Convergence of fundamental limitations in feedback communication, estimation, and feedback control over Gaussian channels[*]

JIALING LIU[†] AND NICOLA ELIA

This paper presents a unifying perspective on communication, estimation, and control. A feedback communication system over a Gaussian channel is reformulated as an estimation system (i.e. a Kalman filter) and also as a feedback control system. Therefore, a fundamental limitation applicable in one system can be translated into a fundamental limitation applicable in another system and vice versa. Specifically, the achievable rate in the communication system is equal to half the Cramer-Rao bound decay rate or minimum mean-square error decay rate in the estimation system, and also equal to the Bode integral in the control system. Furthermore, the optimal trade-offs for each of the following three pairs of quantities are equal: first, the pair of power and rate in communication, second, the pair of the (causal) prediction performance and (anti-causal) smoothing performance in estimation, and third, the pair of the output regulation performance and disturbance rejection measure in control. All these trade-offs can be interpreted as the tradeoff between causality and anti-causality. Extensions to more general channels are briefly discussed.

KEYWORDS AND PHRASES: Fundamental limitations; confluence of feedback communication, estimation, and control; Kalman filtering; minimum mean-square error; Bode integral

1. Introduction

There is an increasing trend to investigate the interactions of information, estimation, and control; see [1–13] and references therein. Specifically, the communication system in which the transmitter has access to noiseless feedback of channel outputs has been a central focus since it possesses aspects of a communication system, an estimation system, and a control system. One of the most important case of such is the system over a Gaussian channel with noiseless feedback. The *fundamental limitation* of this system, from a communication perspective, is characterized by the feedback capacity and capacity-achieving codes and has been studied using several approaches. These approaches are categorized roughly (by no means strictly, since they are intrinsically related) as follows. 1) *Estimation theory related approaches.* This approach utilizes concepts such as maximum likelihood (ML) or minimum mean-square error (MMSE) estimates in constructing the coding schemes, such as the capacity-achieving Schalkwijk-Kailath (SK) codes for additive white Gaussian noise (AWGN) channels motivated by the Robbins-Monro stochastic approximation and recursive ML algorithm [1, 2]. See also [14–17]. 2) *Information theoretic approaches*, most notably those based on the Cover-Pombra scheme (CP scheme for short) built upon the asymptotic equipartition property, the mutual information between the message and channel outputs, and the directed information from the input to output of a channel. In addition, Kim [18] proved the capacity theorem and obtained the capacity expression. See also [6, 19–23]. 3) *Control theory related approaches*, which regards the feedback communication problems as optimal control problems (cf. e.g. [3–6, 22, 24]). Tatikonda and Mitter [6] reformulated the capacity problem as an optimal control problem and proposed dynamic programming based solutions. Sahai and Mitter [5] investigated the problem of tracking unstable sources over a channel and introduced the notion of anytime capacity, revealing the connections of the fundamental limitations between communication and control. Elia [4] established an equivalence between reliable communication and stabilization over Gaussian channels, showed that the achievable transmission rate is equal to the Bode integral (BI) of the associated control system, and generalized the SK codes via control-oriented methods. See also [25, 26] and so on for a related study of analyzing a control system under communication constraints. There also exist studies utilizing more than one approach, such as Ardestanizadeh *et al* [27, 28] showed the connections of the information-theoretic capacity and MMSE exponent with a certain type of control policies.

As we have seen, different approaches have been shown useful in addressing the fundamental limitations associated with the Gaussian feedback communication system. This paper attempts to present a converging point for the studies of this system based on a perspective that unifies information, estimation, and control. We demonstrate that the communication problem can be reformulated as an optimal *estimation* problem or an optimal *control* problem. Specifically, we will show that many coding schemes in the existing literature have to utilize the Kalman filter (KF) algorithm or are reformulations of Kalman filters: First, the CP scheme needs to implement a KF to achieve power efficiency (defined as achieving minimum channel input power to support a given rate) and optimal receovery of the transmitted message (see Section 4.1); second, the SK code can be easily obtained from a KF of an appropriate form (see Sections 2 and 5.1); and third, the control-oriented schemes can be derived from a KF by the duality between control and estimation [29, 30] (see Section 5.1). Therefore, we will conclude that the three approaches described above are intimately connected, and consequently, one can be used to complement another when addressing various problems over Gaussian channels.

As a result, we can establish certain corresponding relationship for the fundamental limitations in feedback communication, estimation, and feedback control. The achievable rate of the feedback communication system is alternatively given by the decay rate of the Cramer-Rao bound (CRB) (or equivalently, decay rate of the MMSE) for the associated estimation system, or given by the BI of the associated control system, in both finite horizon or infinite horizon (see Sections 6.2 and 7.2). In addition, the fundamental limitations in terms of the *optimal tradeoffs in feedback communication, estimation, and feedback control coincide, all of which may be interpreted as the tradeoff between causality and anti-causality.* In feedback communication, this fundamental limitation is the capacity, i.e., the optimal tradeoff between the input power (i.e. the cost) causally generated by the feedback loop and information rate (i.e. the payoff) which is independent of feedback; in the associated estimation system, it is characterized by the optimal tradeoff between the causal mean-square error (MSE) and (anti-causal) smoothing MSE decay rate; and in the associated control system, by the optimal tradeoff between the variance of a regulated output generated by causal feedback (a control performance measure) and the BI (or degree of anti-causality or degree of instability, which is the disturbance rejection measure). For each of the tradeoff pairs, if one wishes to keep the cost (i.e. the first in the pair) low, the payoff (i.e. the second in the pair) cannot exceed some constant determined only by the underlying channel and the cost (see Section 6.3).

The above-mentioned equivalence relationship leads to a simple equivalence between generalized SK codes and the KF, which provides a convenient way to obtain a feedback communication scheme from an estimation problem. This leads to an *information theoretic characterization of KF*; that is, the KF is not only a device to provide sufficient statistics (which was shown in [22]), but also a device to ensure the power efficiency and to recover the message optimally.

Different from papers studying channel capacity problems and establishing capacity theorems or coding theorems, this paper is focused on discussing the confluence of information, estimation, and control, such as how one problem may be transformed into another, and how the fundamental limitations in these problems are related. Hence only some aspects of the channel capacity that are pertinent to estimation and control will be discussed in this paper. We do not attempt to address the capacity problem per se (e.g. proving the coding theorem), and the interested reader is refer to [18, 31] and related papers for the characterization of capacity.

Throughout the paper, for simplicity, *channel* refers to *feedback channel*, *Gaussian channel* refers to *Gaussian channel with memory and feedback*, and *capacity* refers to *feedback capacity*, unless otherwise specified. We use underlines to specify vectors, and boldface to specify matrices. We represent transpose by $'$. $\boldsymbol{K} > 0$ (or ≥ 0) means that the symmetric matrix $\boldsymbol{K}$ is positive definite (or positive semi-definite). We represent time indices by subscripts, such as y_t. We denote by $\underline{y}^T$ the vector $[y_0, y_1, \ldots, y_T]'$, and $\|\underline{y}^T\|$ be the vector 2-norm of $\underline{y}^T$. We assume that the starting time of a process is 0. A finite-dimensional linear time-invariant (FDLTI) system is denoted using a scripted symbol, e.g., $\mathcal{Z}(z)$, and as it can be alternatively captured by a matrix, we represent the matrix associated with the linear system $\mathcal{Z}(z)$ by $\boldsymbol{\mathcal{Z}}$ (boldface scripted symbol). We use $(\boldsymbol{A}, \underline{B}, \underline{C}', D)$ to represent system

$$(1) \qquad \begin{cases} \underline{x}_{t+1} = \boldsymbol{A}\underline{x}_t + \underline{B}u_t \\ y_t = \underline{C}'\underline{x}_t + Du_t. \end{cases}$$

2. Motivating example: Kalman filter based optimal communication over AWGN channel

To help the reader understand the intuition behind our study, we present a simple example over an AWGN channel before we study the general Gaussian channels with memory. First, we introduce a simple KF system (see Fig. 1 (a)), followed by a straightforward rewrite of it (see Fig. 1 (b)), which then may be used for information transmission at the Shannon capacity

rate. Finally, we show that this communication system is equivalent [1] to the SK scheme in the sense that both follow the same procedures and dynamics and generate identical corresponding signals for any time horizon. This implies that the celebrated SK scheme in fact performs the KF algorithm. It motivates the further exploration of the connections among feedback communication, estimation, and feedback control.

2.1. A Kalman filter problem

Consider a standard KF problem for a first-order unstable LTI system with noisy measurements:

(2)   to-be-estimated system: $\begin{cases} x_{t+1} = ax_t \\ r_t = cx_t \\ \bar{y}_t = r_t + w_t, \end{cases}$

for time $t = 0, 1, \ldots$, where x_0 is a continuous random variable with zero mean and finite variance, $|a| > 1$ (namely the system is unstable), a and c are known, and $w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ is the AWGN. The KF provides the linear MMSE estimate of $\{x_t\}$ based on the noisy measurement process $\{\bar{y}_t\}$, and if x_0 is Gaussian, the estimate is the MMSE estimate. The (steady-state) [2] KF is described as (see Fig. 1 (a) for the block diagram)

(3)   Kalman filter: $\begin{cases} \widehat{x}_{t+1} = a\widehat{x}_t + le_t \\ \widehat{r}_t = c\widehat{x}_t \\ e_t = \bar{y}_t - c\widehat{x}_t, \end{cases}$

where

(4) $$l := \frac{a\mathbf{\Sigma}c}{1 + c^2\mathbf{\Sigma}}$$

is the asymptotic *Kalman filter gain*, and $\mathbf{\Sigma}$ is the asymptotic MSE for $\widehat{x}_t$ (i.e. $\mathbf{\Sigma} = \lim_{t \to \infty} \mathbf{E}(x_t - \widehat{x}_t)^2$), which is the positive solution to the discrete-time

---

[1]In here and subsequent places, "equivalence" is according to the notion described in Appendix A.1.

[2]Though $\{\bar{y}_t\}$ is neither stationary nor even asymptotically stationary, a time-varying or time-invariant (steady-state) KF can be built to guarantee *bounded* error variance for estimating x_t, and the difference between the time-varying one and time-invariant one vanishes as time increases, as pointed out in Chapter 14 of [29].

algebraic Riccati equation (DARE)

$$(5) \qquad\qquad \mathbf{\Sigma} = a^2\mathbf{\Sigma} - \frac{a^2c^2\mathbf{\Sigma}^2}{1 + c^2\mathbf{\Sigma}}.$$

Solving the DARE, we obtain

$$(6) \qquad\qquad \mathbf{\Sigma} = \frac{a^2 - 1}{c^2}, \quad l = \frac{a^2 - 1}{ac}.$$
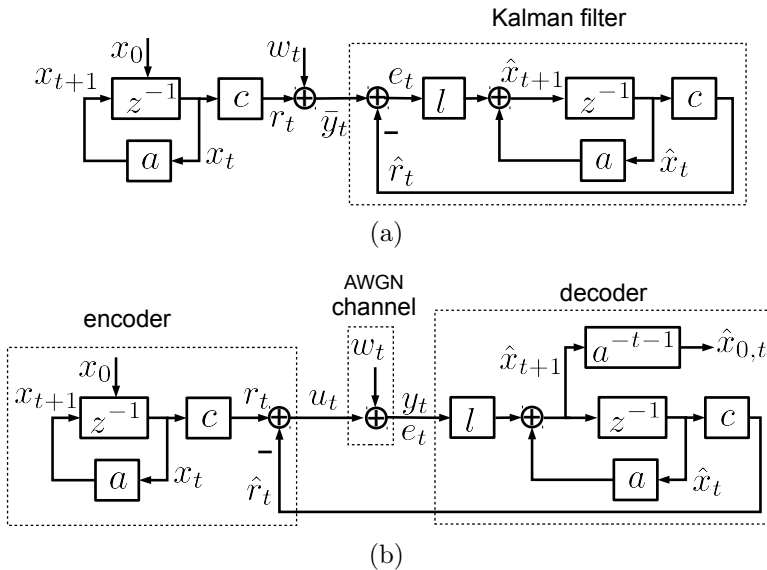

Figure 1: (a) A KF problem. (b) A KF-based coding scheme.

2.2. KF-based feedback communication

Next, as illustrated in Fig. 1 (b), we introduce a *feedback communication coding scheme over an AWGN channel* by slightly changing the KF problem shown in Fig. 1 (a). Rather than closing the loop to the right of the AWGN w_t (i.e. adding $(-\widehat{r}_t)$ to $\bar{y}_t$ to obtain e_t) as done in the KF, in Fig. 1 (b), the loop is closed to the left of the AWGN w_t (i.e. adding $(-\widehat{r}_t)$ to r_t to obtain u_t). It is straightforward to verify that all corresponding signals remain the same after the change and hence the two systems are considered as equivalent.

(Note that in Fig. 1 (b), $y_t = e_t$.) As indicated in Fig. 1 (b), one can identify a communication system with an encoder, AWGN channel, and a decoder, all of which will be described in the following for time $t = 0, 1, \ldots$.

$$(7) \qquad \text{AWGN channel:} \qquad y_t = u_t + w_t,$$

where u_t is the channel input, $w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ is the channel noise, and y_t is the channel output.

At time t, the encoder can access $\widehat{r}_t$ (generated from $\underline{y}^{t-1}$) via the noiseless feedback link and operates according to:

$$(8) \qquad \text{encoder dynamics:} \qquad \begin{cases} x_{t+1} = a x_t \\ r_t = c x_t \\ u_t = r_t - \widehat{r}_t \end{cases}$$

where a and c are encoder design parameters such that $|a| > 1$ and $c \neq 0$. The *encoding procedure* is as follows. Fix the number of channel uses to be $(T + 1)$ (i.e. the system runs from time 0 to time T) and a power budget $\mathcal{P} > 0$. Given a set of K_T equally likely messages, map the kth message into a message point $M^{(k)}$ according to

$$(9) \qquad M^{(k)} := \frac{\sqrt{\mathcal{P}}}{c} \left(-1 + \frac{2k}{K_T - 1} \right); \ k = 0, 1, \ldots, K_T - 1.$$

This is known to both the transmitter and receiver *a priori*. Suppose the kth message is to be transmitted across the channel. To encode, let $x_0 := M^{(k)}$, the message point representing a selected message.

The decoder operates according to

$$(10) \qquad \text{decoder dynamics:} \qquad \begin{cases} \widehat{x}_{t+1} = a \widehat{x}_t + l y_t \\ \widehat{r}_t = c \widehat{x}_t \\ \widehat{x}_{0,t} = a^{-t-1} \widehat{x}_{t+1}, \end{cases}$$

and the *decoding procedure* is to simply map $\widehat{x}_{0,T}$ (or $\widehat{x}_{0,T}/(1 - a^{-2T-2})$) into the closest message point referred to as $\widehat{M}^{(k)}$.

In summary, the KF-based communication system dynamics is:

$$
(11) \quad
\begin{array}{c}
\text{KF-based} \\
\text{communication}
\end{array}
\left\{
\begin{array}{l}
x_{t+1} = a x_t \\
r_t = c x_t \\
u_t = r_t - \widehat{r}_t \\
y_t = u_t + w_t = e_t \\
\widehat{x}_{t+1} = a \widehat{x}_t + l e_t \\
\widehat{r}_t = c \widehat{x}_t \\
\widehat{x}_{0,t} = a^{-t-1} \widehat{x}_{t+1}
\end{array}
\right.
\begin{array}{l}
\left. \begin{array}{c} \\ \\ \end{array} \right\} \text{encoder} \\
\\
\quad\quad \text{AWGN channel} \\
\left. \begin{array}{c} \\ \\ \end{array} \right\} \text{decoder}
\end{array}
$$

The following proposition shows that, the Kalman filter in (3), an optimal estimator for a process generated by a continuous random variable and perturbed by AWGN, can be transformed into an optimal communication system to transmit digital messages over an AWGN channel.

Proposition 1. *The feedback communication system together with the described encoding and decoding procedures can achieve the feedback capacity*

$$
(12) \quad\quad\quad C_{fb}(\mathcal{P}) = \frac{1}{2} \log(1 + \mathcal{P}),
$$

under an average channel input power constraint

$$
(13) \quad\quad\quad \frac{1}{T+1} \mathbf{E} \| \underline{u}^T \|^2 \le \mathcal{P}
$$

for any T.

Proof. Fix any $\epsilon > 0$, an arbitrarily small slack from the capacity C_{fb}. Then let $a := \sqrt{1 + \mathcal{P}}$, $c \ne 0$, $K_T := \lfloor a^{(T+1)(1-\epsilon)} \rfloor$, and follow the above-described encoding and decoding dynamics and procedures. One can derive that

$$
(14) \quad\quad\quad \mathbf{E}(x_t - \widehat{x}_t)^2 = a^{-2t} \mathbf{E}(x_0)^2 + l^2 \sum_{j=0}^{t-1} a^{-2j},
$$

and then straightforward computations show that the channel input power satifies

$$
(15) \quad \mathbf{E}(u_t)^2 = c^2 \mathbf{E}(x_t - \widehat{x}_t)^2 = a^{-2t}(c^2 \mathbf{E}(x_0)^2 - \mathcal{P}) + \mathcal{P} \le \mathcal{P}
$$

since $x_0 := M^{(k)}$ and $|c M^{(k)}| \le \sqrt{\mathcal{P}}$ for any k according to (9). Therefore, (13) holds.

