AWGN) channels, which achieve the asymptotic feedback capacity (denoted C_∞) and greatly reduce the coding complexity and delay. [5] provided a rather general coding structure to achieve the finite-horizon feedback capacity (denoted C_T) for channels with memory; however, it involves prohibitive computation complexity as the coding length $(T+1)$ increases.

[9] proved that the maximum directed information is the feedback capacity; reformulated the problem of finding C_T as a stochastic control optimization problem; and proposed a dynamic programming based solution. This idea was further explored in [10], which uncovered the Markov property of the optimal input distribution for Gaussian channels with memory and eventually reduced the finite-horizon stochastic control optimization problem to a manageable size. Moreover, under a *stationarity conjecture* that C_∞ equals the stationary feedback capacity (denoted C_s, the maximum rate over all *stationary* input distributions), C_∞ is given by the solution of a finite dimensional optimization problem. This is the first computationally efficient [2] method to calculate C_∞ or C_T for general Gaussian channels. [13] studied first-order moving-average Gaussian channels with feedback and discovered the closed-form expression for C_∞.

[11] investigated the tracking of unstable sources over a channel and proposed the notion of *anytime capacity* to capture the fundamental limitations in that problem, which reveals connections between communication and control and brings new insights to feedback communication problems. Furthermore, [12] established the *equivalence* between feedback communication and feedback stabilization over Gaussian channels with memory, showed that the achievable transmission rate is given by the Bode sensitivity integral, and presented an optimization problem based on robust control to compute lower bounds of C_T. [12] also extended the codes in [1], [2] to achieve these lower bounds.

As we can see, it remains an open problem to build a coding scheme with reasonable complexity to achieve C_∞ or C_s for a Gaussian channel with memory; note that no practical codes have been found based on the optimal signalling strategy in [10]. In this paper, we propose a coding scheme for Gaussian channels with noiseless feedback. This coding scheme achieves C_s, the *stationary* feedback capacity of the channel; utilizes the Kalman filter algorithm; simplifies the coding processes; and shortens coding delay. The optimal coding structure is essentially a finite-dimensional linear time-invariant (FDLTI) system, and leads to a further

[1]This research was supported by NSF under Grant ECS-0093950. The authors would like to thank Anant Sahai, Sekhar Tatikonda, Sanjoy Mitter, Murti Salapaka, Zhengdao Wang, Shaohua Yang, and Young-Han Kim for useful discussion.

[2]Here we do not mean their optimization problem is convex. In fact the computation complexity for C_T is $O(T)$, and for C_∞ the complexity is determined mainly by the channel order.

simplification of the optimal stationary signalling strategy in [10]. The construction of the coding system amounts to solving a finite-dimensional optimization problem. Our solution holds for AWGN channels with intersymbol interference (ISI) where the ISI is modeled as a stable and minimum-phase FDLTI system. [3]

The problem of achieving C_∞ remains open, because a proof confirming the stationarity conjecture is missing. However, our study of achieving C_s, the main focus of this paper, is justified by its great simplifications in the coding systems design and operation and by the numerical evidence that C_s indeed equals C_∞.

We remark that our optimal coding design may be derived by applying the control-oriented approach in [12] to the results in [10]. To highlight other important aspects of the coding design, however, we follow a less direct route to derive the scheme, that is, we first present finite-horizon analysis of the feedback communication problem, and then let the horizon tend to infinity.

In our finite-horizon analysis, we establish the necessity of the Kalman filter: The Kalman filter is not only a device to provide sufficient statistics (which was shown in [10]), but also a device to ensure the power efficiency and to recover the message optimally. Additionally, the presence of Kalman filter in our coding scheme reveals the intrinsic connections between feedback communication, estimation, and control. In particular, we show that the feedback communication problem over a Gaussian channel is essentially an optimal estimation problem, and the achievable rate of the feedback communication system is alternatively given by the decay rate of the Cramer-Rao bound (CRB) for the associated estimation system. Invoking the Bode sensitivity characterization of achievable rate [12], we conclude that the fundamental limitations in feedback communication, estimation, and control coincide. We then extend the horizon to infinity and characterize the steady-state of the feedback communication problem. We finally show that our optimal scheme achieves C_s.

We denote by y^T the vector $\{y_0, y_1, \cdots, y_T\}$, and $\{y_t\}$ the sequence $\{y_t\}_{t=0}^{\infty}$. For a random vector y^T, we denote its covariance matrix as $K_y^{(T)}$. We denote "defined to be" as ":=".

II. Channel model

Fig. 1 shows a single-input single-output AWGN channel with ISI, denoted as $\mathcal{F}$. It is described in state-space as

$$\mathcal{F}: \begin{cases} s_{t+1} &= F s_t + G u_t, \quad s_0 = 0 \\ y_t &= H s_t + u_t + N_t, \end{cases} \tag{1}$$

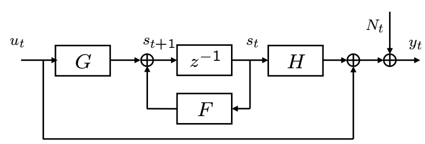


Fig. 1. State-space description of $\mathcal{F}$.

where $F \in \mathbb{R}^{m\times m}$, m is the *dimension* or *order* of $\mathcal{F}$, u_t is the channel input, s_t is the channel state, N_t is AWGN and $N_t \sim \mathcal{N}(0,1)$, and y_t is the channel output. The transfer function from u to y, denoted $\mathcal{Z}(z)^{-1}$, is stable and minimum-phase. [4] In matrix form, we have

$$\mathcal{F}: \; y^T = \mathcal{Z}_T^{-1} u^T + N^T \tag{2}$$

for any block size $(T+1)$, where $\mathcal{Z}_T^{-1} \in \mathbb{R}^{(T+1)\times(T+1)}$ is a lower-triangular Toeplitz matrix of impulse response of $\mathcal{Z}(z)^{-1}$, and has diagonal elements all equal to 1 (and thus is invertible). We may abuse the notation $\mathcal{Z}^{-1}$ for both $\mathcal{Z}(z)^{-1}$ and $\mathcal{Z}_T^{-1}$. We focus on the case $m \geq 1$; the case $m = 0$ was solved in [1], [2].

[3]Through the equivalence shown in [9], [10], this is equivalent to colored Gaussian channels with rational noise power spectrums and without ISI; the rationalness assumption is not too restrictive since any power spectrum can be arbitrarily approximated by rational ones.

[4]We use $\mathcal{Z}^{-1}$ here to reserve $\mathcal{Z}$ for the filter generating colored noise in a colored Gaussian channel, for future use purpose.

III. PROBLEM FORMULATION IN STEADY-STATE AND THE SOLUTION

Before formulating the steady-state communication problem, we distinguish among the three scenarios: Finite-horizon (i.e. finite coding length), infinite-horizon (i.e. infinite coding length), and steady state. Finite-horizon problems often have time-dependent (i.e. time-varying), horizon-dependent solutions, see e.g. finite-horizon Kalman filtering. The horizon-dependence may be removed in the infinite-horizon scenario, and furthermore, the time-dependence may be removed in the steady-state scenario. Therefore, we focus on *finding a (stationary, time-invariant) steady-state solution*, and truncate it and employ the truncation if the practical problem is in finite-horizon but the horizon is large enough. This truncated solution would greatly simplify the implementation while having a performance sufficiently close to finite-horizon optimality.

A. Problem formulation

For a Gaussian channel with feedback, the channel input has the form

$$u_t = \gamma_t u^{t-1} + \eta_t y^{t-1} + \xi_t \tag{3}$$

for any $\gamma_t \in \mathbb{R}^{1\times t}$, $\eta_t \in \mathbb{R}^{1\times t}$, and zero-mean Gaussian random variable $\xi_t \in \mathbb{R}$ which is independent of u^{t-1} and y^{t-1} [9], [10]. The channel inputs are allowed to depend on the channel outputs in a strictly causal manner. Our objective in this paper is to *design encoder/decoder to achieve the stationary feedback capacity*, given by

$$C_s := C_s(\mathcal{P}) := \sup \lim_{T\to\infty} \frac{1}{T+1} I(u^T \to y^T), \text{ subject to } P_\infty := \lim_{T\to\infty} \frac{1}{T+1}\mathbf{E}u^{T\prime}u^T \leq \mathcal{P} \tag{4}$$

for any stationary $\{u_t\}$ in the form of (3). Here $\mathcal{P} > 0$ is the power budget and $I(u^T \to y^T)$ is the directed information from u^T to y^T [9]. Note that C_s is well defined [10].

The problem of solving C_s may be equivalently formulated as minimizing the average channel input power while keeping the rate bounded from below, namely for $\mathcal{R} > 0$,

$$P_{\min}(\mathcal{R}) := \inf \lim_{T\to\infty} \frac{1}{T+1}\mathbf{E}u^{T\prime}u^T, \text{ subject to } \lim_{T\to\infty} \frac{1}{T+1} I(u^T \to y^T) \geq \mathcal{R} \tag{5}$$

for any stationary $\{u_t\}$ in the form of (3). Therefore $P_{\min}(\mathcal{R})$ is the inverse function of $C_s(\mathcal{P})$, i.e., $C_s(P_{\min}(\mathcal{R})) = \mathcal{R}$.

It is conjectured that a stationary sequence $\{u_t\}$ achieves C_∞ ($C_\infty := \lim_{n\to\infty} C_T$ exists by superadditivity of C_T [13]). However, a rigorous proof is not available. Our study of the steady-state problem avoids that technical difficulty and leads to a horizon-independent, time-invariant solution, greatly reducing the implementation complexity.

B. The coding scheme

The encoder/decoder structure: In state-space, the encoder and decoder are described as

$$\text{Encoder:}\left\{\begin{array}{rcl} x_{t+1} &=& Ax_t,\ x_0 := W \\ r_t &=& Cx_t \\ u_t &=& r_t - \hat{r}_t \end{array}\right. \qquad \text{Decoder:}\left\{\begin{array}{rcl} \hat{s}_{t+1} &=& F\hat{s}_t + L_2 e_t,\ \hat{s}_0 = 0 \\ e_t &=& y_t - H\hat{s}_t \\ \hat{x}_{t+1} &=& A\hat{x}_t + L_1 e_t,\ \hat{x}_0 = 0 \\ \hat{r}_t &=& C\hat{x}_t \\ \hat{W}_t &=& A^{-t}\hat{x}_t, \end{array}\right. \tag{6}$$

where $A \in \mathbb{R}^{(n+1)\times(n+1)}$, $C \in \mathbb{R}^{1\times(n+1)}$, $L_1 \in \mathbb{R}^{n+1}$, $L_2 \in \mathbb{R}^m$, and $W \sim \mathcal{N}(0, I_{n+1})$. We call $(n+1)$ the *encoder dimension*. See Fig. 2 for the block diagram. Note that $-\hat{r}_t$ is the feedback from the decoder based on the channel output y^{t-1}, and $-\hat{r}^t = \mathcal{G}_t^* y^t$ where $\mathcal{G}_t^*$ is a strictly lower triangular Toeplitz matrix. Here A, C, u_t, etc. depends on n, however, we do not specify the dependence explicitly to simplify notations.