Additionally, for any fixed x_0, the decoder generates

$$(16) \qquad \widehat{x}_{0,T} = (1 - a^{-2T-2})x_0 + a^{-2T-2}\sum_{j=0}^{T} a^{j+1}lw_j,$$

which maps to the correct message point if

$$(17) \qquad |x_0 - \widehat{x}_{0,T}| < d_T := \frac{\sqrt{\mathcal{P}}}{c(K_T - 1)},$$

where d_T is half of the spacing between any two neighboring message points. One can see that $(x_0 - \widehat{x}_{0,T})$ is Gaussian conditioned on x_0, with mean $\mu_T := a^{-2T-2}x_0$ and variance $\sigma_T^2 := a^{-2T-2}(1 - a^{-2T-2})\mathcal{P}/c^2$. Then the probability of the decoded message point $\widehat{M}$ not equal to the selected message point M can be shown to satisfy

$$(18) \qquad \begin{aligned} PE_{T|M} &:= \Pr(\widehat{M} \neq M|M) \\ &\leq \Pr(\widehat{x}_{0,T} - x_0 \geq d_T|x_0) + \Pr(\widehat{x}_{0,T} - x_0 \leq -d_T|x_0) \\ &= Q\left(\frac{d_T - \mu_T}{\sigma_T}\right) + Q\left(\frac{d_T + \mu_T}{\sigma_T}\right) \\ &= Q\left(\frac{1}{\sqrt{1 - a^{-2T-2}}}\left(\frac{ca^{T+1}}{K_T - 1} + a^{-T-1}\frac{cx_0}{\sqrt{\mathcal{P}}}\right)\right) \\ &\quad + Q\left(\frac{1}{\sqrt{1 - a^{-2T-2}}}\left(\frac{ca^{T+1}}{K_T - 1} - a^{-T-1}\frac{cx_0}{\sqrt{\mathcal{P}}}\right)\right) \\ &< 2Q\left(\frac{ca^{T+1}}{\xi_T a^{(1-\epsilon)(T+1)}} - a^{-T-1}\right) \\ &\leq 2Q\left(ca^{\epsilon(T+1)} - a^{-T-1}\right), \end{aligned}$$

where $\xi_T := K_T/a^{(1-\epsilon)(T+1)} \leq 1$. The first inequality (as opposed to equality) is because when, say, $x_0 := \sqrt{\mathcal{P}}/c$, then any noise such that $\widehat{x}_{0,T} > x_0$ would not result in a decoding error. The last two inequalities are because the Q-function is strictly decreasing and $c|x_0| \leq \sqrt{\mathcal{P}}$. Since $a > 1$, straightforward computation can show that as T tends to infinity, the above upper bound of $PE_{T|M}$, which is independent of x_0, decreases to zero. Then the probability of decoding error $PE_T := \Pr(\widehat{M} \neq M)$ also vanishes to zero. Hence the rate

$$(19) \qquad \lim_{T \to \infty} \frac{1}{T+1}\log K_T = (1 - \epsilon)C_{fb}(\mathcal{P})$$

is achievable (also cf. [32] for a similar proof for a slightly different setting), and we have proven that (11) equipped with the described encoding and decoding procedures is capacity achieving.   □

Note that in the KF problem described in Section 2.1, the initial condition x_0 is a continuous random variable with zero mean and finite variance, whereas in the KF-based communication problem described in this subsection, the initial condition is a discrete random variable uniformly chosen from the K_T message points. However, as suggested by (16) and (17), asymptotically the distribution of x_0 does not play any role in recovering x_0 from the noisy measurements and hence the KF-based communication scheme can be used to transmit any unknown x_0, regardless of its distribution. See [33] (page 481) for similar comments regarding a common scheme used for transmitting either a Gaussian random variable or digital data. As we will see in Section 2.4, the common scheme in [33] is a rewrite of our KF-based communication scheme.

2.3. Connections with the SK coding scheme

The coding scheme in Fig. 1(b) is a simple reformulation of the celebrated SK coding scheme known to achieve capacity. To see this, a slight variation of the original SK scheme (cf. [2]) is illustrated in Fig. 2 [3] . In this figure, one can identify the encoder, AWGN channel, decoder, and the feedback link with one-step delay.

To see the connection between the two coding schemes, note that in the SK scheme, it holds that

$$(20) \quad \text{SK scheme} \begin{cases} u_t = ga^t(\widehat{x}_{0,t-1} - x_0) & \text{encoder} \\ y_t = u_t + w_t & \text{AWGN channel} \\ \widehat{x}_{0,t} = \dfrac{g^2+1}{a^2}\widehat{x}_{0,t-1} - a^{-t-2}gy_t & \text{decoder} \end{cases}$$

---

[3]The SK variation here performs the same operations every step, as opposed to the scheme in [2] whose initialization step differs from later steps, and the message points $M^{(k)}$'s are defined differently in this paper and in [2]. However, for any time $k > 1$, both the variation in Fig. 2 and the original SK scheme perform identical operations. The resulting signal differences can be shown to vanish as time increases and affect neither the power constraint nor the decay of the error probability (details omitted for brevity). See also Sec. IV-B for [4] for similar comparison and [24] for further comparison of a few SK-type schemes and their variations.
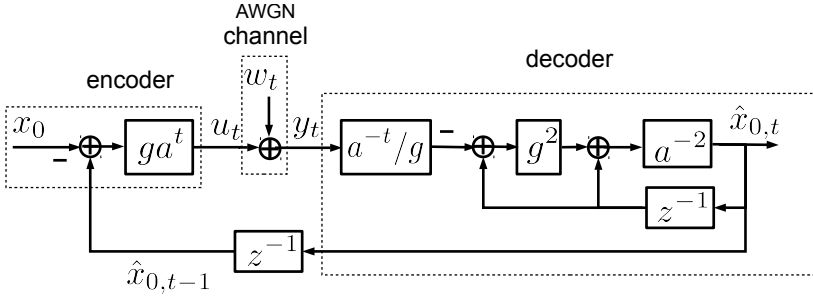
Figure 2:   The SK coding scheme.

(To help the reader verify, note that the signals x_0, w_t, y_t, $\widehat{x}_{0,t}$ in Fig. 2 correspond to θ, Z_{i1}, $Y_{i1}(X_{i1})$, $X_{i+1,1}$ in [2], resp.) If we let $g := -\sqrt{\mathcal{P}}$, then (20) leads to

(21)
$$u_t = \sqrt{\mathcal{P}}a^t(x_0 - \widehat{x}_{0,t-1})$$
$$\widehat{x}_{0,t} = \widehat{x}_{0,t-1} + a^{-t-2}\sqrt{\mathcal{P}}y_t.$$

Straightforward computations can show that the KF-based coding scheme also follows (21) if we let $c := \sqrt{\mathcal{P}}$ (and hence $l = \sqrt{\mathcal{P}}/a$), namely, the KF-based coding scheme and the SK scheme generate identical channel inputs u_t, outputs y_t, and decoder estimates $\widehat{x}_{0,t}$ respectively at each time.

2.4. A stable variation of the KF-based coding scheme

The KF-based coding scheme illustrated in Fig. 1 generates unbounded signals such as r_t (noticing $r_t = ca^t x_0$ where $|a| > 1$). The SK coding scheme illustrated in Fig. 2 generates unbounded values such as a^t. A simple reformulation, however, leads to a stable coding scheme with all involved signals and values stable. See Fig. 3 for the stable formulation.

The system dynamics is

(22)   KF-based stable coding scheme
$$\left\{ \begin{array}{l} \widetilde{x}_t = a\widetilde{x}_{t-1} - ly_{t-1} \\ u_t = c\widetilde{x}_t \\ y_t = e_t = u_t + w_t \\ \widehat{x}_{0,t} = \widehat{x}_{0,t-1} + a^{-t-1}ly_t \end{array} \right. \begin{array}{l} \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{ encoder} \\ \\ \text{AWGN channel} \\ \text{decoder} \end{array}$$

where $\widetilde{x}_{-1} := M/a$ if M is the selected message point (i.e. $\underline{\widetilde{x}}_0 = M$), $y_{-1} := 0$, and $\widehat{x}_{0,-1} = 0$. To see how this formulation may be derived from (11), simply
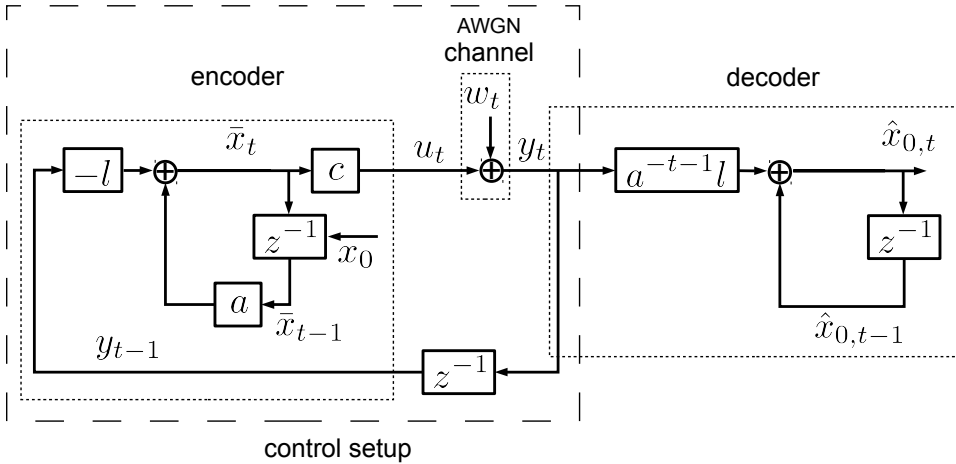
Figure 3: A stable variation of the KF-based coding scheme.

let

$$\widetilde{x}_t := x_t - \widehat{x}_t \tag{23}$$

in (11) and then (22) will follow.

To see that it is stable, note that

$$\begin{aligned}
\widetilde{x}_t &= a\widetilde{x}_{t-1} - l(c\widetilde{x}_{t-1} + w_{t-1}) \\
&= (a - lc)\widetilde{x}_{t-1} - lw_{t-1} \\
&= a^{-1}\widetilde{x}_{t-1} - lw_{t-1},
\end{aligned} \tag{24}$$

where we have used the definition of l in (6). As $|a^{-1}| < 1$, (24) is stabilized and hence all signals and parameters in (22) are bounded.

We remark that the stable variation in (22) coincides with the common scheme studied by Gallager (pages 480 to 482, [33]). The only difference being that, in [33], the decoder calculates the unbiased estimate $\widehat{x}_{0,T}/(1 - a^{-2T-2})$ and then maps it to the closest message point to receover the transmitted message (as opposed to mapping $\widehat{x}_{0,T}$ directly to a message point); however the difference vanishes sufficiently fast and hence (22) and the common scheme have the same asymptotic behavior.

Moreover, this formulation is related to a control problem. To see this, consider the optimal control problem for (22), assuming a and c are given,

l is the controller gain (i.e., $-ly_{t-1}$ is the control input based on noisy observation y_{t-1}):

$$(25) \qquad \min_{l:(22) \text{ is stable}} \quad \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} \mathbf{E}(u_t)^2;$$

that is, we want to minimize the power of the regulated output u by appropriately choosing l (namely to minimize the power of the "error" signal $u := c\widetilde{x}_t$, or simply put, to make the state $\widetilde{x}_t$ small, through proper controller design). Since the objective function does not assign any direct penalty to the controller's output $-ly_t$, i.e., there is no need to reduce the control effort, this is a cheap control problem. See [34]; also see Appendix A.2 for the discussion of cheap control and closely related expensive control problems. It is straightforward to see that, the average channel input power $\mathbf{E}u^2$ is minimized if and only if the control gain l takes the value according to (6). For this reason, the dotted box in Fig. 3 is called a control setup.

Thus, we have seen that the SK scheme can be seen as a reformulation of a KF system or cheap control solution. The SK scheme is an optimal communication system; the KF system is an optimal estimation system; and the cheap control solution is an optimal control system. This draws connections among communication, estimation, and control. These observations can be generalized for Gaussian channels with memory, which we now turn to.

3. The channel model and CP scheme

In this section, we introduce the Gaussian channel with memory, which includes the AWGN channel studied in the previous section as a special case. The channel may be formulated as one with AWGN and inter-symbol interference (ISI), or one with colored Gaussian noise but without ISI. We then discuss the CP scheme over the channel with colored Gaussian noise. Finally we reformulate the CP scheme for the channel with AWGN and ISI and introduce a state-space representation of the scheme.

3.1. Channel models

Fig. 4 depicts the channel $\mathcal{F}$ with AWGN and ISI. In state-space the channel is given by

$$(26) \qquad \text{channel } \mathcal{F} : \begin{cases} \underline{s}_{t+1} = \boldsymbol{F}\underline{s}_t + \underline{G}u_t \\ y_t = \underline{H}'\underline{s}_t + u_t + w_t, \end{cases}$$

where $\underline{s}_0 = \underline{0}$, $\boldsymbol{F} \in \mathbb{R}^{m \times m}$, the system $(\boldsymbol{F}, \underline{G}, \underline{H}', 1)$ is stable and minimum-phase, and $w_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. Here m is called the order (or dimension) of the channel. The input-output representation of $\mathcal{F}$, in matrix form and for the time duration $t = 0, 1, \ldots, T$, can be written as

$$(27) \qquad \mathcal{F} : \quad \underline{y}^T = \boldsymbol{\mathcal{Z}}_T^{-1} \underline{u}^T + \underline{w}^T,$$

where $\boldsymbol{\mathcal{Z}}_T^{-1} \in \mathbb{R}^{(T+1) \times (T+1)}$ is a unique lower triangular Toeplitz matrix of the impulse response of $(\boldsymbol{F}, \underline{G}, \underline{H}', 1)$ with all-one diagonal elements.
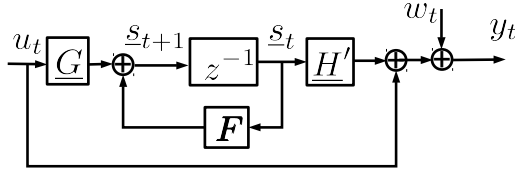

Figure 4:   The channel $\mathcal{F}$ in state space.

This channel $\mathcal{F}$ can be transformed to a colored Gaussian noise channel without ISI ([6, 22]), called $\mathcal{F}'$, defined as

$$(28) \quad \mathcal{F}' : \quad \widetilde{\underline{y}}^T := \boldsymbol{\mathcal{Z}}_T \underline{y}^T = \underline{u}^T + \boldsymbol{\mathcal{Z}}_T \underline{w}^T := \underline{u}^T + \underline{Z}^T, \quad \text{i.e.} \quad \widetilde{\underline{y}}^T = \underline{u}^T + \underline{Z}^T,$$

where $\underline{Z}^T := \boldsymbol{\mathcal{Z}}_T \underline{w}^T$ is the colored noise process generated by the FDLTI system $\boldsymbol{\mathcal{Z}}_T$ driven by AWGN $\underline{w}^T$. The channel $\mathcal{F}$ is said to be equivalent to the channel (28) in the sense that $\underline{y}^T$ and $\widetilde{\underline{y}}^T$ causally determine each other.

3.2. The CP scheme over $\mathcal{F}'$

Over the channel $\mathcal{F}'$, the CP scheme is obtained by letting

$$(29) \qquad \underline{u}^T := \boldsymbol{\mathcal{B}}_T \underline{Z}^T + \underline{v}^T,$$

where $\boldsymbol{\mathcal{B}}_T \in \mathbb{R}^{(T+1) \times (T+1)}$ is strictly lower triangular and $\underline{v}^T$ is Gaussian and independent of $\underline{Z}^T$. That is, the channel input $\underline{u}^T$ is generated based on the strictly causal feedback of the channel noise $\underline{Z}^T$ and an independent information-bearing signal $\underline{v}^T$. Based on the CP scheme, the concept of "finite-horizon information capacity" was introduced in [31] as the rate associated with the maximum mutual information between $\underline{v}^T$ and the corresponding channel output $\underline{y}^T$ for all admissible $\boldsymbol{\mathcal{B}}_T$ and $\boldsymbol{K}_{\underline{v}}^{(T)} := \mathbf{E}\underline{v}^T \underline{v}^{T\prime}$.

Formally, let $\underline{u}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$ and $\underline{y}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$ be an input sequence and an output sequence generated using $(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$, and $I(\underline{v}_T; \underline{y}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T))$ be the mutual information between $\underline{v}^T$ and $\underline{y}^T$.

Definition 1. The *finite-horizon information capacity* $C_T(\mathcal{P})$ for any power budget $\mathcal{P} > 0$ is defined as

$$(30) \qquad C_T(\mathcal{P}) := \sup R_T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T) := \sup \frac{1}{T+1} I(\underline{v}_T; \underline{y}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T))$$

subject to the power constraint

$$(31) \qquad P_T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T) := \frac{1}{T+1} \mathbf{E}\|\underline{u}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)\|^2 \leq \mathcal{P}.$$

One can also define the inverse function of $C_T(\mathcal{P})$ as $P_T(\mathcal{R})$, the infimum power for a given rate constraint:

$$(32) \qquad P_T(\mathcal{R}) = \inf_{s.t. \ R_T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T) \geq \mathcal{R}} P_T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T).$$

The finite-horizon C_T is different from the capacity notion defined for infinite horizon in the sense of Shannon, but in [31] a channel coding theorem was proven for the case of Gaussian channels based on the sequence of C_T. Furthermore, if the Gaussian noise process $\{Z_t\}$ is stationary, then C_T converges to the feedback capacity in the sense of Shannon [18]. Therefore, it is meaningful to study C_T. In addition, the results in [31] also imply that any mutual information rate is achievable by appropriate codes, and hence it is suffcent to study mutual information rate for the Gaussian channels.

3.3. The CP scheme over $\mathcal{F}$

We now introduce the CP scheme for $\mathcal{F}$ based on the CP scheme for $\mathcal{F}'$. Let

$$(33) \qquad \underline{r}^T := (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1} \underline{v}^T;$$

i.e., $\underline{r}^T$ and $\underline{v}^T$ can causally determine each other for any given $\boldsymbol{\mathcal{B}}_T$ and hence are in one-to-one correspondence with each other. However, $\underline{r}^T$ turns out to bear some significance since the mutual information between $\underline{r}^T$ and $\underline{y}^T$ across the channel is determined *only* by $\underline{r}^T$ (in particular, the mutual

information or its rate is independent of $\boldsymbol{\mathcal{B}}_T$, a property not held for $\underline{v}^T$); see Proposition 2 and Remark 1. By $\underline{Z}^T = \boldsymbol{\mathcal{Z}}_T \underline{w}^T$ and $\widetilde{\underline{y}}^T = \boldsymbol{\mathcal{Z}}_T \underline{y}^T$, we have

$$
(34) \qquad \begin{aligned}
\underline{u}^T &= \boldsymbol{\mathcal{B}}_T \boldsymbol{\mathcal{Z}}_T \underline{w}^T + (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)\underline{r}^T \\
\underline{y}^T &= \boldsymbol{\mathcal{Z}}_T^{-1}(\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)(\boldsymbol{\mathcal{Z}}_T \underline{w}^T + \underline{r}^T).
\end{aligned}
$$

This further follows that, the channel input $\underline{u}^T$ can be represented as

$$
(35) \qquad \underline{u}^T = (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\boldsymbol{\mathcal{B}}_T \boldsymbol{\mathcal{Z}}_T \underline{y}^T + \underline{r}^T = \boldsymbol{\mathcal{B}}_T(\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\boldsymbol{\mathcal{Z}}_T \underline{y}^T + \underline{r}^T,
$$

which we call the CP scheme over $\mathcal{F}$ and is illustrated in Fig. 5 (a). We point out that in this formulation of the CP scheme, the channel input $\underline{u}^T$ is generated from the strictly causal feedback of the channel output $\underline{y}^T$ (instead of the channel noise $\underline{Z}^T$ in (29)) and message-bearing signal $\underline{r}^T$ (instead of $\underline{v}^T$ in (29)).

3.4. State-space representation of the CP scheme over $\mathcal{F}$

A state-space representation of the CP scheme over $\mathcal{F}$, denoted $\mathbb{S} := \mathbb{S}(\mathcal{F})$, can be obtained as shown in Fig. 5 (b), and its dynamics is shown in (36). The corresponding relation between $\mathbb{S}$ and the CP scheme over $\mathcal{F}$ will be rigorously established in Lemma 1.

$$
(36) \quad \begin{array}{l} \text{Coding} \\ \text{scheme } \mathbb{S}: \end{array} \left\{ \begin{aligned}
\underline{x}_{t+1} &= \boldsymbol{A}\underline{x}_t \\
r_t &= \underline{C}'\underline{x}_t \\
u_t &= r_t - \widehat{r}_t \\
\underline{s}_{t+1} &= \boldsymbol{F}\underline{s}_t + \underline{G}u_t \\
y_t &= \underline{H}'\underline{s}_t + u_t + w_t \\
-\widehat{\underline{r}}^T &= \boldsymbol{\mathcal{G}}_T \underline{y}^T
\end{aligned} \right. \quad \begin{array}{l} \left.\vphantom{\begin{aligned}\underline{x}_{t+1}\\r_t\\u_t\end{aligned}}\right\} \text{encoder} \\[2ex] \left.\vphantom{\begin{aligned}\underline{s}_{t+1}\\y_t\end{aligned}}\right\} \text{channel } \mathcal{F} \\[1ex] \text{feedback generator} \end{array}
$$

where $\underline{x}_0 := \underline{M} \sim \mathcal{N}(\underline{0}, \boldsymbol{I}_{n+1})$, $0 \leq n \leq T$, $(n+1)$ is called the encoder dimension or order, and $\boldsymbol{\mathcal{G}}_T$ will be characterized in state-space in the next section. Note that the initial condition $\underline{x}_0$ is assumed to be Gaussian since here we focus on the study of mutual information, which is maximized if $\underline{x}_0$ is Gaussian.

Let $\mathbb{T}_{T,n} := (\boldsymbol{A} \in \mathbb{R}^{(n+1)\times(n+1)}, \underline{C} \in \mathbb{R}^{n+1}, \boldsymbol{\mathcal{G}}_T \in \mathbb{R}^{(T+1)\times(T+1)})$. The triple $\mathbb{T}_{T,n}$ is said to be admissible if the following Assumption (**A1**) holds:

(**A1**) $(\boldsymbol{A}, \underline{C}')$ is observable, $\boldsymbol{\mathcal{G}}_T$ is strictly lower triangular, and $0 \leq n \leq T$.
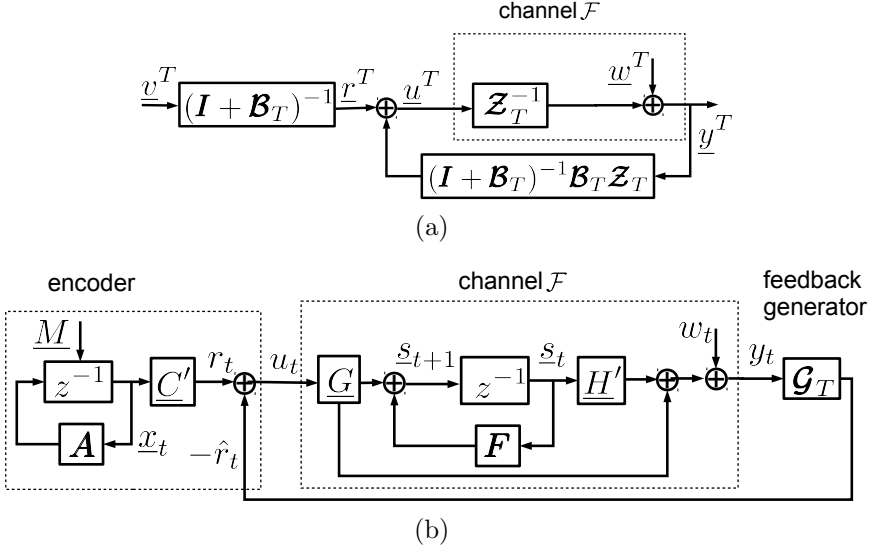
(a)

(b)

Figure 5: (a) The CP scheme over the channel $\mathcal{F}$. (b) A state-space representation of (a), called the coding scheme $\mathbb{S}$ over $\mathcal{F}$. The feedback generator is a portion of the decoder.

Throughout the paper, Assumption (**A1**) is assumed unless otherwise specified. Note that the feedback generator's operation at time t is fully captured in the $(t+1)$st row of the matrix $\boldsymbol{\mathcal{G}}_T$. Then the fact that $\boldsymbol{\mathcal{G}}_T$ is strictly lower triangular implies that the feedback generator is strictly causal. Assumption (**A1**) allows the feedback generator $\boldsymbol{\mathcal{G}}_T$ to be either time invariant (if $\boldsymbol{\mathcal{G}}_T$ is Toeplitz) or time varying (if $\boldsymbol{\mathcal{G}}_T$ is not Toeplitz).

Let $\underline{u}^T(\mathbb{T}_{T,n})$ and $\underline{y}^T(\mathbb{T}_{T,n})$ be any input and output sequences generated with the triple $\mathbb{T}_{T,n}$ and so on (the dependency on $\mathbb{T}_{T,n}$ may be dropped for notational convenience), and

(37)
$$R_{T,n}(\mathbb{T}_{T,n}) := \frac{1}{T+1} I(\underline{M}; \underline{y}^T(\mathbb{T}_{T,n})),$$
$$P_{T,n}(\mathbb{T}_{T,n}) := \frac{1}{T+1} \mathbf{E}\|\underline{u}^T(\mathbb{T}_{T,n})\|^2.$$

The system $\mathbb{S}$ with a triple $\mathbb{T}_{T,n}$ generates

(38)
$$\underline{u}^T(\mathbb{T}_{T,n}) = -\underline{\widehat{r}}^T + \underline{r}^T = \boldsymbol{\mathcal{G}}_T \underline{y}^T + \boldsymbol{\Gamma}_T \underline{M},$$

where we have defined

$$(39) \qquad\qquad \underline{r}^T := \mathbf{\Gamma}_T \underline{M}$$

and

$$(40) \qquad\qquad \mathbf{\Gamma}_T(\mathbb{T}_{T,n}) := [\underline{C}, \mathbf{A}'\underline{C}, \ldots, \mathbf{A}^{T'}\underline{C}]'$$

is the observability matrix with full rank by Assumption (**A1**) (cf. [35]). Therefore, the covariance matrix of $\underline{r}^T$, i.e.,

$$(41) \qquad\qquad \boldsymbol{K}_{\underline{r}}^{(T)} := \mathbf{E}\underline{r}^T \underline{r}^{T\prime} = \mathbf{\Gamma}_T \mathbf{\Gamma}_T{}',$$

is of full rank.

In the case of $n = T$, the strictly lower triangular matrix $\boldsymbol{\mathcal{G}}_T$ can be chosen such that it can generate any strictly causal feedback from $\underline{y}^T$, and $\underline{r}^T$ can be or approach any given covariance matrix. Thus, comparing the above $\underline{u}^T(\mathbb{T}_{T,n})$ generated by the system $\mathbb{S}$ with (35), one can conclude that any input that the CP scheme can generate can also be generated or approached by the system $\mathbb{S}$, and hence C_T can be attained or approached by the system $\mathbb{S}$.

Definition 2. Consider $\mathbb{S}$ in (36). Define the constrained capacity

$$(42) \qquad C_{T,n} := C_{T,n}(\mathcal{P}) := \sup_{P_{T,n}(\mathbb{T}_{T,n}) \le \mathcal{P}, (A1)} R_{T,n}(\mathbb{T}_{T,n})$$

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

$$(43) \qquad P_{T,n} := P_{T,n}(\mathcal{R}) := \inf_{R_{T,n}(\mathbb{T}_{T,n}) \ge \mathcal{R}, (A1)} P_{T,n}(\mathbb{T}_{T,n}).$$

In other words, $C_{T,n}$ is the finite-horizon information capacity over all admissible $\mathbb{T}_{T,n}$ with a *fixed encoder dimension* $(n + 1)$. The following lemma connects $C_{T,T}$ and C_T based on the relation between $\mathbb{S}$ and the CP scheme over either $\mathcal{F}'$ or $\mathcal{F}$.

Lemma 1. *i) For any given triple $\mathbb{T}_{T,n}$, there is $(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$ such that $\underline{u}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T) = \underline{u}^T(\mathbb{T}_{T,n})$;*
ii) For any given pair $(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$ with $\boldsymbol{K}_{\underline{v}}^{(T)} > 0$, there exists an admissible triple $\mathbb{T}_{T,T}$ such that $\underline{u}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T) = \underline{u}^T(\mathbb{T}_{T,T})$; for any given pair

$(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$ *with* $\boldsymbol{K}_{\underline{v}}^{(T)} \geq 0$ *but* $\boldsymbol{K}_{\underline{v}}^{(T)} \not\succ 0$, *there exists a sequence of admissible triples* $\{\mathbb{T}_{T,T,i}\}_{i=1}^{\infty}$ *such that* $\underline{u}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T) = \lim_{i \to \infty} \underline{u}^T(\mathbb{T}_{T,T,i})$;
    iii) It holds that $C_T(\mathcal{P}) = C_{T,T}(\mathcal{P})$ *and* $P_T(\mathcal{R}) = P_{T,T}(\mathcal{R})$.

Proof. See Appendix A.3. □

Thus, to achieve C_T, it is sufficient to study either the CP scheme with all possible $(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$ or the system $\mathbb{S}$ with all admissible $\mathbb{T}_{T,T}$. Since the set formed by all admissible $\mathbb{T}_{T,T}$ is a subset of the set formed by all admissible $\mathbb{T}_{T,n}$ with $n \leq T$, we can study the system $\mathbb{S}$ with all admissible $\mathbb{T}_{T,n}$ with $n \leq T$ with no loss of generality or optimality. However, allowing n to be not necessarily equal to T gives us the flexibility of increasing T (i.e. time horizon) for a fixed n (i.e. encoder dimension), a useful step towards the KF characterization of the feedback communication problem. In what follows, several refinements to the coding scheme $\mathbb{S}$ with $\mathbb{T}_{T,n}$ will be presented.

4. Necessity of KF for optimal communication

In this section, we refine the CP scheme by showing that the scheme $\mathbb{S}$ necessarily implements the KF algorithm in order to minimize the channel input power (i.e., to achieve power efficiency) and to generate an optimal estimate of the transmitted message (e.g., in the sense of MMSE estimation). Therefore, the optimal feedback communication system over the channel $\mathcal{F}$ (optimal in the sense of capacity achieving) has to contain a KF, either explicitly or implicitly.

4.1. The KF for channel input power minimization

We first compute the mutual information in the aforementioned coding scheme $\mathbb{S}$.

Proposition 2. *Let* $(\boldsymbol{A}, \underline{C}, \boldsymbol{\mathcal{G}}_T)$ *be an admissible triple for* $\mathbb{S}$ *in (36). Then it holds that*

$$(44) \qquad I(\underline{M}; \underline{y}^T) = I(\underline{r}^T; \underline{y}^T) = \frac{1}{2} \log |\boldsymbol{K}_{\underline{y}}^{(T)}|$$
$$= \frac{1}{2} \log |\boldsymbol{I} + \boldsymbol{\mathcal{Z}}_T^{-1} \boldsymbol{K}_{\underline{r}}^{(T)}(\boldsymbol{A}, \underline{C}) \boldsymbol{\mathcal{Z}}_T^{-1'}|,$$

independent of $\boldsymbol{\mathcal{G}}_T$.

Proof. From $\underline{y}^T = \boldsymbol{Z}_T^{-1}\underline{u}^T + \underline{w}^T$ and $\underline{u}^T = \underline{r}^T - \boldsymbol{G}_T\underline{y}^T$, one obtains

$$(45) \qquad \underline{y}^T = (\boldsymbol{I} - \boldsymbol{Z}_T^{-1}\boldsymbol{G}_T)^{-1}(\boldsymbol{Z}_T^{-1}\underline{r}^T + \underline{w}^T).$$

Therefore, we have

$$(46) \qquad \begin{aligned} I(\underline{M};\underline{y}^T) &= h(\underline{y}^T) - h(\underline{y}^T|\underline{M}) \\ &= h(\underline{y}^T) - h\left((\boldsymbol{I} - \boldsymbol{Z}_T^{-1}\boldsymbol{G}_T)^{-1}(\boldsymbol{Z}_T^{-1}\underline{r}^T + \underline{w}^T)|\underline{M}\right) \\ &\overset{(a)}{=} \frac{1}{2}\log|2\pi e\boldsymbol{K}_{\underline{y}}^{(T)}| - h(\underline{w}^T) \\ &= \frac{1}{2}\log|\boldsymbol{K}_{\underline{y}}^{(T)}| \overset{(b)}{=} \frac{1}{2}\log|\boldsymbol{I} + \boldsymbol{Z}_T^{-1}\boldsymbol{K}_{\underline{r}}^{(T)}\boldsymbol{Z}_T^{-1'}|, \end{aligned}$$

where (a) is due to $\underline{r}^T = \boldsymbol{\Gamma}_T\underline{M}$ and $|\boldsymbol{I} - \boldsymbol{Z}_T^{-1}\boldsymbol{G}_T|^{-1} = 1$, (b) is due to $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$. Thus, $I(\underline{M};\underline{y}^T)$ is independent of the feedback generator $\boldsymbol{G}_T$ and depends only on $(\boldsymbol{A}, \underline{C})$ or equivalently $\boldsymbol{K}_{\underline{r}}^{(T)}(\boldsymbol{A}, \underline{C})$. One can also easily seen that

$$(47) \qquad I(\underline{r}^T; \underline{y}^T) = \frac{1}{2}\log|\boldsymbol{K}_{\underline{y}}^{(T)}|$$

following similar derivations. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Remark 1. The proposition implies that, for the given channel $\mathcal{F}$, *a fixed* $(\boldsymbol{A}, \underline{C})$ *or fixed* $\boldsymbol{K}_{\underline{r}}^{(T)}(\boldsymbol{A}, \underline{C})$ *leads to a fixed mutual information regardless of the feedback generator.* Thus, the information rate $R_{T,n}(\boldsymbol{A}, \underline{C}, \boldsymbol{G}_T)$ can be written as $R_{T,n}(\boldsymbol{A}, \underline{C})$ since it does not depend on $\boldsymbol{G}_T$, in other words the mutual information (and information rate) is invariant under any causal feedback, and is equal to the mutual information (and information rate, resp.) in the system without feedback. Consequently, these quantities can be computed offline before running the closed-loop system, and hence may be interpreted as anti-causal quantities.

A further implication is that *the feedback generator $\boldsymbol{G}_T$ has to be chosen to minimize the average channel input power* in order to achieve the capacity (recalling that the capacity can be expressed as the minimal power for any giving rate (43)). The infinite-horizon counterpart of this result was proven in [4].