Optimal choice of parameters: Fix a desired rate $\mathcal{R}$. Let $DI := 2^{\mathcal{R}}$ and $n := m-1$, and solve the optimization problem

$$[\boldsymbol{a}_f^{opt}, \Sigma^{opt}] := \underset{\substack{\boldsymbol{a}_f \in \mathbb{R}^n \\ s.t.\ \Sigma = \mathbb{A}\Sigma\mathbb{A}' - \mathbb{A}\Sigma\mathbb{C}'\mathbb{C}\Sigma\mathbb{A}'/(\mathbb{C}\Sigma\mathbb{C}'+1)}}{\arg\inf} \mathbb{D}\Sigma\mathbb{D}', \tag{7}$$

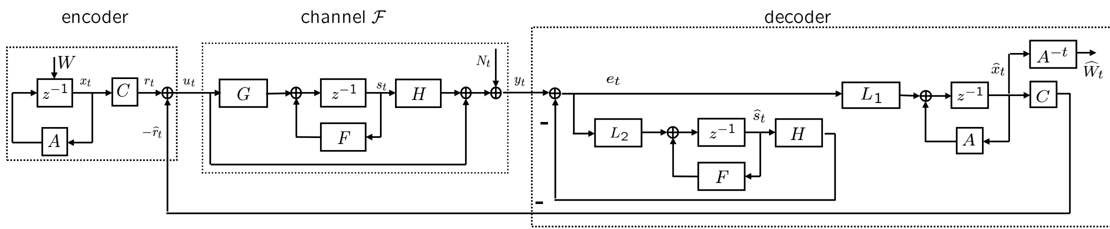


Fig. 2. The encoder/decoder structure for $\mathcal{F}$.

where

$$\mathbb{A} := \begin{bmatrix} A & 0 \\ GC & F \end{bmatrix}, \mathbb{C} := [\, C \quad H \,], \mathbb{D} := [\, C \quad 0 \,], A := \left[\begin{array}{c|c} 0_{n\times 1} & I_n \\ \hline \pm DI & \boldsymbol{a}_f \end{array} \right], C := \left[\, 1 \quad 0_{1\times n} \,\right], \quad (8)$$

Note that we need to solve the problem twice (one for $+DI$ in A and one for $-DI$ in A), and choose the optimal solution as the one with the smaller objective function value. Then we obtain the optimal A^{opt} according to $\boldsymbol{a}_f^{opt}$, and let (n^*+1) be the number of unstable eigenvalues in A^{opt}, where $n^* \geq 0$. Let $n := n^*$, solve the above optimization problem again, let A^* be the newly obtained A^{opt}, and form $\mathbb{A}^*$. Using Σ^* (the newly obtained Σ^{opt}) and $\mathbb{C}^* := [1, 0_{1\times n^*}, H]$, we obtain

$$L^* := [L_1^{*\prime}, L_2^{*\prime}]' := \mathbb{A}^*\Sigma^*\mathbb{C}^{*\prime}/(\mathbb{C}^*\Sigma^*\mathbb{C}^{*\prime} + 1). \quad (9)$$

It holds that (A^*, C^*) is observable, and A^* has exactly (n^*+1) unstable eigenvalues.

We assign the encoder/decoder parameters by letting

$$A := A^*, C := C^* := [1, 0_{1\times n^*}], L_1 := L_1^*, L_2 := L_2^*. \quad (10)$$

We then drive the initial condition s_0 of channel $\mathcal{F}$ to zero. Now we are ready to communicate at a rate $\mathcal{R}$ using power $P_{\min}(\mathcal{R}) = \mathbb{D}^*\Sigma^*\mathbb{D}^{*\prime}$ where $\mathbb{D}^* := [C^*, 0]$.

Encoding/Decoding process: The designed communication system can transmit either an analog source or a digital message. In the former case, we assume that the encoder wishes to convey a Gaussian random vector through the channel and the decoder wishes to learn the random vector, which is a rate-distortion problem. The coding process is as follows. Assume that the to-be-conveyed message W is distributed as $\mathcal{N}(0, I_{n^*+1})$ (noting that any non-degenerate (n^*+1)-variate Gaussian vector W can be transformed in this form). Assume that the coding length is $(T+1)$. To encode, let $x_0 := W$. Then run the system till time epoch T. To decode, let $\hat{W}_t := \hat{x}_{0,t}$ for $t = 0, 1, \cdots, T$. The quantities of interest include the square-error distortion $\text{MSE}(\hat{W}_t) := \mathbf{E}(W - \hat{W}_t)(W - \hat{W}_t)'$. In the case of transmitting a digital message, the encoding/decoding can be done in a partitioned hypercube, see e.g. [1], [12].

C. Coding theorem

Theorem 1. *Construct the encoder/decoder shown in Fig. 2 using n^*, A^*, C^*, L_1^*, and L_2^*. Then under the power constraint $\mathbf{E}u^2 \leq \mathcal{P}$,*

i) The coding scheme transmits an analog source $W \sim \mathcal{N}(0, I_{n^*+1})$ from the encoder to the decoder at rate $C_s(\mathcal{P})$, with MSE distortion $D(C_s(\mathcal{P}))$, where $D(\cdot)$ is the distortion-rate function;

ii) The coding scheme can transmit digital message from the encoder to the decoder at a rate arbitrarily close to $C_s(\mathcal{P})$, with PE_T decays to zero doubly exponentially.

The proof of the theorem will be developed in the next four sections. In Section IV, we consider a general coding structure in finite-horizon which may be viewed as a generalization of our optimal coding structure. We show that this general structure essentially contains a Kalman filter. The presence of the Kalman filter links the feedback communication problem to an estimation problem and a control problem, and hence we rewrite the information rate in terms of estimation theory quantities and control theory quantities; see Section V. Sections IV and V are focused on finite-horizon. In Section VI, we extend the horizon to infinity and characterize the steady-state behaviors. Then in Section VII, we show that our optimal encoder/decoder design is actually the solution to the steady-state communication problem.

IV. Necessity of Kalman filter in optimal coding

In this section, we consider a finite-horizon coding structure that includes our optimal design in Section III as a special case. This general structure is useful since: 1) searching over all possible parameters in the general structure achieves C_s; 2) we can show that to ensure power efficiency (to be explained), the general structure necessarily contains a Kalman filter.