Next we solve the optimal feedback generator for a fixed $(\boldsymbol{A}, \underline{C})$ in finite horizon. Denote the *optimal feedback generator* for a given $(\boldsymbol{A}, \underline{C})$ as

$\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$, namely

$$(48) \quad \boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C}) := \arg\inf_{\boldsymbol{\mathcal{G}}_T} P_{T,n}(\boldsymbol{A}, \underline{C}, \boldsymbol{\mathcal{G}}_T) = \arg\inf_{\boldsymbol{\mathcal{G}}_T} \mathbf{E}\|\underline{r}^T + \boldsymbol{\mathcal{G}}_T(\boldsymbol{A}, \underline{C})\underline{y}^T\|^2;$$

that is, the optimal feedback generator $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$ achieves power efficiency. This problem can be solved by studying the estimation system shown in Fig. 6 (a), which contains an estimator $\widehat{\boldsymbol{\mathcal{G}}}_T$ mapping the channel output $\bar{y}^T$ into $\widehat{\underline{r}}^T$, an estimate of $\underline{r}^T$, i.e., $\widehat{\underline{r}}^T = \widehat{\boldsymbol{\mathcal{G}}}_T \bar{y}^T$. As one can see, similar to the case studied in Section 2, a communication system with feedback can be constructed from an estimation system, and *the minimization of channel input power in the communication system can be obtained from the minimization of MSE in the estimation problem*, which necessitates a KF. In other words, the observation regarding converting a KF over an AWGN channel to a communication system (cf. Fig. 1) can be generalized to more general Gaussian channels. More specifically, from the estimation system Fig. 6 (a), one can subtract the estimate $\widehat{r}_t$ from the channel input and add back $(\boldsymbol{\mathcal{Z}}_T^{-1}\widehat{r}_t)$ to the channel output, which leads to the communication system with the feedback generator $\boldsymbol{\mathcal{G}}_T$ shown in Fig. 6 (b) without affecting the input, state, and output of the estimator $\widehat{\boldsymbol{\mathcal{G}}}_T$. If the estimate $\widehat{r}_t$ minimizes the MSE, then the power of the channel input $u_t := r_t - \widehat{r}_t$ is minimized in the communication system, i.e., the optimal feedback generator achieves power efficiency.

Proposition 3. *Consider the coding scheme $\mathbb{S}$ in Fig. 5 (b). Then it holds that i) (recall the capacity definition $P_{T,n}(\mathcal{R})$ in (43))*

$$(49) \qquad P_{T,n}(\mathcal{R}) = \inf_{R_{T,n}(\boldsymbol{A}, \underline{C}) \geq \mathcal{R}} P_{T,n}(\boldsymbol{A}, \underline{C}, \boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C}))$$

and ii) the optimal feedback generator $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$ is given by

$$(50) \qquad \boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C}) = -\widehat{\boldsymbol{\mathcal{G}}}_T^*(\boldsymbol{A}, \underline{C})(\boldsymbol{I} - \boldsymbol{\mathcal{Z}}_T^{-1}\widehat{\boldsymbol{\mathcal{G}}}_T^*(\boldsymbol{A}, \underline{C}))^{-1},$$

where $\widehat{\boldsymbol{\mathcal{G}}}_T^*(\boldsymbol{A}, \underline{C})$ is the one-step prediction MMSE estimator (Kalman filter) of $\underline{r}^T$ based on the noisy observation $\bar{y}^T := \boldsymbol{\mathcal{Z}}_T^{-1}\underline{r}^T + \underline{w}^T$ (i.e. the optimal one-step prediction is $\widehat{\underline{r}}^T = \widehat{\boldsymbol{\mathcal{G}}}_T^*(\boldsymbol{A}, \underline{C})\bar{y}^T$), given by

$$(51) \qquad \widehat{\boldsymbol{\mathcal{G}}}_T^*(\boldsymbol{A}, \underline{C}) \quad := \quad \arg\min_{\widehat{\boldsymbol{\mathcal{G}}}_T} \frac{1}{T+1}\mathbf{E}\|\underline{r}^T - \widehat{\boldsymbol{\mathcal{G}}}_T \bar{\underline{y}}^T\|^2,$$

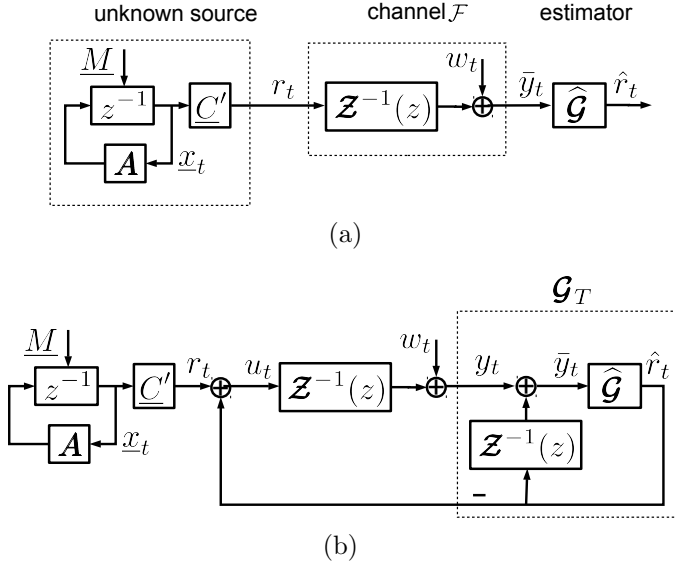*where $\widehat{\boldsymbol{\mathcal{G}}}_T$ is strictly lower triangular.*

(a)

(b)

Figure 6: Transforming an estimation system (a) to a communication system (b).

Fig. 7 (a) shows the KF $\widehat{\boldsymbol{\mathcal{G}}}_T^*(\boldsymbol{A}, \underline{C})$, and (b) the state-space representation of the optimal feedback generator $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$ (the parameters $\underline{L}_{1,t}$ and $\underline{L}_{2,t}$ will be characterized in (64)).

Remark 2. Proposition 3 formally states that, for any fixed $(\boldsymbol{A}, \underline{C})$, the minimization of the channel input power in a feedback communication problem is equivalent to the minimization of MSE in an estimation problem. Thus, *one essential role of the feedback generator* $\boldsymbol{\mathcal{G}}$ *for any fixed* $(\boldsymbol{A}, \underline{C})$ *is to minimize the input power*, and the optimal feedback generator $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$ is given by the KF.

Proof. i) By Proposition 2, for any fixed $(\boldsymbol{A}, \underline{C})$, $R_{T,n}(\boldsymbol{A}, \underline{C})$ is fixed. Then from the definition of $P_{T,n}(\mathcal{R})$, we have

$$(52) \qquad P_{T,n}(\mathcal{R}) = \inf_{\substack{\boldsymbol{A}, \underline{C}, \boldsymbol{\mathcal{G}}_T \\ s.t.\ R_{T,n}(\boldsymbol{A}, \underline{C}) \geq \mathcal{R}}} P_{T,n}(\boldsymbol{A}, \underline{C}, \boldsymbol{\mathcal{G}}_T)$$

$$= \inf_{\substack{\boldsymbol{A}, \underline{C} \\ s.t.\ R_{T,n}(\boldsymbol{A}, \underline{C}) \geq \mathcal{R}}} \inf_{(\boldsymbol{A}, \underline{C})\ \text{fixed}, \boldsymbol{\mathcal{G}}_T} P_{T,n}(\boldsymbol{A}, \underline{C}, \boldsymbol{\mathcal{G}}_T).$$

Kalman filter $\widehat{\boldsymbol{\mathcal{G}}}_T^*$

(a)

$$\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C}')$$

(b)
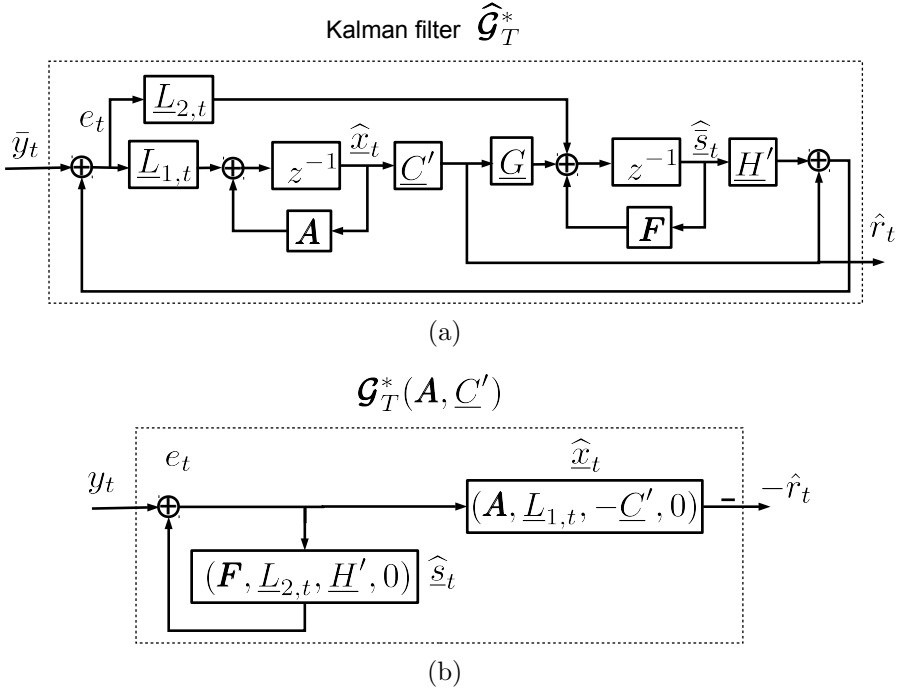
Figure 7: (a) The KF $\widehat{\boldsymbol{\mathcal{G}}}_T^*(\boldsymbol{A}, \underline{C})$. (b) The KF-based feedback generator $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$ in state space. $(\boldsymbol{A}, \underline{L}_{1,t}, -\underline{C}, 0)$ with $\widehat{\underline{x}}_t$ denotes a state-space representation with $\widehat{\underline{x}}_t$ being its state at time t, and initial condition $\widehat{\underline{x}}_0$ being 0.

Then i) follows from the definition of $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$. For ii), note that for the coding scheme $\mathbb{S}$, it holds that

(53) $$\underline{u}^T = \underline{r}^T + (-\widehat{\underline{r}}^T) = \underline{r}^T + \boldsymbol{\mathcal{G}}_T \underline{y}^T.$$

Then, letting

(54) $$\widehat{\boldsymbol{\mathcal{G}}}_T := -\boldsymbol{\mathcal{G}}_T (\boldsymbol{I} - \boldsymbol{\mathcal{Z}}_T^{-1} \boldsymbol{\mathcal{G}}_T)^{-1}$$

and $\underline{\bar{y}}^T := \boldsymbol{\mathcal{Z}}_T^{-1} \underline{r}^T + \underline{w}^T$, we have $\boldsymbol{\mathcal{G}}_T \underline{y}^T = -\widehat{\boldsymbol{\mathcal{G}}}_T \underline{\bar{y}}^T$. Therefore,

(55) $$\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C}) = \arg\inf_{\boldsymbol{\mathcal{G}}_T} \frac{1}{T+1} \mathbf{E} \|\underline{r}^T + \boldsymbol{\mathcal{G}}_T \underline{y}^T\|^2$$

$$= \arg\inf_{\widehat{\boldsymbol{\mathcal{G}}}_T} \frac{1}{T+1} \mathbf{E} \|\underline{r}^T - \widehat{\boldsymbol{\mathcal{G}}}_T \underline{\bar{y}}^T\|^2.$$

The last equality implies that the optimal solution $\widehat{\boldsymbol{\mathcal{G}}}_T^*$ is the strictly causal MMSE estimator (with one-step prediction) of $\underline{r}^T$ given $\bar{\underline{y}}_T$ which can be implemented recursively in state-space as a KF ([29]). Finally, from the relation between $\boldsymbol{\mathcal{G}}_T$ and $\widehat{\boldsymbol{\mathcal{G}}}_T$, we obtain (50). The state-space representation of $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$, as illustrated in Fig. 7 (c), can be obtained from straightforward computation, as shown in Appendix A.1.   □

4.2. The presence of the KF algorithm for optimal message recovery

Given the transmitted message $\underline{x}_0 := \underline{M}$, denote the decoder-side estimate of the transmitted message at time t as $\widehat{\underline{x}}_{0,t}$, and denote $\widehat{\underline{M}}_T := \widehat{\underline{x}}_{0,T}$. The estimate is desired to satisfy: 1) no loss of information during the recovery process, i.e. $I(\underline{M}; \widehat{\underline{M}}_T) = I(\underline{M}; \underline{y}^t)$, and 2) minimization of the error $\widetilde{\underline{M}} := \underline{M} - \widehat{\underline{M}}_T$ in some sense, e.g., in terms of MMSE or minimum error entropy (MEE, see e.g. [36]). It is known that to recovery $\underline{M}$, a (an anti-causal) fixed-point smoothing filter can be used. The smoothing filter, typically based on the KF innovations, is given by (see Problem 10.1 in [29])

$$(56) \qquad \widehat{\underline{x}}_{0,t} := \mathbf{E}(\underline{M}|\underline{y}^t) = \widehat{\underline{x}}_{0,t-1} + \boldsymbol{\phi}'(t)\underline{C}K_{e,t}^{-1}e_t,$$

where $\boldsymbol{\phi}(t) := \boldsymbol{A}_{cl}(t-1)\boldsymbol{A}_{cl}(t-2)\cdots\boldsymbol{A}_{cl}(0)$ if $t > 0$ and $\boldsymbol{\phi}(0) := \boldsymbol{I}$, $\boldsymbol{A}_{cl}(t) := \boldsymbol{A} - \underline{L}_{1,t}\underline{C}'$, e_t is the KF innovation (see Fig. 7 and Section 5.1). Since this estimate is the conditional-mean estimate and the system is linear with Gaussian random variables, straightforward computation shows it satisfies the above-mentioned optimal criteria 1) and 2). Hence, the KF algorithm can provide the optimal estimate of the transmitted message.

4.3. The complete characterization of the roles of KF algorithm in feedback communication

To summarize, the KF for a process generated by an unstable linear system driven by its unknown initial condition and observed through a Gaussian channel with memory (e.g. $\underline{r}^T$ in Fig. 6 (a)), when reformulated in an appropriate form, is optimal in the sense of power efficiency in information transmission and information recovery. Searching over all admissible unknown processes $\underline{r}^T$ can achieve the capacity $C_T(\mathcal{P})$. The *power efficiency* in communication, i.e. the minimization of the channel input power for a fixed information rate, is guaranteed by the strictly causal one-step prediction operation in Kalman filtering (i.e. the operation to recursively

generate $\mathbf{E}(r_t|\underline{y}^{t-1})$ at time t); and the *optimal recovery of the transmitted message* (optimal in the sense of MMSE or MEE between the message and the estimate of the message) is guaranteed by the anti-causal smoothing operation in Kalman filtering (i.e. the operation to generate $\mathbf{E}(\underline{M}|\underline{y}^T)$). We may view this characterization as the optimality of KF in the sense of *information transmission* with feedback, which is a complement to the existing characterization that KF is optimal in the sense of *information processing* established by Mitter and Newton in [12].

5. Connections among feedback communication, estimation, and feedback control

We have shown that the coding scheme $\mathbb{S}$ necessarily contains a KF-based feedback generator. The KF immediately links the feedback communication problem to estimation and control problems. In this section, we present a *unified representation* of the coding scheme $\mathbb{S}^*$ (defined as $\mathbb{S}$ with the optimal $\boldsymbol{\mathcal{G}}^*(\boldsymbol{A}, \underline{C})$ replacing $\boldsymbol{\mathcal{G}}$), its estimation theory counterpart, and its control theory counterpart.

5.1. Unified representation of feedback coding system, KF, and cheap control

5.1.1. Coding scheme $\mathbb{S}^*$. The coding scheme $\mathbb{S}^*$ which contains the optimal feedback generator $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$ can be described as

$$
(57) \quad \begin{array}{c} \text{coding} \\ \text{scheme } \mathbb{S}^* \end{array} \left\{ \begin{array}{l} \underline{x}_{t+1} = \boldsymbol{A}\underline{x}_t \\ r_t = \underline{C}'\,\underline{x}_t \\ u_t = r_t - \widehat{r}_t \\ \underline{s}_{t+1} = \boldsymbol{F}\underline{s}_t + \underline{G}u_t \\ y_t = \underline{H}'\,\underline{s}_t + u_t + w_t \\ \widehat{\underline{s}}_{t+1} = \boldsymbol{F}\widehat{\underline{s}}_t + \underline{L}_{2,t}e_t \\ e_t = y_t - \underline{H}'\,\widehat{\underline{s}}_t \\ \widehat{\underline{x}}_{t+1} = \boldsymbol{A}\widehat{\underline{x}}_t + \underline{L}_{1,t}e_t \\ -\widehat{r}_t = -\underline{C}'\,\widehat{\underline{x}}_t \end{array} \right\} \begin{array}{l} \text{encoder} \\ \\ \\ \text{channel } \mathcal{F} \\ \\ \text{optimal feedback} \\ \text{generator } \boldsymbol{\mathcal{G}}^*(\boldsymbol{A}, \underline{C}) \end{array}
$$

with $\underline{x}_0 = \underline{M}$ known to the encoder only, $\underline{s}_0 = \widehat{\underline{s}}_0 = \underline{0}$, and $\widehat{\underline{x}}_0 = \underline{0}$. Here $\underline{L}_{1,t}$ and $\underline{L}_{2,t}$ are the time-varying optimal gains of the optimal feedback generator (which is readily solved by studying the innovations representation of the associated estimation problem; refer to Section 5.1.3). See Appendix A.1 for the derivation of a state-space representation of $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$. One can see that

this is an extension of the KF-based coding scheme over an AWGN channel shown in (11) [4]. The encoding and decoding procedures, omitted here for brevity, are also extensions of the AWGN ones as described in [4] (which also showed that any mutual information rate is achievable using the encoding and decoding procedures).

5.1.2. The estimation system. Fig. 6 (a) shows the estimation system, in which $(\boldsymbol{A}, \underline{C})$ is fixed and known to the estimator. (Note the randomness in $\underline{r}^T$ comes from the initial condition for generating $\underline{r}^T$.) The system is described as

$$(58) \quad \begin{matrix} \text{estimation} \\ \text{system:} \end{matrix} \quad \begin{cases} \underline{x}_{t+1} = \boldsymbol{A}\underline{x}_t \\ r_t = \underline{C}' \, \underline{x}_t \\ \underline{\bar{s}}_{t+1} = \boldsymbol{F}\underline{\bar{s}}_t + \underline{G}r_t \\ \underline{\bar{y}}_t = \underline{H}'\underline{\bar{s}}_t + r_t + w_t \\ \underline{\hat{x}}_{t+1} = \boldsymbol{A}\underline{\hat{x}}_t + \underline{L}_{1,t}e_t \\ \underline{\hat{r}}_t = \underline{C}' \, \underline{\hat{x}}_t \\ \underline{\hat{\bar{s}}}_{t+1} = \boldsymbol{F}\underline{\hat{\bar{s}}}_t + \underline{G}\widehat{r}_t + \underline{L}_{2,t}e_t \\ e_t = \underline{\bar{y}}_t - \underline{H}' \underline{\hat{\bar{s}}}_t - \widehat{r}_t \end{cases} \quad \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{unknown source} \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{channel } \mathcal{F} \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} \begin{matrix} \text{Kalman filter} \\ \widehat{\boldsymbol{\mathcal{G}}}^*(\boldsymbol{A}, \underline{C}) \end{matrix} \end{matrix}$$

with $\underline{x}_0 = \underline{M}$, $\underline{\bar{s}}_0 = \underline{\hat{\bar{s}}}_0 = \underline{0}$, and $\underline{\hat{x}}_0 = \underline{0}$. To write this in a more compact form, define

$$(59) \quad \overline{\overline{\mathbb{X}}}_t := \begin{bmatrix} x_t \\ \bar{s}_t \end{bmatrix}, \quad \widehat{\overline{\mathbb{X}}}_t := \begin{bmatrix} \hat{x}_t \\ \hat{\bar{s}}_t \end{bmatrix}, \quad \mathbb{A} := \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{0} \\ \hline \underline{G}\underline{C}' & \boldsymbol{F} \end{array} \right]$$

$$\mathbb{C} := \left[\frac{\underline{C}}{\underline{H}} \right], \quad \underline{L}_t := \begin{bmatrix} \underline{L}_{1,t} \\ \underline{L}_{2,t} \end{bmatrix}.$$

Then we have

$$(60) \quad \begin{matrix} \text{estimation} \\ \text{system:} \end{matrix} \quad \begin{cases} \overline{\overline{\mathbb{X}}}_{t+1} = \mathbb{A}\overline{\overline{\mathbb{X}}}_t \\ \underline{\bar{y}}_t = \mathbb{C}'\overline{\overline{\mathbb{X}}}_t + w_t \\ \widehat{\overline{\mathbb{X}}}_{t+1} = \mathbb{A}\widehat{\overline{\mathbb{X}}}_t + \underline{L}_t e_t \\ e_t = \underline{\bar{y}}_t - \mathbb{C}'\widehat{\overline{\mathbb{X}}}_t. \end{cases} \quad \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \begin{matrix} \text{unknown source} \\ \text{and channel } \mathcal{F} \end{matrix} \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} \begin{matrix} \text{Kalman filter} \\ \widehat{\boldsymbol{\mathcal{G}}}^*(\boldsymbol{A}, \underline{C}) \end{matrix} \end{matrix}$$

with $\overline{\overline{\mathbb{X}}}_0 = [\underline{M}', \underline{0}']'$ and $\widehat{\overline{\mathbb{X}}}_0 = \underline{0}$. This is clearly an extension of the 1st-order KF system shown in (2).

---

[4]Note that (57) is time-varying whereas (11) is time-invariant. Therefore, to be more rigorous, (57) is an extension of the time-varying counterpart of (11), and the steady-state of (57) is an extension of (11). See Section 7.1 for the existence of the steady-state of (57).

It can be easily shown that r_t, $\widehat{r}_t$, e_t, $\underline{x}_t$, and $\widehat{\underline{x}}_t$ in (58) and (57) are equal, respectively, and it holds for any t that $\underline{s}_t - \widehat{\underline{s}}_t = \bar{\underline{s}}_t - \widehat{\bar{\underline{s}}}_t$, which leads to the following unified representation as a control system.

5.1.3. The unified representation: A cheap control problem. Define

$$
\widetilde{\underline{s}}_t := \underline{s}_t - \widehat{\underline{s}}_t = \bar{\underline{s}}_t - \widehat{\bar{\underline{s}}}_t, \quad \widetilde{\underline{x}}_t := \underline{x}_t - \widehat{\underline{x}}_t,
$$

(61)

$$
\mathbb{X}_t := \begin{bmatrix} \widetilde{\underline{x}}_t \\ \widetilde{\underline{s}}_t \end{bmatrix} = \bar{\mathbb{X}}_t - \widehat{\mathbb{X}}_t, \quad \mathbb{X}_0 := \begin{bmatrix} M \\ 0 \end{bmatrix}, \quad \mathbb{D} := \begin{bmatrix} C \\ 0 \end{bmatrix}.
$$

Note that $\mathbb{X}_t$ is the estimation error for $\bar{\mathbb{X}}_t$. Substituting (61) into (58) and (57), we obtain that both systems become

(62) $\quad$ control system: $\quad \begin{cases} \mathbb{X}_{t+1} = (\mathbf{A} - \underline{L}_t\mathbb{C}')\mathbb{X}_t - \underline{L}_t w_t & \text{(state evolution)} \\ \qquad\quad = \mathbf{A}\mathbb{X}_t - \underline{L}_t e_t & \\ e_t = \mathbb{C}'\,\mathbb{X}_t + w_t & \text{(noisy measurement)} \\ u_t = \mathbb{D}'\,\mathbb{X}_t & \text{(regulated output)} \end{cases}$

It is a control system where we want to minimize the power of the regulated output u by appropriately choosing $\underline{L}_t$ (namely to minimize the power of the "error" signal $\mathbb{X}$ weighted by $\mathbb{D}$, or simply put, to make the state $\mathbb{X}$ small, through proper controller design). More specifically, one may view e_t as the noisy measurement which is also the input to the controller, $(-\underline{L}_t)$ as the time-varying controller gain, $(-\underline{L}_t e_t)$ as the controller's output which is also the input to the system with state $\mathbb{X}_t$. The objective is to minimize $\mathbf{E}\|\underline{u}^T\|^2$; formally we want to solve

(63) $\qquad P_{T,n}(\mathbf{A}, \underline{C}) := \min_{\substack{\underline{L}_0, \dots, \underline{L}_t \\ s.t.\ (62),(\mathbf{A},\underline{C})\text{fixed}}} \frac{1}{T+1}\mathbf{E}\|\underline{u}^T(\mathbf{A}, \underline{C}, \underline{L}_0, \dots, \underline{L}_t)\|^2$

in which $\underline{M}$ is unknown to the controller. Note that the control effort is "free" as there is no direct penalty on the controller's output $(-\underline{L}_t e_t)$, and therefore, it is a *cheap control* problem, similar to the AWGN case studied in Section 2.

The signal e_t in (62) is the *KF innovation* or simply *innovation*, which forms a white process (cf. [29]), i.e., its covariance matrix $\mathbf{K}_{\underline{e}}^{(T)}$ is a diagonal matrix. Also $\underline{e}^T$ and $\underline{y}^T$ (and thus $\bar{\underline{y}}^T$) determine each other causally: $\underline{e}^T = \mathbf{V}\underline{y}^T$ where $\mathbf{V}$ is lower triangular with all-one diagonal elements. Hence $h(\underline{e}^T) = h(\underline{y}^T)$ and $|\mathbf{K}_{\underline{y}}^{(T)}| = |\mathbf{K}_{\underline{e}}^{(T)}|$.

For each t, the optimal $\underline{L}_t$ is determined as

$$(64) \qquad \underline{L}_t := \left[\begin{array}{c} \underline{L}_{1,t} \\ \underline{L}_{2,t} \end{array} \right] = \frac{\mathbb{A}\boldsymbol{\Sigma}_t\mathbb{C}}{K_{e,t}},$$

where $\boldsymbol{\Sigma}_t := \mathbf{E}\underline{\mathbb{X}}_t\underline{\mathbb{X}}_t'$, $K_{e,t} := \mathbf{E}(e_t)^2 = \mathbb{C}'\boldsymbol{\Sigma}_t\mathbb{C} + 1$. The error covariance matrix $\boldsymbol{\Sigma}_t$ satisfies the Riccati recursion

$$(65) \qquad \boldsymbol{\Sigma}_{t+1} = \mathrm{Ric}(\mathbb{A}, \mathbb{C}, \boldsymbol{\Sigma}_t)$$

with initial condition $\boldsymbol{\Sigma}_0 := \mathrm{diag}(\boldsymbol{I}_{n+1}, \boldsymbol{0})$, where the Riccati operator $\mathrm{Ric}(\cdot, \cdot, \cdot)$ is defined as

$$(66) \qquad \mathrm{Ric}(\mathbb{A}, \mathbb{C}, \boldsymbol{\Sigma}_t) := \mathbb{A}\boldsymbol{\Sigma}_t\mathbb{A}' - \frac{\mathbb{A}\boldsymbol{\Sigma}_t\mathbb{C}\,\mathbb{C}'\boldsymbol{\Sigma}_t\mathbb{A}'}{\mathbb{C}'\boldsymbol{\Sigma}_t\mathbb{C} + 1}.$$

This completes the description of the (KF-based) optimal feedback generator for a given $(\boldsymbol{A}, \underline{C})$.

The existence of one unified expression for three different systems (58), (57), and (62) is because the first two are actually two different non-minimal realizations (cf. [35]) of the third. The input-output mappings from $\underline{w}^T$ to $\underline{e}^T$ in the three systems are equivalent (see Appendix A.1). Thus we say that the three problems, the optimal estimation problem, the optimal feedback generator problem, and the cheap control problem, are *equivalent* in the sense that, if any one of the problems is solved, then the other two are solved.

6. Connections of fundamental limitations

In this section, the fundamental limitations and their connections are discussed. These limitations involve the information rate and channel input power in the feedback communication system, the Fisher information, MSE, entropy rate, and CRB in the estimation system, and the Bode integral and regulated output variance in the feedback control system. The limitation in one system may be expressed in terms of that in other systems, thanks to the equivalence established above. For an incomplete list of references on similar connections, see, e.g., [5, 6, 8, 10, 11, 13, 27, 28, 30, 37, 38]. Specifically, Mitter and Newton presented a rigorous theory on the duality between estimation and control in [30, 38]. Van Trees [37] (pp. 501-511), de Bruijn, and Guo *et al* [11] (and therein references and subsequent works) discussed filtering versus smoothing as well as their relation to entropy and mutual information.

6.1. Fisher information matrix (FIM), CRB, entropy rate, and Bode-type sensitivity integral (sum)

In estimation theory, it always holds, as a *fundamental limitation*, that

$$\text{(67)} \qquad \text{MSE}_{\underline{M}} \geq \text{CRB}_{\underline{M}} = \boldsymbol{\mathcal{I}}_{\underline{M}}^{-1},$$

regardless of how one designs the estimator [37], where $\text{MSE}_{\underline{M}}$ is the MSE for estimating $\underline{M}$, $\text{CRB}_{\underline{M}}$ is the (Bayesian) CRB, and $\boldsymbol{\mathcal{I}}_{\underline{M}}$ is the Fisher information matrix (FIM).

In feedback control, the Bode integral is a *fundamental limitation* (typically stated in steady state). Simply put, for any feedback design, the sensitivity of the output to exogenous disturbance cannot be made small uniformly over all frequencies, since the sensitivity transfer function's power spectrum in log scale sums up (integrates) to be constant [8]. Various forms of the Bode integral and the like were derived in e.g. [10, 39]. A similar limitation holds in finite horizon as we now show for the system $\mathbb{S}$ in (36). For any feedback generator $\boldsymbol{\mathcal{G}}_T$, the outputs of the closed-loop system is

$$\text{(68)} \qquad \underline{y}^T = (\boldsymbol{I} - \boldsymbol{\mathcal{Z}}_T^{-1}\boldsymbol{\mathcal{G}}_T)^{-1}(\boldsymbol{\mathcal{Z}}_T^{-1}\underline{r}^T + \underline{w}^T).$$

Since $\underline{e}^T = \boldsymbol{V}\underline{y}^T$, one may treat either $\underline{y}^T$ or $\underline{e}^T$ as the measurement outputs. Then the sensitivity of the measurements $\underline{e}^T$ against noise $\underline{w}^T$ is $\mathcal{S}_T := \boldsymbol{V}(\boldsymbol{I} - \boldsymbol{\mathcal{Z}}_T^{-1}\boldsymbol{\mathcal{G}}_T)^{-1}$. It is easily seen that, if the spectrum of $\mathcal{S}_T\mathcal{S}_T'$ is $\{\lambda_i\}_{i=1}^{T+1}$, then

$$\text{(69)} \qquad \sum_{i=1}^{T+1} \log \lambda_i = 0,$$

which holds valid regardless of the choice of feedback generator $\boldsymbol{\mathcal{G}}_T$, including the case of open loop (i.e. there is no feedback). Thus, the effect of noise $\underline{w}^T$ cannot be made uniformly small over time in the noisy measurements $\underline{e}^T$ by any feedback, which may be viewed as a fundamental limitation of noise (or disturbance) suppression.

We also introduce another way to quantify the disturbance rejection capability. Let

$$\text{(70)} \qquad \text{BI}_T := \frac{1}{2(T+1)} \log \frac{|\boldsymbol{K}_{\underline{y}}^{(T)}|}{|\boldsymbol{K}_{\underline{w}}^{(T)}|} = \frac{1}{2(T+1)} \sum_{i=1}^{T+1} \log \lambda_i(\boldsymbol{K}_{\underline{y}}^{(T)})$$

$$= \frac{1}{2(T+1)} \sum_{t=0}^{T} \log K_{e,t},$$

which, by Proposition 2 and $|\boldsymbol{K}_{\underline{y}}^{(T)}| = |\boldsymbol{K}_{\underline{e}}^{(T)}|$, is *independent of any causal feedback* $\boldsymbol{\mathcal{G}}_T$ and may be seen as the finite-horizon counterpart of the widely known Bode integral of infinite-horizon; that is, causal feedback can reshape the power spectrum for the noisy measurements $\underline{e}^T$ or channel outputs $\underline{y}^T$, but it cannot change the sum (in log scale) of the measurement power spectrum. Hence BI_T is an open-loop system property determined by $\mathcal{Z}_T$ and $\underline{r}^T$, which may be seen as an anti-causality measure invariant with or independent of the causal feedback. Note that $\underline{u}^T = \underline{r}^T$ in open loop for the system in (62); that is, $\underline{u}^T$ or equivalently $\underline{r}^T$ is the regulated output associated with $\underline{L}_t = 0$ for all t, and is also the initial-condition response for (62).

6.2. Expressions for mutual information and channel input power

Proposition 4. *Consider the coding scheme $\mathbb{S}$ with fixed $(\boldsymbol{A}, \underline{C})$ satisfying* *(**A1**). Then it holds that i)*

$$(71) \quad I(\underline{M}; \underline{y}^T) = (T+1)\mathrm{BI}_T = \frac{1}{2}\sum_{t=0}^{T}\log K_{e,t} = \frac{1}{2}\sum_{t=0}^{T}\log(\mathbb{C}'\boldsymbol{\Sigma}_t\mathbb{C}+1)$$

$$= \frac{1}{2}\log|\mathrm{MMSE}_{\underline{M},T}|^{-1} = \frac{1}{2}\log|\boldsymbol{\mathcal{I}}_{\underline{M},T}| = \frac{1}{2}\log|\mathrm{CRB}_{\underline{M},T}|^{-1};$$

and ii)

$$(72) \qquad P_{T,n}(\boldsymbol{A}, \underline{C}) = \frac{1}{T+1}\sum_{t=0}^{T}\mathbb{D}'\boldsymbol{\Sigma}_t\mathbb{D} = \frac{1}{T+1}\mathrm{trace}(\mathrm{CMMSE}_{\underline{r},T})$$

$$= \frac{1}{T+1}\sum_{t=0}^{T}\underline{C}'\boldsymbol{A}^t\mathrm{MMSE}_{\underline{M},t}\boldsymbol{A}^{t'}\underline{C},$$

where $MMSE_{\underline{M},T}$ is the minimum MSE of $\underline{M}$ at time T, $CMMSE_{\underline{r},T}$ is the causal minimum MSE of $\underline{r}^T$, $\boldsymbol{\mathcal{I}}_{\underline{M},T}$ is the FIM of $\underline{M}$ for the estimation system (58), and $CRB_{\underline{M},T}$ is the Bayesian CRB of $\underline{M}$.