A. A general coding structure

Fig. 3 illustrates the general coding structure, including the encoder and the *feedback generator*, a portion of the decoder. Below, we fix the time horizon to be $\{0, 1, \cdots, T\}$ and describe the coding structure.

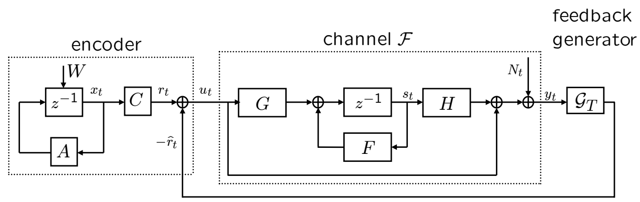


Fig. 3. A general coding structure for $\mathcal{F}$.

Encoder: Let the encoder dimension $(n+1)$ satisfy $0 \le n \le T$. We assume that $W \sim \mathcal{N}(0, I_{n+1})$, $A \in \mathbb{R}^{(n+1)\times(n+1)}$, $C \in \mathbb{R}^{1\times(n+1)}$, (A, C) is observable, and none of the eigenvalues of A are in the unit circle or at the locations of the eigenvalues of F. We denote the observability matrix for (A, C) as $\Gamma_n := [C', A'C', \cdots, A^{n\prime}C']'$, and let $\Gamma := [C', A'C', \cdots, A^{T\prime}C']'$ and $K_r^{(T)} := \mathbf{E} r^T r^{T\prime}$. Therefore, Γ_n is invertible, Γ has rank $(n+1)$, $r^T = \Gamma W$, and $K_r^{(T)} = \Gamma\Gamma'$.

Feedback generator: The feedback signal $-\hat{r}_t$ is generated through $\mathcal{G}_T$, the *feedback generator*, i.e. $-\hat{r}^T = \mathcal{G}_T y^T$. We assume that $\mathcal{G}_T \in \mathbb{R}^{(T+1)\times(T+1)}$ is a strictly lower triangular matrix. Clearly, the optimal encoder/decoder can be viewed as a special case of the general structure. Throughout the paper, the above assumptions are always assumed.

Definition 1. *Consider the system shown in Fig. 3. Define*

$$C_{T,n} := C_{T,n}(\mathcal{P}) := \sup_{\substack{A\in\mathbb{R}^{(n+1)\times(n+1)}, C, \mathcal{G}_T \\ s.t.\ \mathbf{E} u^{T\prime} u^T/(T+1) \le \mathcal{P}}} \frac{1}{T+1} I(W; y^T) \tag{11}$$

and define its inverse function as $P_{T,n}(\mathcal{R})$.

In other words, $C_{T,n}$ is the finite-horizon capacity (in the information sense) for a fixed transmitter dimension. It is not hard to show that $C_{n,n} = C_n$ and hence $\lim_{n\to\infty} C_{n,n} = C$. Moreover, $\lim_{T\to\infty} C_{T,n} := C_{\infty,n}$ is well defined and $\lim_{n\to\infty} C_{\infty,n} = C_s$ (details skipped for brevity).

B. The presence of Kalman filter

We first compute the mutual information in the general coding structure.

Proposition 1. *Consider the general coding structure in Fig. 3. Fix any* $0 \le n \le T$. *For any fixed* (A, C) *and* $\mathcal{G}_T$, *it holds that*

$$I(W; y^T) = I(u^T \to y^T) = \frac{1}{2} \log |I + \mathcal{Z}_T^{-1} K_r^{(T)} \mathcal{Z}_T^{-1\prime}|. \tag{12}$$

Proof: Since (A, C) is observable, W and r^T determine each other, so $I(W; y^T) = I(r^T; y^T)$. Note that $y^T = (I - \mathcal{Z}_T^{-1}\mathcal{G}_T)^{-1}(\mathcal{Z}_T^{-1} r^T + N^T)$ and $|I - \mathcal{Z}_T^{-1}\mathcal{G}_T| = 1$. Then the result follows from direct computation. ■

Proposition 1 says that $I(W; y^T)$ is independent of $\mathcal{G}_T$, and dependent only on $K_r^{(T)}$ or equivalently on (A, C). Thus, fixed (A, C) implies fixed rate, and hence the feedback generator

$\mathcal{G}_T$ has to be chosen to minimize the average channel input power, which turns out to be a Kalman filter for an associated estimation problem. Let us define $R_T(A,C) := I(W;y^T)/(T+1)$ for a fixed (A,C).

Proposition 2. *Consider the general coding structure in Fig. 3. Fix any $0 \le n \le T$,*

i)

$$P_{T,n}(\mathcal{R}) = \inf_{\substack{A,C,\mathcal{G}_T := \mathcal{G}_T^*(A,C)\\ s.t.\ R_T(A,C)\ge\mathcal{R}}} \frac{1}{T+1}\mathbf{E}u^{T\prime}u^T \tag{13}$$

where $\mathcal{G}_T^*(A,C)$ is the optimal feedback generator for a given (A,C), defined as

$$\mathcal{G}_T^*(A,C) := \arg\inf_{(A,C)\text{fixed},\mathcal{G}_T} \frac{1}{T+1}\mathbf{E}u^{T\prime}u^T. \tag{14}$$

ii)

$$\mathcal{G}_T^*(A,C) = -\widehat{\mathcal{G}}_T^*(A,C)(I - \mathcal{Z}_T^{-1}\widehat{\mathcal{G}}_T^*(A,C))^{-1}, \tag{15}$$

where $\widehat{\mathcal{G}}_T^*(A,C)$ is the strictly causal MMSE estimator (Kalman filter) of r^T given $\bar{y}_T$, i.e.,

$$\widehat{\mathcal{G}}_T^*(A,C) := \arg\inf_{\widehat{\mathcal{G}}_T\in\mathbb{R}^{(T+1)\times(T+1)}} \frac{1}{T+1}\mathbf{E}(r^T - \widehat{\mathcal{G}}_T\bar{y}^T)(r^T - \widehat{\mathcal{G}}_T\bar{y}^T)', \tag{16}$$

where $\widehat{\mathcal{G}}_T$ is strictly lower triangular and $\bar{y}^T := \mathcal{Z}_T^{-1}r^T + N^T$. See Fig. 4 (a) for the associated estimation problem and (b) for $\mathcal{G}_T^*(A,C)$.