Note that $\mathrm{CMMSE}_{\underline{r},T} := \mathbf{E}\|\widehat{\underline{r}}^T - \underline{r}^T\|^2$, in which $\widehat{\underline{r}}^T = [\widehat{r}_0, \ldots, \widehat{r}_T]$ contains the strictly causal estimates with one-step prediction $\widehat{r}_t := \underline{C}'\mathbf{E}(\underline{x}_t|\underline{y}^{t-1})$ for $t = 0, \ldots, T$.

Remark 3. We have described before and also re-described in this proposition that the mutual information (or information rate) cannot be made higher or lower than a constant regardless of the feedback generator choice;

the disturbance rejection measure BI cannot be made smaller than a constant regardless of the feedback controller design; the estimate error cannot be made smaller than a constant regardless of the estimator design. Furthermore, this proposition connects these limitations by showing they are essentially equivalent, and one limitation can be translated into another. For example, the amount of mutual information we may obtain is limited by the disturbance rejection ability in control and by how well the estimate can be in estimation. We remark these limitations are expressed in finite horizon, as opposed to the infinite-horizon limitations studied in [4]. It is also interesting to point out that a capacity notion can be introduced based on the MMSE decay rate and is closely related to the Shannon capacity for several Gaussian channels; see [28].

Proof: i) First we note that $h(\underline{y}^T) = h(\underline{e}^T)$ and $K_{e,t} = \underline{\mathbb{C}}'\Sigma_t\underline{\mathbb{C}} + 1$. Next, to find MMSE of $\underline{M}$, note that the linear model in Fig. 7 (a) is $\bar{\underline{y}}^T = \boldsymbol{\mathcal{Z}}_T^{-1}\boldsymbol{\Gamma}_T\underline{M} + \underline{w}^T$ with $\underline{M} \sim \mathcal{N}(\underline{0}, \boldsymbol{I})$, $\underline{w}^T \sim \mathcal{N}(\underline{0}, \boldsymbol{I})$. Thus we have

$$\text{(73)} \qquad \text{MMSE}_{\underline{M},t} = (\boldsymbol{I} + \boldsymbol{\Gamma}_T'\boldsymbol{\mathcal{Z}}_T^{-1\prime}\boldsymbol{\mathcal{Z}}_T^{-1}\boldsymbol{\Gamma}_T)^{-1},$$

yielding

$$\text{(74)} \qquad |\text{MMSE}_{\underline{M},t}| = |\boldsymbol{I} + \boldsymbol{\mathcal{Z}}_T^{-1}\boldsymbol{\Gamma}_T\boldsymbol{\Gamma}_T'\boldsymbol{\mathcal{Z}}_T^{-1\prime}|^{-1} = |\boldsymbol{I} + \boldsymbol{\mathcal{Z}}_T^{-1}\boldsymbol{K}_{\underline{r}}^{(T)}\boldsymbol{\mathcal{Z}}_T^{-1\prime}|^{-1}$$
$$= |\boldsymbol{K}_{\underline{r}}^{(T)} + \boldsymbol{K}_{\underline{Z}}^{(T)}|^{-1}.$$

In addition, from Sec. 2.4 in [37] we can directly compute the FIM of $\underline{M}$ to be $(\boldsymbol{I} + \boldsymbol{\Gamma}_T'\boldsymbol{\mathcal{Z}}_T^{-1\prime}\boldsymbol{\mathcal{Z}}_T^{-1}\boldsymbol{\Gamma}_T)$. Then i) follows from Proposition 2 and the system dynamics (62). ii) Since $u_t = \underline{\mathbb{D}}'\,\mathbb{X}_t = \underline{C}'\,\widetilde{x}_t = r_t - \widehat{r}_t$ and $\mathbf{E}\widetilde{x}_t\widetilde{x}_t' = \boldsymbol{A}^t\text{MMSE}_{\underline{M},t}\boldsymbol{A}^{t\prime}$, we have $\mathbf{E}(u_t)^2 = \underline{\mathbb{D}}'\boldsymbol{\Sigma}_t\underline{\mathbb{D}} = \underline{C}'\mathbf{E}\widetilde{x}_t\widetilde{x}_t'\underline{C} = \mathbf{E}(r_t - \widehat{r}_t)^2$, and then ii) follows. $\square$

6.3. Connections of the fundamental tradeoffs

The above fundamental limitations are based on *one fixed* $(\boldsymbol{A}, \underline{C})$ with $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Searching over *all* admissible $(\boldsymbol{A}, \underline{C})$ with $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ for all $n \leq T$, one obtains the optimal tradeoffs for feedback communication, estimation, and feedback control. By the optimal tradeoff in feedback communication we mean the optimal power-rate relation. The optimal tradeoffs in estimation and control will be characterized below, and it is expected that they correspond to the optimal tradeoff in feedback communication. Indeed, the optimal tradeoffs obtained by searching over all admissible $(\boldsymbol{A}, \underline{C})$ are optimal

over all (possibly nonlinear) feedback communication designs [5], estimator designs, and feedback control designs, as elaborated below.

The fundamental tradeoff in the feedback communication problem over the channel $\mathcal{F}$ for the finite-horizon from time 0 to time T is the capacity $C_T(\mathcal{P})$ (or $P_T(\mathcal{R})$, see definitions in (30) and (32)) expressed as the optimal power-rate pair where the power is viewed as the cost and the rate is the utility. Though it is desired to make the cost low and the utility high, the cost cannot be arbitrarily low for any given utility, and the utility cannot be arbitrarily high for any given cost, regardless of how one may design the coding scheme. In other words, although one wishes to obtain a high rate using a low power, one can only obtain no more than a certain rate with the given power, and cannot arbitrarily lower the power in order to achieve a desired rate. To capture this, let $\boldsymbol{f} := (\boldsymbol{A}, \underline{C}, \boldsymbol{\mathcal{G}}_T)$ be any coding scheme over the channel $\mathcal{F}$, $R_T(\boldsymbol{f}) := \frac{1}{T+1} I\left(\underline{u}^T(\boldsymbol{f}) \to \underline{y}^T(\boldsymbol{f})\right)$ be the information rate with $\boldsymbol{f}$, and $P_T(\boldsymbol{f}) := \frac{1}{T+1}\mathbf{E}\|\underline{u}^T(\boldsymbol{f})\|^2$ be the average channel input power with $\boldsymbol{f}$. Here we used the concept of directed information (see [6]) from $\underline{u}^T$ to $\underline{y}^T$ defined as

$$(75) \qquad I(\underline{u}^T \to \underline{y}^T) := \sum_{t=0}^{T} I(\underline{u}^t; y_t | \underline{y}^{t-1}).$$

Optimal Feedback Communication Tradeoff: Consider the feedback communication problem over the channel $\mathcal{F}$ with one-step delayed output feedback. **i)** Given any $\mathcal{P} > 0$, then the rate $R_T(\boldsymbol{f})$ cannot be higher than a constant (the constant being $C_T(\mathcal{P})$) for any $\boldsymbol{f}$ such that $P_T(\boldsymbol{f}) \leq \mathcal{P}$. **ii)** Given any $\mathcal{R} > 0$, then the power $P_T(\boldsymbol{f})$ cannot be lower than a constant (the constant being $P_T(\mathcal{R})$) for any $\boldsymbol{f}$ such that $R_T(\boldsymbol{f}) \geq \mathcal{R}$.

In the power versus rate tradeoff, one may see the power as a causal quantity whereas the rate as an anti-causal quantity. This is because the average input power depends on the strictly causal feedback from the channel output. On the other hand, the information rate is independent of the causal feedback and can be obtained *a priori* before executing the feedback system; therefore, one may view the rate as a measure of anti-causality of the system. Hence the power versus rate tradeoff can be seen as a causality versus anti-causality tradeoff.

---

[5]This can be seen from Lemma 1 and Proposition 3: Linear $(\boldsymbol{A}, \underline{C})$ can achieve $P_{T,n}(\mathcal{R})$ and hence $C_T(\mathcal{P})$.

A fundamental tradeoff for the estimation problem over the channel $\mathcal{F}$ is the causal estimation performance versus anti-causal estimation performance. Consider *any* unknown process $\underline{r}^T$ with known covariance matrix, and $\underline{r}^T$ passes through the channel $\mathcal{F}$ to generate noisy measurements $\bar{y}^T$. Let us assume $\underline{r}^T$ is generated from the normalized, whitened signal $\underline{M}$ and $\underline{M}$ is to be estimated. That is, it holds that $\underline{M} := \mathbf{R}^{-1}\underline{r}^T$, where $\mathbf{R} := \left(\boldsymbol{K}_{\underline{r}}^{(T)}\right)^{1/2}$ if $\boldsymbol{K}_{\underline{r}}^{(T)}$ is of full rank; otherwise $\mathbf{R}$ is such that $\boldsymbol{K}_{\underline{M}} = \boldsymbol{I}$ holds that $\mathrm{rank}(\boldsymbol{K}_{\underline{M}}) = \mathrm{rank}(\boldsymbol{K}_{\underline{r}}^{(T)})$. Therefore we have a linear model (which includes the linear model in Fig. 6 (a) as a special case)

$$(76) \qquad \bar{y}^T = \boldsymbol{\mathcal{Z}}_T^{-1}\mathbf{R}\underline{M} + \underline{w}^T.$$

Let $\boldsymbol{f}$ be any estimator which outputs a causal, one-step prediction $\widehat{r}_{t+1|t}(\boldsymbol{f})$ at each t and a smoothed estimate $\widehat{\underline{M}}_{T|T}(\boldsymbol{f})$ at time T. Then the time-averaged causal MSE with one-step prediction, for the source $\underline{r}^T$ with the estimator $\boldsymbol{f}$, is

$$(77) \qquad \mathrm{CMSE}(\underline{r}^T, \boldsymbol{f}) := \frac{1}{T+1}\sum_{t=0}^{T}\mathbf{E}\left(r_t - \widehat{r}_{t|t-1}(\boldsymbol{f})\right)^2,$$

and the decay rate of the anti-causal, smoothing MSE, for the source $\underline{r}^T$ with the estimator $\boldsymbol{f}$, is

$$(78) \qquad \mathrm{R\text{-}MSE}(\underline{r}^T, \boldsymbol{f})$$
$$:= \frac{1}{2(T+1)}\log|\mathrm{MSE}(\underline{r}^T, \boldsymbol{f})|^{-1}$$
$$= -\frac{1}{2(T+1)}\log\left|\mathbf{E}\left((\underline{M} - \widehat{\underline{M}}_{T|T}(\boldsymbol{f}))(\underline{M} - \widehat{\underline{M}}_{T|T}(\boldsymbol{f}))'\right)\right|.$$

The CMSE is viewed as the cost since the error is desired to be small, and R-MSE the utility since the error is desired to decay fast.

Optimal Estimation Tradeoff: Consider the estimation problem over the channel $\mathcal{F}$. **i)** Given any $\mathcal{P} > 0$, then the smoothing rate R-MSE$(\underline{r}^T, \boldsymbol{f})$ cannot be larger than a constant (the constant being $C_T(\mathcal{P})$), *for any source $\underline{r}^T$ and any estimator $\boldsymbol{f}$ such that* CMSE$(\underline{r}^T, \boldsymbol{f}) \leq \mathcal{P}$. **ii)** Given any $\mathcal{R} > 0$, then CMSE$(\underline{r}^T, \boldsymbol{f})$ cannot be smaller than a constant (the constant being $P_T(\mathcal{R})$), *for any source $\underline{r}^T$ and any estimator $\boldsymbol{f}$ such that* R-MSE$(\underline{r}^T, \boldsymbol{f})$ $\geq \mathcal{R}$.

Since CMSE is causal but the smoothing MSE is anti-causal and independent of the causal operation (if any) done by the estimator, this prediction versus smoothing tradeoff is a causality versus anti-causality tradeoff. Note that the statement above "for any source $\underline{r}^T$ and any estimator $\boldsymbol{f}$" is equivalent to "for any admissible $(\boldsymbol{A}, \underline{C}, \underline{L}_0, \ldots, \underline{L}_T)$" as we mentioned that there is no loss of generality in considering linear systems only.

A fundamental tradeoff for the cheap control problem (62) over the channel $\mathcal{F}$ is the closed-loop control performance (regulated output variance, in this case the variance of the channel input $u(\boldsymbol{f})$) versus the Bode integral (the disturbance rejection measure in (70)) $\mathrm{BI}_T(\boldsymbol{f})$, for any control design $\boldsymbol{f} := (\boldsymbol{A}, \underline{C}, \underline{L}_0, \ldots, \underline{L}_T)$. The control performance is defined as $P_u(\boldsymbol{f}) := \frac{1}{T+1} \sum_{t=0}^{T} \mathbf{E}(u_t(\boldsymbol{f}))^2$, which is viewed as the cost since one wishes to keep it small. The disturbance rejection metric is desired to be high and is the utility since a higher $\mathrm{BI}_T(\boldsymbol{f})$ implies a relatively smaller contribution of the noise.

Optimal Feedback Control Tradeoff: Consider the control problem over the channel $\mathcal{F}$. **i)** Given any $\mathcal{P} > 0$, then the Bode integral $\mathrm{BI}_T(\boldsymbol{f})$ cannot be larger than a constant (the constant being $C_T(\mathcal{P})$) for any control design $\boldsymbol{f}$ such that the regulated output variance $P_u(\boldsymbol{f}) \leq \mathcal{P}$. **ii)** Given any $\mathcal{R} > 0$, then $P_u(\boldsymbol{f})$ cannot be smaller than a constant (the constant being $P_T(\mathcal{R})$), for any control design $\boldsymbol{f}$ such that $\mathrm{BI}_T(\boldsymbol{f}) \geq \mathcal{R}$.

Note that this specifies the relation between the (closed-loop) control performance achievable via causal feedback and the (open-loop) anti-causality of the system.

To summarize, all three tradeoffs are equivalent to each other and are essentially the *fundamental tradeoff between causality and anti-causality*, which manifests itself in the three different but closely related problems. The causal entities, e.g. the channel inputs in feedback communication, one-step prediction in estimation, and regulated output in control, are closed-loop entities generated in a causal, successive way by the causal feedback, and vary as the causal feedback varies. In contrast, the anti-causal entities, e.g. the information rate (and the decoded message) in communication, the smoothed estimate in estimation, and the BI in control, are invariant regardless of whether the systems are in open-loop or closed-loop or how the closed-loop is done. Note that this does not claim that the open-loop capacity is equal to closed-loop capacity, since the capacity is in fact the pair of rate and power.

7. Asymptotic analysis of the feedback system

By far we have completed our analysis in finite-horizon. Below, we briefly consider the steady-state communication problem, by studying the limiting behavior (T going to infinity) of the finite-horizon feedback system $\mathbb{S}^*$ while fixing the encoder dimension to be $(n + 1)$. (The infinite-horizon capacity problem, solved by Kim in [18], is beyond the main scope of this paper and will not be discussed.) We make the following assumption in $\mathbb{S}^*$:

(**A2**) $(\boldsymbol{A}, \underline{C}')$ is observable, and $\boldsymbol{A}$ has no eigenvalue on the unit circle or at the locations of the eigenvalues of $\boldsymbol{F}$.

7.1. Convergence to steady-state

The time-varying KF in (62) converges to a steady-state, namely (62) is *stabilized* in closed loop. That is, asymptotically (62) becomes an LTI system

$$(79) \qquad \text{steady-state:} \begin{cases} \mathbb{X}_{t+1} = (\mathbb{A} - \underline{L}\,\underline{\mathbb{C}}')\mathbb{X}_t - \underline{L}w_t = \mathbb{A}\mathbb{X}_t - \underline{L}e_t \\ e_t = \underline{\mathbb{C}}'\,\mathbb{X}_t + w_t \\ u_t = \underline{\mathbb{D}}'\,\mathbb{X}_t, \end{cases}$$

where

$$(80) \qquad \underline{L} := \frac{\mathbb{A}\boldsymbol{\Sigma}\mathbb{C}}{K_e},$$

$K_e = \underline{\mathbb{C}}'\boldsymbol{\Sigma}\underline{\mathbb{C}} + 1$, and $\boldsymbol{\Sigma}$ is the unique stabilizing solution to the Riccati equation $\boldsymbol{\Sigma} = \mathrm{Ric}(\mathbb{A}, \mathbb{C}, \boldsymbol{\Sigma})$.

This LTI system is sometimes easy to analyze and to implement. For instance, the cheap control (cf. [24] and [34]) of an LTI system asserts that the transfer function from w to e is an *all-pass* function in the form of

$$(81) \qquad \mathcal{T}_{we}(z) = \prod_{i=0}^{k} \frac{z - \lambda_i}{z - \lambda_i^{-1}}$$

where $\lambda_0, \ldots, \lambda_k$ are the *unstable eigenvalues* of $\boldsymbol{A}$ or $\mathbb{A}$ (noting that $\boldsymbol{F}$ is stable; see (59)), which also implies the whiteness of innovations process $\{e_t\}$ since the all-pass function $\mathcal{T}_{we}(z)$ holds that $|\mathcal{T}_{we}(\mathrm{e}^{j2\pi\theta})|$ is constant for all θ. It also shows that the closed-loop eigenvalues are the reciprocals of the open-loop ones.

Proposition 5. *Consider the Riccati recursion $\Sigma_{t+1} = \text{Ric}(\mathbb{A}, \mathbb{C}, \Sigma_t)$ and the system (62). Assume (A2) and that $\boldsymbol{A}$ has $(k+1)$ unstable eigenvalues. Then i) Starting from any initial condition $\Sigma_0 := \text{diag}(\boldsymbol{I}_{n+1}, \boldsymbol{0})$, the sequence $\{\Sigma_t\}$ converges to Σ_∞, the unique stabilizing solution to the Riccati equation $\Sigma = \text{Ric}(\mathbb{A}, \mathbb{C}, \Sigma)$, and Σ_∞ has rank $(k+1)$. ii) The time-varying system (62) converges to the unique steady-state as given in (79).*

Proof. See Appendix A.4.   □

7.2. Steady-state fundamental limitations

Now we fix $(\boldsymbol{A}, \underline{C})$ and let the horizon T in the coding scheme $\mathbb{S}^*$ go to infinity. Let $\mathbb{N}(\widetilde{\underline{M}_T}) := e^{\frac{2}{n+1} h(\widetilde{\underline{M}_T})}$ be the *entropy power* of the smoothing error $\widetilde{\underline{M}_T} := \underline{M} - \widetilde{\underline{M}}_{T|T}$,

$$(82) \qquad\qquad DI(\boldsymbol{A}) := \prod_{i=0}^{k} |\lambda_i|$$

be the *degree of instability* or the *degree of anti-causality* of $\boldsymbol{A}$ (as λ_i's are in the right-half plane), and $S(e^{j2\pi\theta}) := e(e^{j2\pi\theta})/w(e^{j2\pi\theta})$ be the spectrum of the sensitivity function of system (79) (cf. [4]). Denote the *asymptotic linear growth rate* of a sequence $\{a_T\}$ as

$$(83) \qquad\qquad \sigma(a_T) := \lim_{T\to\infty} \frac{a_T}{T+1}$$

and the *asymptotic exponential growth rate* of a sequence $\{a_T\}$ as

$$(84) \qquad\qquad \rho(a_T) := \lim_{T\to\infty} \frac{\log a_T}{T+1}.$$

Note that in this paper we only encounter the cases where the above limits exist.

Then the limiting result of Proposition 4 is summarized in the next proposition. See Remark 1 and (37) for the notation $R_{T,n}(\boldsymbol{A}, \underline{C})$, which is extended to the infinite horizon below.

Proposition 6. *Consider the coding scheme $\mathbb{S}^*$. Fix any $n \geq 0$ and $(\boldsymbol{A}, \underline{C}')$ with $\boldsymbol{A} \in \mathbb{R}^{(n+1)\times(n+1)}$ satisfying (A2). Then the asymptotic information rate*

is given by

$$(85) \qquad R_{\infty,n}(\boldsymbol{A}, \underline{C}) := \sigma(I(\underline{M}; \underline{y}^T))$$

$$= \frac{1}{2} \log K_e = \sigma(h(e_T)) - \frac{1}{2} \log 2\pi e = \frac{1}{2} \log(\mathbb{C}'\boldsymbol{\Sigma}\mathbb{C} + 1)$$

$$= \frac{1}{2} \rho(|\boldsymbol{\mathcal{I}}_{M,T}|) = -\frac{1}{2}\rho(|\mathrm{MSE}_{\underline{M},T}|) = -\frac{1}{2}\rho(|\mathrm{CRB}_{\underline{M},T}|)$$

$$= -\frac{1}{2}\rho(\mathbb{N}(\widetilde{\underline{M}}_T)) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log |S(e^{j2\pi\theta})| d\theta = \log DI(\boldsymbol{A}),$$

and the average channel input power is given by

$$(86) \qquad P_{\infty,n}(\boldsymbol{A}, \underline{C}) := \sigma(\mathbf{E}\|\underline{u}^T\|^2) = \mathbb{D}'\boldsymbol{\Sigma}\mathbb{D}.$$

Proof. Proposition 5 leads to that, the limits of the results in Proposition 4 are well defined. Then

$$(87) \qquad R_{\infty,n}(\boldsymbol{A}, \underline{C}) = \lim_{T\to\infty} \frac{1}{2(T+1)} \sum_{t=0}^{T} \log K_{e,t}$$

$$= \frac{1}{2} \log K_{e,\infty} = \sigma(h(e_T)) - \frac{1}{2} \log 2\pi e,$$

where the second equality is due to the Cesaro mean, and the last equality follows from the definition of entropy rate of a Gaussian process. Now by (81), $\{e_t\}$ has a flat power spectrum with magnitude $DI(\boldsymbol{A})^2$. Then $R_{\infty,n}(\boldsymbol{A}, \underline{C}) = \log DI(\boldsymbol{A})$. Since

$$(88) \qquad h(\widetilde{\underline{M}}_T) = \frac{1}{2} \log(2\pi e)^{n+1} \left| \boldsymbol{K}_{\underline{\widetilde{M}}}^{(T)} \right| = \frac{1}{2} \log(2\pi e)^{n+1} |\mathrm{MSE}_{\underline{M},T}|,$$

we have $\rho(\mathbb{N}(\widetilde{\underline{M}}_T)) = \rho(\mathrm{MSE}_{\underline{M},T})$. The Bode integral follows from [4]. The other equalities are the direct applications of the Cesaro mean to the results in Proposition 4. $\square$

   This proposition shows that all the fundamental limitations for the information rate obtained in Proposition 4 become the logarithm of the degree of anti-causality of $\boldsymbol{A}$. One consequence of this is stable eigenvalues in $\boldsymbol{A}$ do not affect the rate (also pointed out in [4]). Stable eigenvalues do not affect $P_{\infty,n}(\boldsymbol{A}, \underline{C})$, either, as shown in the following corollary. Therefore, the feedback communication problem studied in this paper can be reduced to a problem of tracking an anti-stable source over the Gaussian channel ([4, 5]).

Proposition 7. *Suppose that $(\boldsymbol{A}, \underline{C})$ with $\boldsymbol{A} \in \mathbb{R}^{(n+1)\times(n+1)}$ satisfies (A2). Suppose further that $\boldsymbol{A}$ has $(k+1)$ unstable eigenvalues denoted $\lambda_0, \ldots, \lambda_k$ where $0 \leq k \leq (n+1)$. Then there exists an observable pair $(\boldsymbol{A}_k, \underline{C}'_k)$ with $\boldsymbol{A}_k \in \mathbb{R}^{(k+1)\times(k+1)}$ being anti-stable such that $R_{\infty,n}(\boldsymbol{A}, \underline{C}) = R_{\infty,k}(\boldsymbol{A}_k, \underline{C}_k)$ and $P_{\infty,n}(\boldsymbol{A}, \underline{C}) = P_{\infty,k}(\boldsymbol{A}_k, \underline{C}_k)$.*

Proof. See Appendix A.5.   □

It is straightforward to extend the finite-horizon connections between the fundamental tradeoffs for feedback communication, estimation, and feedback control, presented in Section 6.3, to infinite horizon, since the limits exist. The detail is skipped for brevity.

8. More general channels

In this section, we will extend some previous results over the Gaussian channel to more general channels. We will present an inequality of the mutual information rate and estimation error decay rates, and will show that the causal MMSE predictor is necessary for power efficiency for general channels with feedback.

8.1. Information rate versus estimation error decay rate

Consider a channel with input $\underline{u}^T$ and output $\underline{y}^T$ for the time horizon spanning from 0 to T, in which each u_t can depend on the message signal $\underline{M}$ and past outputs $\underline{y}^{t-1}$ for any t between 0 and T. Assume the message signal $\underline{M}$ of size $n \times 1$ is fixed for any time horizon, and the recovered message is $\widehat{\underline{M}}_T$. Assume all the involved differential entropies and limits exist and are finite. Let 1) $\widetilde{\underline{M}}_T := \underline{M}_T - \widehat{\underline{M}}_T$ be the estimation error; 2) $h_{\widetilde{\underline{M}}_{T,MEE}}$ be the minimum (infimum) error entropy (MEE), i.e. $h_{\widetilde{\underline{M}}_{T,MEE}} := \inf h_{\widetilde{\underline{M}}_T}$ for any possible $\widetilde{\underline{M}}_T$ [36]; 3) $H_{\widetilde{\underline{M}},MEE} := \sigma(h_{\widetilde{\underline{M}}_{T,MEE}})$, i.e. $H_{\widetilde{\underline{M}},MEE}$ is the entropy rate associated with the MEE estimate; 4) $\rho_{\widetilde{\underline{M}},MEE} := \rho(\mathbb{N}(\widetilde{\underline{M}}_{T,MEE}))$, i.e. $\rho_{\widetilde{\underline{M}},MEE}$ is the exponential increasing rate of the entropy power for $\widetilde{\underline{M}}_{T,MEE}$; 5) $\rho_{\widetilde{\underline{M}},MMSE} := \rho(|\text{MMSE}_{\underline{M}_T}|)$ where $\text{MMSE}_{\underline{M}_T}$ is the MMSE for estimating $\underline{M}_T$; and 6) $R := \sigma(I(\underline{M}; \underline{y}^T))$ be the information rate across the channel.

Proposition 8. *It holds that*

$$(89) \qquad R \geq -H_{\widetilde{\underline{M}},MEE} = -\frac{n}{2}\rho_{\widetilde{\underline{M}},MEE} \geq -\frac{1}{2}\rho_{\widetilde{\underline{M}},MMSE}$$

and the first equality holds if and only if $\sigma(I(\widetilde{\underline{M}}_{T,MEE}; \underline{y}^T)) = 0$.

A sufficient condition for $\sigma(I(\widetilde{\underline{M}}_T; \underline{y}^T)) = 0$ is $I(\widetilde{\underline{M}}_T; \underline{y}^T) = 0$, namely the estimation error is *independent* (not just orthogonal) to the observations. In this case it is easily seen that the first and last equalities in (89) hold (since if $I(\widetilde{\underline{M}}_T; \underline{y}^T) = 0$ then the corresponding $\widehat{\underline{M}}_T$ is the MMSE estimate). Therefore, the proposition states that *one cannot expect the MEE (or MMSE, or any) estimator error to decay at a rate higher than the information rate, and if the estimation error can be made independent of the observations, the estimation error decays at a rate equal to how fast information is transmitted over the channel*; here the decay rate may be the linear decay rate of the entropy, the exponential decay rate of the entropy power, or the exponential decay rate of the mean squares.

Proof. For any time horizon, it holds that

$$(90) \qquad I(\underline{M}; \underline{y}^T) = h(\underline{M}) - h(\underline{M}|\underline{y}^T) = h(\underline{M}) - h(\underline{M} - \widehat{\underline{M}}_T|\underline{y}^T)$$
$$= h(\underline{M}) - h(\widetilde{\underline{M}}_T|\underline{y}^T) \overset{(a)}{=} h(\underline{M}) - h(\widetilde{\underline{M}}_T) + I(\widetilde{\underline{M}}_T; \underline{y}^T)$$
$$\overset{(b)}{\geq} h(\underline{M}) - h(\widetilde{\underline{M}}_T).$$

From (b), one obtains that for any estimation error $\widetilde{\underline{M}}_T$, it holds that

$$(91) \qquad \sigma(I(\underline{M}; \underline{y}^T)) \geq \sigma(h(\underline{M})) - \sigma(h(\widetilde{\underline{M}}_T))$$
$$= -\sigma(h(\widetilde{\underline{M}}_T)) = -\frac{n}{2}\rho(\mathbb{N}(\widetilde{\underline{M}}_T));$$

by straightforward computation. Clearly, the above inequality holds for the MEE estimate. From (a), it can be seen that the equality can be achieved if $\sigma(I(\widetilde{\underline{M}}_T; \underline{y}^T)) = 0$.

By the definition of MEE estimate, it holds that $h_{\widetilde{\underline{M}}_{T,MEE}} \leq h_{\widetilde{\underline{M}}_{T,MMSE}}$ $\leq \frac{1}{2}\log(2\pi e)^n |MMSE_{\underline{M},T}|$ and thus

$$(92) \qquad \sigma\left(h_{\widetilde{\underline{M}}_{T,MEE}}\right) \leq \sigma\left(\frac{1}{2}\log(2\pi e)^n |MMSE_{\underline{M},T}|\right) = \frac{1}{2}\rho(|MMSE_{\underline{M},T}|).$$

Then the proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

8.2.　Necessity of the MMSE predictor for general channels with feedback

Assume an arbitrary additive channel

$$
(93) \qquad \underline{y}^T = \mathbf{H}\underline{u}^T + \underline{Z}^T
$$

with an average power constraint $\mathbf{E}\|\underline{u}^T\|^2 \le (T+1)\mathcal{P}$, where $\underline{Z}^T$ is an arbitrary additive noise process. Assuming one-step delayed channel output feedback, the next Proposition states that the optimal feedback generator has to be an MMSE one-step predictor. Let $I(\underline{u}^T \to \underline{y}^T)$ be the directed information from $\underline{u}^T$ to $\underline{y}^T$ of some appropriate form that can be used to capture the information rate across the channel [6].

Proposition 9. *Let* $\widetilde{u}_t := u_t - \mathbf{E}(u_t|\underline{y}^{t-1})$ *and* $\widetilde{\underline{y}}^T := \mathbf{H}\widetilde{\underline{u}}^T + \underline{Z}^T$. *Then* $I(\underline{u}^T \to \underline{y}^T) = I(\widetilde{\underline{u}}^T \to \widetilde{\underline{y}}^T)$ *and* $\mathbf{E}\|\widetilde{\underline{u}}^T\| \le \mathbf{E}\|\underline{u}^T\|^2$.

Proof. Note that $\mathbf{E}(u_t|\underline{y}^{t-1})$ can be generated and added back to the channel output at the receiver side and hence the directed information across the channel is the same using either $\underline{u}^T$ or $\widetilde{\underline{u}}^T$ as the channel inputs. The average power of using $\widetilde{\underline{u}}^T$ is no larger since the conditional mean $\mathbf{E}(u_t|\underline{y}^{t-1})$ has the minimum MSE. $\qquad\qquad\square$

Simple as it is, this necessary condition for optimality is rather universal. A corollary is that in the optimal feedback coding scheme the current channel input u_t is independent of all past channel outputs $\underline{y}^{t-1}$ by the Projection Theorem. Moreover, since $\mathbf{E}\widetilde{u}_t = \mathbf{E}(u_t - \mathbf{E}(u_t|\underline{y}^{t-1})) = 0$ by the law of total expectation, it is a center-of-gravity encoding rule (cf. [16, 22, 40]). It is also straightforward to see that if the channel output feedback delay is d steps, then an MMSE d-step predictor is needed for optimality. [6]

9.　Conclusions and future work

We proposed a perspective that integrates information transmission (communication), information processing (estimation), and information utilization (feedback control). We established a certain equivalence and explored

---

[6]We note that in [17] the authors demonstrated the significance of a principle called Posterior Matching in feedback communications and proposed optimal feedback communication based on the principle; its connections with this paper will be addressed in future study.

fundamental limitations in feedback communication, estimation, and feedback control over Gaussian channels with memory. We demonstrated that a simple reformulation of the Kalman filter becomes the celebrated Schalkwijk-Kailath codes, and the well-studied Cover-Pombra scheme necessarily contains a Kalman filter in order to be optimal. We characterized the roles of Kalman filtering in an optimal feedback communication system as to ensure power efficiency and to optimally recover the transmitted codewords. We showed that the fundamental limitations/tradeoffs in these three systems also coincide: The power versus rate tradeoff in feedback communication, the causal prediction MSE versus smoothing MSE tradeoff in estimation, and the control performance versus Bode integral tradeoff in control, are equivalent and in essence, are a causality versus anti-causality tradeoff. Some of the results were extended to the cases with more general channels.

Our perspective has been employed in [41] to uniformly address the fundamental limits of several classes of feedback communication problems, and we envision that this perspective can generate a new avenue for studying more general feedback communication problems, such as multiuser feedback communications. Future research may include extending our proposed scheme to address the optimality of other feedback communication problems (e.g. MIMO systems with output feedback). We also anticipate that the perspective and the approaches developed in this paper be extended and help build a theoretically and practically sound paradigm that unifies information, estimation, and control.

Appendix A.

A.1. Systems representations and equivalence

The concepts of linear system representations and the equivalence between different representations are used in this paper. We introduce a finite-horizon input-output equivalence below; for a thorough treatment (including both finite- and infinite-horizon systems), see e.g. [35].

Definition 3. Fix $0 \leq T < \infty$. Two linear mappings $\mathcal{M}_{i,T} : \mathbb{R}^{q(T+1)} \to \mathbb{R}^{p(T+1)}$, $i = 1, 2$, are said to be *equivalent* if for any $\underline{u}^T \in \mathbb{R}^{q(T+1)}$, it holds that $\mathcal{M}_{1,T}(\underline{u}^T) = \mathcal{M}_{2,T}(\underline{u}^T)$.