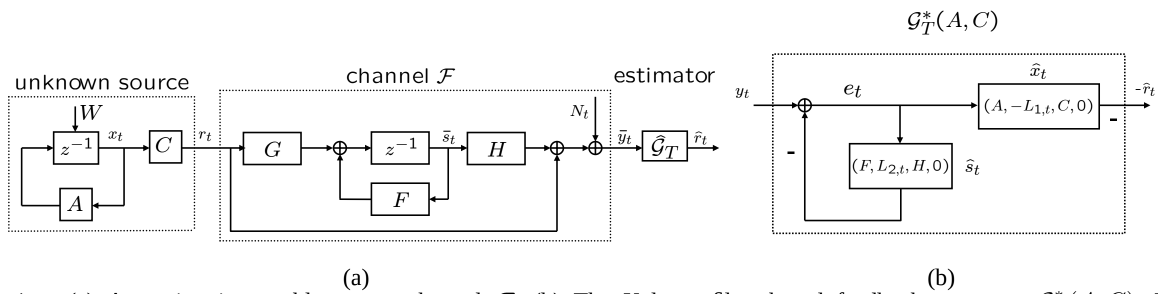


Fig. 4. (a) An estimation problem over channel $\mathcal{F}$. (b) The Kalman filter based feedback generator $\mathcal{G}_T^*(A,C)$. Here $(A,-L_{1,t},C,0)$ with $\hat{x}_t$ denotes a state-space representation with $\hat{x}_t$ being its state at time t, and $\hat{x}_0$ being 0; see (17) and (19) for $L_{1,t}$ and $L_{2,t}$.

Proposition 2 i) says that, we may reformulate the feedback capacity problem as, in step 1, fixing (A,C), i.e. fixing the rate, and minimizing the input power by searching over $\mathcal{G}$, and in step 2, searching over all possible (A,C) subject to the rate constraint. The *role of the feedback generator* $\mathcal{G}$ for any fixed (A,C) is to minimize the input power. Then ii) solves the optimal feedback generator $\mathcal{G}_T^*(A,C)$ by considering the *equivalent* optimal estimation problem in Fig. 4 (a), whose solution is the Kalman filter. Notice that we also obtain the MMSE estimate of W by (**??**), so the Kalman filter leads to both *power efficiency* and *best estimate of the message*.

Proof: i) follows from the definitions of $P_{T,n}(\mathcal{R})$ and $\mathcal{G}_T^*(A,C)$ and Proposition 1. ii) Noting that $u^T = r^T - \hat{r}^T = r^T - \widehat{\mathcal{G}}_T\bar{y}^T$, $\mathbf{E}u^Tu^{T\prime}$ is the MSE of estimating r^T based on observation $\bar{y}^T$. Thus $\widehat{\mathcal{G}}_T^*$ must be the strictly causal MMSE estimator (with one-step prediction). ∎

V. Feedback rate, CRB, and Bode integral

We have shown that in the general coding structure, to ensure power efficiency for a fixed (A,C), we need to design a Kalman-filter based feedback generator. The necessity of the Kalman filter is not surprising given the previous indications in [2], [4], [9], [11], [14], etc. However, the Kalman filter immediately links the feedback communication problem to estimation and control problems. In this section, we present a *unified representation* for the general coding structure (with $\mathcal{G} := \mathcal{G}^*(A,C)$), its estimation theory counterpart, and its control theory counterpart. Then we will establish connections among the information theory quantities, estimation theory quantities, and control theory quantities.

In Fig. 3, fix (A,C) and let $\mathcal{G} := \mathcal{G}^*(A,C)$. Define $\tilde{x}_t := x_t - \hat{x}_t$ and $\tilde{s}_t := s_t - \hat{s}_t$, and

$$\mathbb{A} := \begin{bmatrix} A & 0 \\ GC & F \end{bmatrix}, L_t := \begin{bmatrix} L_{1,t} \\ L_{2,t} \end{bmatrix}, \mathbb{C} := \begin{bmatrix} C & H \end{bmatrix}, \mathbb{D} := \begin{bmatrix} C & 0 \end{bmatrix}, \mathbb{X}_t := \begin{bmatrix} \tilde{x}_t \\ \tilde{s}_t \end{bmatrix}. \quad (17)$$

The equivalence between the communication system in Fig. 3 and the estimation system in Fig. 4 leads to the unified representation of the two systems:

$$\begin{cases} \mathbb{X}_{t+1} &= (\mathbb{A} - L_t\mathbb{C})\mathbb{X}_t - L_t N_t = \mathbb{A}\mathbb{X}_t - L_t e_t, \ \mathbb{X}_0 := [W', 0']' \\ e_t &= \mathbb{C}\mathbb{X}_t + N_t \\ u_t &= \mathbb{D}\mathbb{X}_t. \end{cases} \quad (18)$$

(18) may also be viewed as a control system where we want to minimize the power of u by appropriately choosing L_t. This is a *minimum energy control* problem [15]. Here e_t is called the *innovation* (this innovation differs from those defined in [5] or [10]), which plays a significant role in Kalman filtering. One fact is that $\{e_t\}$ is a white process, that is, its covariance matrix $K_e^{(T)}$ is a diagonal matrix. Another fact is that e^T and y^T determine each other causally, and thus $h(e^T) = h(y^T)$ and $|K_y^{(T)}| = |K_e^{(T)}|$. Let $\Sigma_t := \mathbf{E}\mathbb{X}_t\mathbb{X}_t'$, then it holds that

$$\Sigma_{t+1} = \mathbb{A}\Sigma_t\mathbb{A}' - \frac{\mathbb{A}\Sigma_t\mathbb{C}'\mathbb{C}\Sigma_t\mathbb{A}'}{\mathbb{C}\Sigma_t\mathbb{C}'+1}, L_t := \frac{\mathbb{A}\Sigma_t\mathbb{C}'}{K_{e,t}}, K_{e,t} := \mathbf{E}(e_t)^2 = \mathbb{C}\Sigma_t\mathbb{C}' + 1 \quad (19)$$

Proposition 3. *For any fixed* $0 \le n \le T$ *and* (A,C)*, it holds that*

$$\begin{aligned} I(W;y^T) = & \ \frac{1}{2}\sum_{t=0}^{T}\log K_{e,t} &=& \ \frac{1}{2}\sum_{t=0}^{T}\log(\mathbb{C}\Sigma_t\mathbb{C}'+1) &= \frac{1}{2}\log|\mathcal{I}_{W,T}| \\ = & \ \frac{1}{2}\log|\text{CRB}_{W,T}|^{-1} &=& \ \frac{1}{2}\log|\textit{MSE}_{W,T}|^{-1}; & \\ P_{T,n}(A,C) = & \ \frac{1}{T+1}\sum_{t=0}^{T}\mathbb{D}\Sigma_t\mathbb{D}' &=& \ \frac{1}{T+1}\sum_{t=0}^{T}CA^t\textit{MSE}_{W,t}A^{t\prime}C', & \end{aligned} \quad (20)$$

where $\mathcal{I}_{W,T}$ *is the Bayesian Fisher information matrix of* W*, and* $\text{CRB}_{W,T}$ *is the Bayesian CRB of* W *[16].*

This proposition connects the mutual information to the innovation process and to the Fisher information, (minimum) MSE, and CRB. As a consequence, the finite-horizon feedback capacity $C_{T,n}$ is then linked to the smallest possible Bayesian CRB, i.e. the smallest possible estimation error covariance, and thus the fundamental limitation in information theory is linked to the fundamental limitation in estimation theory.

Proof: Note $h(y^T) = h(e^T)$, $K_{e,t} = \mathbb{C}\Sigma_t\mathbb{C}' + 1$, and $\mathbf{E}(u_t)^2 = \mathbb{D}\Sigma_t\mathbb{D}' = C\mathbf{E}(\tilde{x}_t)^2C'$. For the estimation problem $\bar{y}^T = \mathcal{Z}_T^{-1}\Gamma W + N^T$, $\text{MSE}_{W,T}$ can be computed by Th. 12.1 of [17]. ∎

VI. Asymptotic behaviors of the system

By far we have completed our analysis in finite-horizon. We have shown that the optimal encoder/decoder must contain a Kalman filter, and connected the feedback communication problem to an estimation problem and a control problems. Below, we consider the steady-state communication problem, by studying the limiting behavior (T going to infinity) of the finite-horizon solution while fixing the encoder dimension $(n+1)$.

A. Asymptotic behaviors of the systems

The time-varying (singular) Kalman filter in (18) converges to a steady-state (cf. [18]), namely (18) is *stabilized* in closed-loop, u_t, e_t, and y_t will converge to steady-state distributions, and Σ_t, L_t, $\mathcal{G}_t^*(A,C)$, $\widehat{\mathcal{G}}_t^*$, and $K_{e,t}$ will converge to their steady-state values, for example, $L := \mathbb{A}\Sigma\mathbb{C}'/K_e$, $K_e = \mathbb{C}\Sigma\mathbb{C}' + 1$, and Σ is the unique stabilizing solution to the Riccati equation

$$\Sigma = \mathbb{A}\Sigma\mathbb{A}' - \mathbb{A}\Sigma\mathbb{C}'\mathbb{C}\Sigma\mathbb{A}'/(\mathbb{C}\Sigma\mathbb{C}'+1). \quad (21)$$

Thus (18) has the same asymptotic behavior as the LTI system obtained by letting $L_t = L$ for all t. This LTI system is easy to analyze (e.g., it allows transfer function based study) and to implement. The result in the steady-state minimum-energy control problem says that the transfer function from N to e is an *all-pass* function in the form of

$$\mathcal{T}_{Ne}(z) = \prod_{i=0}^{k} \frac{z-a_i}{z-a_i^{-1}} \tag{22}$$

where a_0^k are the unstable eigenvalues of A or $\mathbb{A}$ (noting that F is stable).

Now fix (A, C) and let the horizon T in the general coding structure go to infinity. Let $\mathcal{H}(e)$ be the entropy rate of $\{e_t\}$, $DI(A) := \prod_{i=0}^{k} |a_i|$ be the *degree of instability* of A, and $S(e^{2\pi j\theta})$ be the spectrum of the sensitivity function [12].

Proposition 4. *Consider the general coding structure in Fig. 3. For any $n \geq 0$ and (A, C),*

$$\begin{array}{llllll} & R_{\infty,n}(A,C) & := & \lim_{T\to\infty} \frac{1}{T+1} I(W; y^T) & = & \mathcal{H}(e) \\ = & \log DI(A) & = & \int_{-\frac{1}{2}}^{\frac{1}{2}} \log S(e^{2\pi j\theta}) d\theta & = & \frac{1}{2}\log(\mathbb{C}\Sigma\mathbb{C}' + 1) \\ = & \lim_{T\to\infty} \frac{\log|\mathcal{I}_{W,T}|}{2(T+1)} & = & -\lim_{T\to\infty} \frac{\log|MSE_{W,T}|}{2(T+1)} & = & -\lim_{T\to\infty} \frac{\log|CRB_{W,T}|}{2(T+1)}; \\ & P_{\infty,n}(A,C) & := & \lim_{T\to\infty} \frac{1}{T+1} \mathbf{E} u_T u_T' & = & \mathbb{D}\Sigma\mathbb{D}'. \end{array} \tag{23}$$

Proposition 4 links the asymptotic feedback rate to the entropy rate of the innovation process, to the degree of instability and Bode sensitivity integral [12], to the asymptotic increasing rate of Fisher information, and to the asymptotic decay rate of MSE and of CRB.

The presence of stable eigenvalues in A does not affect the rate (see also [12]). Stable eigenvalues do not affect $P_{\infty,n}(A, C)$, either, since the initial condition response associated with the stable eigenvalues can be tracked with zero power (i.e. zero MSE). So, we can achieve $C_{\infty,n}$ by a sequence of purely unstable (A, C), and hence the communication problem is related to the tracking of unstable source over a communication channel [11], [12].

Proof: i) By (22), the power spectrum of $\{e_t\}$ is flat with magnitude $DI(A)^2$. Then the results follow from [12], the Grenander-Szego theorem, and the Cesaro mean [20]. ■

VII. Achievability of C_s

In this section, we will first prove that $C_{\infty,m-1} = C_s$, which will lead to the optimality of our encoder/decoder design in Section III in the information sense, and then show that our design achieves C_s in the operational sense.

A. The optimal Gauss-Markov signalling strategy

[10] proved that for each input in the form of (3), there exists a Gauss-Markov (GM) input that leads to the same directed information and same input power. The GM input takes the form

$$u_t = d_t' \tilde{s}_{s,t} + \mathcal{E}_t, \tag{24}$$

where $d_t \in \mathbb{R}^m$ is a time-varying gain; $\{\mathcal{E}_t\}$ is a zero-mean white Gaussian process and $\mathcal{E}_t$ is independent on N^{t-1}, u^{t-1}, and y^{t-1}; and $\tilde{s}_{s,t}$ is generated by a Kalman filter

$$\begin{cases} \tilde{s}_{s,t} & := & s_t - \hat{s}_{s,t} \\ \hat{s}_{s,t+1} & = & F\hat{s}_{s,t} + L_{s,t} e_t, \ \hat{s}_{s,0} = 0 \\ e_t & = & y_t - H\hat{s}_{s,t}, \end{cases} \tag{25}$$

If one lets $d_t = t$ and $K_{\mathcal{E}}^{(t)} = K_{\mathcal{E}}$ for all t, then the search over all possible d and $K_{\mathcal{E}}$ solves C_s. We remark that [10] was focused more on the structure of the optimal input distribution and capacity computation, instead of designing a coding scheme; how to encode/decode a message (rather than using a random coding argument) is not clear from [10].

Now we claim that $K_{\mathcal{E}} = 0$, namely $\{\mathcal{E}_t\}$ vanishes in steady-state. [6] This leads to a further simplification of the results in [10].

[6] $K_{\mathcal{E}} = 0$ was also conjectured and numerically verified by Shaohua Yang (personal communication).

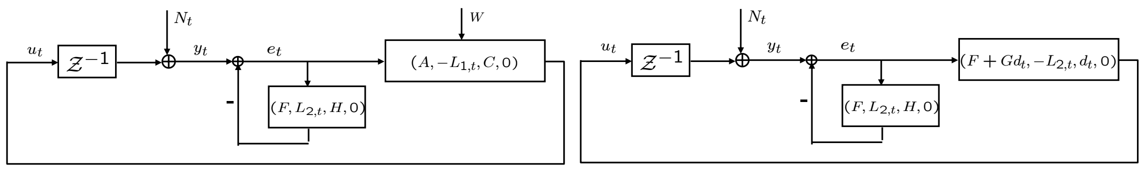
Proposition 5. *For the GM input (24) to achieve C_s, it must hold that $K_{\mathcal{E}} = 0$.*

Proof: (sketch) Assume that for some $K_{\mathcal{E}} \neq 0$, the GM input can achieve C_s. Fix the corresponding optimizing S, T, L_2, and d. We can show that this leads to: 1) The whiteness of $\{\tilde{y}_t\}$; 2) $L_{s,2} = G$; 3) $K_{\mathcal{E}} = 0$ and hence contradiction. ∎

The vanishing of $\{\mathcal{E}_t\}$ in steady-state helps us to show that, in our general coding structure, the encoder dimension needs not be higher than the channel dimension in order to achieve C_s, namely to achieve C_s we need A to have at most m unstable eigenvalues. This also follows that the control-oriented communication scheme in [12] can achieve C_s.

Proposition 6. *For channel $\mathcal{F}$ with order $m \geq 1$, $C_{\infty,n} = C_s$ for $n \geq m-1$.*

Proof: (sketch) We rewrite the general coding structure (with $\mathcal{G} := \mathcal{G}^*(A, C)$) as in Fig. 5 (a), and rewrite the system driven by the GM optimal input (with $\mathcal{E} = 0$) as in Fig. 5 (b). Note that the presence of W does not affect the steady-state. It is then clear that the dimension of A needs not be greater than the dimension of F. ∎



(a) (b)

Fig. 5. (a) A transform of the block diagram of the general coding structure. (b) An equivalent form of the communication system driven by GM inputs.

B. Achieving C_s

In this subsection, we show that our coding scheme achieves C_s in the information sense as well as in the operational sense.

Proposition 7. *For the coding scheme described in Theorem 1, $R_{\infty,n^*}(A^*, C^*) = C_s(\mathcal{P})$ and $P_{\infty,n^*}(A^*, C^*) = \mathcal{P}$.*

Proof: From Proposition 6, the optimization

$$[A^{opt}, C^{opt}, \Sigma^{opt}] := \arg \inf_{\substack{A,C \\ s.t.\ (21), \log DI(A) \geq \mathcal{R}}} \mathbb{D}\Sigma\mathbb{D}', \tag{26}$$

with $n = m-1$ attains $P_{\min}(\mathcal{R})$. Note that the stable eigenvalues (if any) in A^{opt} can be removed without affecting the optimality. Moreover, without loss of generality, we may assume that (A, C) is in the observable canonical form. Additionally, imposing $a_n = \pm 2^{\mathcal{R}}$ guarantees that $\log DI(A) \geq \mathcal{R}$. Then the optimization (7) achieves $P_{\min}(\mathcal{R})$. ∎

Proposition 8. *The system constructed in Theorem 1 transmits the analog source $W \sim \mathcal{N}(0, I)$ at a rate $C_s(\mathcal{P})$, with MSE distortion $D(C_s(\mathcal{P}))$, where $D(\cdot)$ is the distortion-rate function.*

Proof: Note that $\text{MSE}(\hat{W}_t) = A^{-t-1}\Sigma_{x,t+1}A'^{-t-1}$ and hence R is no smaller than $\log|\det A|$. Here $\Sigma_{x,t+1} := [I, 0]\Sigma_{t+1}[I, 0]'$. ∎

Proposition 9. *The system constructed in Theorem 1 transmits a digital message W from the transmitter to the receiver at a rate arbitrarily close to $C_s(\mathcal{P})$ with PE_T decays doubly exponentially.*

Proof: The proof is in essence a Schalkwijk-Kailath type argument [2], [12], [13]. We can also show that

$$PE_T = 1 - \prod_{i=0}^{n}(1 - 2Q(\frac{\sigma_{T,i}^{-\epsilon}}{2})), \tag{27}$$

where $\sigma_{T,i}$ is the ith eigenvalue of $\text{MSE}_{W,T}$. Thus, if $\text{MSE}_{W,T}$ decays to zero exponentially (which is indeed the case), PE_T decays to zero doubly exponentially. ∎

Note that, in both the analog and digital communication case, the coding length needed for a pre-specified performance level can be pre-determined since $\Sigma_{x,T}^*$ can be solved off-line. Moreover, because the probability of error decays doubly exponentially, it leads to much shorter coding length than forward transmission.

VIII. Conclusions and future work

We presented a coding scheme to achieve the stationary feedback capacity for a Gaussian channel with feedback. The scheme is essentially the Kalman filter algorithm, and its construction involves only a finite dimensional optimization problem. We established connections to estimation and control, and in particular, the encoder may be seen as a control system, and the decoder may be seen as an estimation system, as pointed by Sanjoy Mitter and in [11], [21]. We have seen that concepts in estimation theory and control theory, such as MMSE, CRB, minimum-energy control, etc., are useful in studying a feedback communication system. We also verified the results by simulations (not reported here).

Our ongoing research includes convexifying the optimization problem (7) to reduce the computation complexity, and finding a more feasible scheme to fight against feedback noise while keeping the feedback signal bounded. In future, we will further explore the connections among information, estimation, and control in more general setups (such as MIMO channels with feedback).

References

[1] J.P.M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback Part I: No bandwidth constraint. *IEEE Trans. Inform. Theory*, IT-12:172–182, 1966.

[2] J.P.M. Schalkwijk. A coding scheme for additive noise channels with feedback Part II:Bandlimited signals. *IEEE Trans. Inform. Theory*, IT-12:183–189, 1966.

[3] J. Omura. Optimum linear transmission of Analog data for channels with feedback. *IEEE Trans. Inform. Theory*, 14:38-43, 1968.

[4] S.A. Butman. Linear feedback rate bounds for regressive channels. *IEEE Trans. Inform. Theory*, IT-22:363–366, 1976.

[5] T. Cover and S. Pombra. Gaussian feedback capacity. *IEEE Trans. Inform. Theory*, IT-35:37–43, 1989.

[6] L.H. Ozarow. Random coding for Gaussian channels with feedback. *IEEE Trans. Inform. Theory*, 36:17–22, 1988.

[7] G. Kramer. Feedback strategies for white Gaussian interference networks. *IEEE Trans. Inform. Theory*, 48:1423–1438, 2002.

[8] A. Shahar-Doron and M. Feder. On a capacity achieving scheme for the colored Gaussian channel with feedback. *Proc. ISIT*, page 74, 2004.

[9] S. Tatikonda and S. Mitter. The capacity of channels with feedback. *IEEE Trans. Inform. Theory*, submitted 2001.

[10] S. Yang, A. Kavcic, and S. Tatikonda. Feedback capacity of power constrained Gaussian channels with memory. *IEEE Trans. Inform. Theory*, submitted 2003.

[11] A. Sahai. *Anytime Information Theory*. PhD thesis, MIT, 2001.

[12] N. Elia. When Bode meets Shannon: Control-oriented feedback communication schemes. *IEEE Trans. Autom. Contr.*, 49:1477–1488, 2004.

[13] Young-Han Kim. The feedback capacity of the first-order MA Gaussian channel. *http://arxiv.org/abs/cs/0411036*, 2004.

[14] S. Mitter and N. Newton. Information and entropy flow in the Kalman-Bucy filter. *J. of Stat. Phys.*, 118:145-176, 2005.

[15] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems*. John Wiley & Sons, 1972.

[16] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. John Wiley and Sons, 1968.

[17] S. M. Kay. *Fundamentals of Statistical Signal Processing I: Estimation Theory*. Prentice-Hall PTR, 1998.

[18] K. Gallivan, X. Rao, and P. Van Dooren. Singular riccati equations stabilizing large-scale systems. *Lin. Alg. Appl.*, 2005.

[19] T. Kailath, A. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, 2000.

[20] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[21] S. Tatikonda, and S. Mitter. Control over noisy channels. *IEEE Trans. Autom. Contr.*, 49:1196 - 1201, 2004.

[22] N. Elia. The information cost of disturbance rejection. *Mediterranean Conference on Control and Automation*, 2005.