Example: Equivalence between the estimation system (58) and coding scheme $\mathbb{S}$ (57) To show the equivalence, it is sufficient to show that the mappings from the system input $(\underline{M}', \underline{w}^{T'})'$ to system output $(\underline{r}^{T'},$

$\underline{e}^{T\prime}, \widehat{\underline{r}}^{T\prime})\prime$ (or any other choice of output formed by corresponding signals in the two systems) are equal; that is, for each t, the sets of signals r_t, $\widehat{r}_t$, e_t, $\underline{x}_t$, and $\widehat{\underline{x}}_t$ in (58) and (57) are equal, respectively. To this aim, first note that for $t = 0$, the sets of signals are equal, respectively, and that $\underline{s}_0 - \widehat{\underline{s}}_0 = \bar{\underline{s}}_0 - \widehat{\bar{\underline{s}}}_0$. Assume that for $t \le \tau$, the sets of signals are equal, respectively, and that $\underline{s}_\tau - \widehat{\underline{s}}_\tau = \bar{\underline{s}}_\tau - \widehat{\bar{\underline{s}}}_\tau$. Now use induction. Apparently, $\underline{x}_{\tau+1}$ (and hence $r_{\tau+1}$) generated by (58) and (57) are equal, respectively. Then

$$(\text{A.1}) \quad \underline{s}_{\tau+1} - \widehat{\underline{s}}_{\tau+1} = \boldsymbol{F}(\underline{s}_\tau - \widehat{\underline{s}}_\tau) + \underline{G}(r_\tau - \widehat{r}_\tau) - \underline{L}_{2,\tau}e_\tau = \bar{\underline{s}}_{\tau+1} - \widehat{\bar{\underline{s}}}_{\tau+1},$$

and $e_{\tau+1}$ from both (58) and (57) equals $\underline{H}\prime(\bar{\underline{s}}_\tau - \widehat{\bar{\underline{s}}}_\tau) + (r_\tau - \widehat{r}_\tau) + w_{\tau+1}$. Thus we have proven the equivalence.

Likewise, we can show that the estimation system (58), feedback communication system (57), and control system (62) are equivalent.

Example: Derivation of state-space representation of $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$ Consider $\boldsymbol{\mathcal{G}}_T^*(\boldsymbol{A}, \underline{C})$ in (50) in Section 4, where the state-space representations for $\widehat{\boldsymbol{\mathcal{G}}}_T^*(\boldsymbol{A}, \underline{C})$ and $\boldsymbol{\mathcal{Z}}_T^{-1}$ are illustrated in Fig. 7 (a) and Fig. 4. Since (50) suggests a feedback connection of $\widehat{\boldsymbol{\mathcal{G}}}^*$ and $\boldsymbol{\mathcal{Z}}^{-1}$ as shown in Fig. 6 (b), we can write the state-space for $\boldsymbol{\mathcal{G}}^*$ as

$$(\text{A.2}) \quad \left\{ \begin{array}{l} \widehat{\underline{x}}_{t+1} = \boldsymbol{A}\widehat{\underline{x}}_t + \underline{L}_{1,t}e_t \\ \widehat{r}_t = \underline{C}\prime\widehat{\underline{x}}_t \\ \widehat{\bar{\underline{s}}}_{t+1} = \boldsymbol{F}\widehat{\bar{\underline{s}}}_t + \underline{G}\widehat{r}_t + \underline{L}_{2,t}e_t \\ e_t = \bar{\underline{y}}_t - \underline{H}\prime\widehat{\bar{\underline{s}}}_t - \widehat{r}_t \\ \underline{s}_{a,t+1} = \boldsymbol{F}\underline{s}_{a,t} + \underline{G}\widehat{r}_t \\ \bar{\underline{y}}_t = y_t + \underline{H}\prime\underline{s}_{a,t} + \widehat{r}_t \end{array} \right. \begin{array}{l} \text{Kalman filter } \widehat{\boldsymbol{\mathcal{G}}}_T^*(\boldsymbol{A}, \underline{C}) \\ \\ \boldsymbol{\mathcal{Z}}_T^{-1} \end{array}$$

Letting $\widehat{\underline{s}}_t := \widehat{\bar{\underline{s}}}_t - \underline{s}_{a,t}$, the above reduces to

$$(\text{A.3}) \quad \left\{ \begin{array}{l} \widehat{\underline{x}}_{t+1} = \boldsymbol{A}\widehat{\underline{x}}_t + \underline{L}_{1,t}e_t \\ \widehat{r}_t = \underline{C}\prime\widehat{\underline{x}}_t \\ \widehat{\underline{s}}_{t+1} = \boldsymbol{F}\widehat{\underline{s}}_t + \underline{L}_{2,t}e_t \\ e_t = y_t - \underline{H}\prime\widehat{\underline{s}}_t \end{array} \right.$$

which is the dynamics shown in Fig. 7 (b) for $\boldsymbol{\mathcal{G}}^*$, as claimed in the proof for Proposition 3. The above reduction of realization is allowed since it preserves the system equivalence.

A.2. Cheap control

We briefly describe the scalar LTI formulation of the cheap control problem and its related expensive control. Further details may be found in [34], [4], and references therein. Consider a scalar LTI system

(A.4)
$$\tilde{x}_t = a\tilde{x}_{t-1} - ly_{t-1}$$
$$u_t = c\tilde{x}_t$$
$$y_t = u_t + w_t,$$

where $a > 1$ and $c \neq 0$ are given. We need to find the control gain l (i.e., the controller's output is $-ly_t$) to solve the following optimal control problem in infinite horizon:

(A.5)
$$\min_{l:(A.4)\text{ is stable}} \sum_{k=0}^{\infty} \mathbf{E}(u_t)^2,$$

which is called a cheap control problem as the objective function does not assign any direct penalty to the control effort $-ly_t$.

Alternatively, instead of assuming c is given, we may assume l is given in (A.4), and the state feedback control gain c (i.e. the controller's output is $u_t = c\tilde{x}_t$) is to be found to solve the following problem in infinite horizon:

(A.6)
$$\min_{c:(A.4)\text{ is stable}} \sum_{k=0}^{\infty} \mathbf{E}(u_t)^2,$$

which is called an expensive control problem as the sole objective is to reduce the control effort u_t.

In either problem, when (A.4) is stable, it holds that

(A.7)
$$\lim_{t\to\infty} \mathbf{E}(u_t)^2 = c^2 l^2 \sum_{t=0}^{\infty} (a - lc)^{2t} = \frac{c^2 l^2}{1 - (a - lc)^2}.$$

Then straightforward calculations can show that the optimizing solution must satisfy

(A.8)
$$lc = (a^2 - 1)/a,$$

and the resulting closed-loop is stable, with the associated minimum average power being $a^2 - 1$. Therefore, the cheap control problem and expensive control problem in infinite horizon can be reformulated as one another

for (A.4), and this holds similarly for their higher-order LTI counterparts (proofs skipped for brevity). Thus, a communication problem may be connected to a cheap control problem (as done in the present paper) or an expensive control problem (as done in [4]).

It should be pointed out that the above connection between the cheap control and expensive control is shown for the case where the controller gains are static, i.e., the closed-loop systems are LTI. The cheap control and expensive control can also be studied if the controller gains are allowed to be time-varying, i.e., if l_t at time t needs to be solved in (A.5) and if c_t needs to be solved in (A.6), for every t. For the cheap control problem, notice the following iteration

$$
\begin{aligned}
\text{(A.9)} \quad \mathbf{E}(\tilde{x}_t)^2 &= (a - l_t c)^2 \mathbf{E}(\tilde{x}_{t-1})^2 + (l_t)^2 \\
&= (c^2 \mathbf{E}(\tilde{x}_{t-1})^2 + 1)(l_t)^2 - 2ac\mathbf{E}(\tilde{x}_{t-1})^2 l_t + a^2 \mathbf{E}(\tilde{x}_{t-1})^2;
\end{aligned}
$$

therefore, to minimize the average power of u, one needs to minimize $\mathbf{E}(\tilde{x}_t)^2$ at each t by choosing l_t (independently from the choices of l_k, $k \neq t$). By completion of squares (as done in [29] for the Kalman filtering problem), one can solve for the optimal l_t:

$$
\text{(A.10)} \qquad l_t^* = \frac{ac\mathbf{E}(\tilde{x}_{t-1})^2}{c^2\mathbf{E}(\tilde{x}_{t-1})^2 + 1}.
$$

Then the convergence to the LTI cheap control solution can be easily shown, similar to the Kalman filtering problem. On the other hand, the time-varying expensive control problem is a minimum variance linear quadratic Gaussian (LQG) problem, which can be solved using the dynamic programming (details skipped for brevity), and the solution converges to the LTI expensive control solution.

A.3. Corresponding relation between the CP scheme and coding scheme $\mathbb{S}$

i) For any given admissible $\mathbb{T}_{T,n}$ in $\mathbb{S}$, the channel input is equal to

$$
\text{(A.11)} \qquad \underline{u}^T(\mathbb{T}_{T,n}) = \underline{r}^T + \boldsymbol{\mathcal{G}}_T \underline{y}^T = \boldsymbol{\Gamma}_T(\mathbb{T}_{T,n})\underline{M} + \boldsymbol{\mathcal{G}}_T \underline{y}^T.
$$

To find the $(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$ in the CP scheme that can generate the same channel input, let

$$
\text{(A.12)} \qquad
\begin{aligned}
\boldsymbol{\mathcal{B}}_T &:= \boldsymbol{\mathcal{G}}_T \boldsymbol{\mathcal{Z}}_T^{-1}(\boldsymbol{I} - \boldsymbol{\mathcal{G}}_T \boldsymbol{\mathcal{Z}}_T^{-1})^{-1} \\
\underline{v}_T &:= (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)\boldsymbol{\Gamma}_T(\mathbb{T}_{T,n})\underline{M},
\end{aligned}
$$

(i.e. $\boldsymbol{K}_{\underline{v}}^{(T)} := (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)\boldsymbol{\Gamma}_T(\mathbb{T}_{T,n})\boldsymbol{\Gamma}_T(\mathbb{T}_{T,n})'(\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)')$. Then the channel input generated by the CP scheme is

$$(A.13) \qquad \underline{u}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T) = (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\underline{v}^T + (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\boldsymbol{\mathcal{B}}_T\boldsymbol{\mathcal{Z}}_T\underline{y}^T$$
$$= \boldsymbol{\Gamma}_T(\mathbb{T}_{T,n})\underline{M} + \boldsymbol{\mathcal{G}}_T\underline{y}^T = \underline{u}^T(\mathbb{T}_{T,n}).$$

ii) Assume $\boldsymbol{K}_{\underline{v}}^{(T)} > 0$ first. For any fixed $(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$ in the CP scheme, define in $\mathbb{S}$ that

$$\boldsymbol{\mathcal{G}}_T := (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\boldsymbol{\mathcal{B}}_T\boldsymbol{\mathcal{Z}}_T$$

$$(A.14) \qquad \boldsymbol{A} := \boldsymbol{\Gamma}_a^{-1}\left[\begin{array}{c|c} \boldsymbol{0} & \boldsymbol{I}_T \\ \hline * & * \end{array}\right]\boldsymbol{\Gamma}_a := \boldsymbol{\Gamma}_a^{-1}\boldsymbol{A}_a\boldsymbol{\Gamma}_a \quad \in \mathbb{R}^{(T+1)\times(T+1)}$$

$$\underline{C} := \boldsymbol{\Gamma}_a\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}' := \boldsymbol{\Gamma}_a\underline{e}_1,$$

where $\boldsymbol{\Gamma}_a := ((\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\boldsymbol{K}_{\underline{v}}^{(T)}(\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-'})^{\frac{1}{2}} = \boldsymbol{\Gamma}_a'$ is a positive definite square root, $*$ can be any number, and $\boldsymbol{A}_a$ and $\underline{e}_1$ are defined in obvious ways. Then $\boldsymbol{\mathcal{G}}_T$ is strictly lower triangular, and $(\boldsymbol{A}, \underline{C}')$ is observable since its observability matrix is $\boldsymbol{\Gamma}_a$ with full rank. Letting $\underline{M} := (\boldsymbol{\Gamma}_a)^{-1}(\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\underline{v}_T$, then the channel input is

$$(A.15) \quad \underline{u}^T(\mathbb{T}_{T,n}) = \boldsymbol{\Gamma}_T(\mathbb{T}_{T,n})\underline{M} + \boldsymbol{\mathcal{G}}_T\underline{y}^T = \boldsymbol{\Gamma}_a(\boldsymbol{\Gamma}_a)^{-1}(\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\underline{v}_T + \boldsymbol{\mathcal{G}}_T\underline{y}^T$$
$$= (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\underline{v}^T + (\boldsymbol{I} + \boldsymbol{\mathcal{B}}_T)^{-1}\boldsymbol{\mathcal{B}}_T\boldsymbol{\mathcal{Z}}_T\underline{y}^T$$

which is the same as the input generated by $(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$.

Now consider the case that $\boldsymbol{K}_{\underline{v}}^{(T)} \geq 0$ but $\boldsymbol{K}_{\underline{v}}^{(T)} \ngtr 0$. Perform the eigenvalue decomposition, one obtains that $\boldsymbol{K}_{\underline{v}}^{(T)} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$, where without loss of generality, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_q, 0, \ldots, 0)$, $0 < q < T + 1$, and q is the rank of $\boldsymbol{K}_{\underline{v}}^{(T)}$. Define a sequence $\{\boldsymbol{K}_{\underline{v},i}^{(T)}\}_{i=1}^{\infty}$, where $\boldsymbol{K}_{\underline{v},i}^{(T)} := \boldsymbol{U}(\boldsymbol{\Lambda} + \frac{1}{i}\boldsymbol{I} - \frac{1}{i}\frac{T+2}{q}\boldsymbol{I}_q)\boldsymbol{U}'$, $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{I}_q := \text{diag}(1, \ldots, 1, 0, \ldots, 0)$ with rank q. For large enough i, each matrix in the sequence is positive definite, and hence for each pair $(\boldsymbol{K}_{\underline{v},i}^{(T)}, \boldsymbol{\mathcal{B}}_T)$, we can find an admissible triple $\mathbb{T}_{T,T,i}$ corresponding to it per the above construction. It is easily shown that the sequence of triples generate a sequence of inputs that converge to $\underline{u}^T(\boldsymbol{K}_{\underline{r}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$.

iii) By the continuity of the mutual information and power, the power and mutual information generated by $\underline{u}^T(\mathbb{T}_{T,T,i})$ converge to the power and mutual information generated by $\underline{u}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$, respectively. Thus to show iii), it is sufficient to show that $\underline{u}^T(\mathbb{T}_{T,T,i})$ holds the power constraint in Definition 2 if $\underline{u}^T(\boldsymbol{K}_{\underline{v}}^{(T)}, \boldsymbol{\mathcal{B}}_T)$ holds the power constraint in (31). However, this is true as $\text{trace}(\boldsymbol{K}_{\underline{v}_i}^{(T)}) < \text{trace}(\boldsymbol{K}_{\underline{v}}^{(T)})$. Then the result follows.

A.4. Proof of Proposition 5: convergence to steady-state

We show that system (62) converges to the steady-state given by (79). To this aim, we first transform the Riccati recursion into a new coordinate system, then show that it converges to a limit, and finally prove that the limit is the unique stabilizing solution of the Riccati equation. The coordinate transformation is given by

(A.16)
$$\widetilde{\mathbb{A}} := \boldsymbol{\Psi}\mathbb{A}\boldsymbol{\Psi}^{-1} := \begin{bmatrix} \boldsymbol{A} & 0 \\ 0 & \boldsymbol{F} \end{bmatrix},$$

$$\widetilde{\mathbb{C}} := \boldsymbol{\Psi}^{-1\prime}\mathbb{C} = \begin{bmatrix} \underline{C} + \boldsymbol{\phi}'\underline{H} \\ \underline{H} \end{bmatrix}, \quad \widetilde{\mathbb{D}} := \boldsymbol{\Psi}^{-1\prime}\mathbb{D} = \begin{bmatrix} \underline{C} \\ 0 \end{bmatrix},$$

$$\widetilde{\boldsymbol{\Sigma}}_t := \boldsymbol{\Psi}\widetilde{\boldsymbol{\Sigma}}_t\boldsymbol{\Psi}', \quad \widetilde{\boldsymbol{\Sigma}}_0 := \boldsymbol{\Psi}\boldsymbol{\Sigma}_0\boldsymbol{\Psi}' = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{\psi}' \\ -\boldsymbol{\psi} & \boldsymbol{\psi}\boldsymbol{\psi}' \end{bmatrix},$$

where

(A.17)
$$\boldsymbol{\Psi} := \begin{bmatrix} \boldsymbol{I}_{n+1} & \boldsymbol{0} \\ -\boldsymbol{\psi} & \boldsymbol{I}_m \end{bmatrix} \quad \text{(i.e. } \boldsymbol{\Psi}^{-1} = \begin{bmatrix} \boldsymbol{I}_{n+1} & \boldsymbol{0} \\ \boldsymbol{\psi} & \boldsymbol{I}_m \end{bmatrix} \text{)},$$

and $\boldsymbol{\psi}$ is the unique solution to the Sylvester equation $\boldsymbol{F}\boldsymbol{\psi} - \boldsymbol{\psi}\boldsymbol{A} = -\underline{G}\,\underline{C}'$. The existence and uniqueness of $\boldsymbol{\psi}$ are guaranteed by Assumption (**A2**) that $\lambda_i(-\boldsymbol{A}) + \lambda_j(\boldsymbol{F}) \neq 0$ for any i and j.

Assume $k = n$ for the rest of the proof; i.e., $\boldsymbol{A}$ is anti-stable. For the case $k < n$, we can further transform $\widetilde{\mathbb{A}}$, $\widetilde{\mathbb{C}}$, and $\widetilde{\boldsymbol{\Sigma}}$ into $\check{\mathbb{A}}$, $\check{\mathbb{C}}$, and $\check{\boldsymbol{\Sigma}}$ such that $\check{\mathbb{A}} = \text{diag}[\boldsymbol{A}_+, \boldsymbol{A}_-, \boldsymbol{F}]$ where $\boldsymbol{A}_+ \in \mathbb{R}^{(k+1)\times(k+1)}$ is anti-stable and $\boldsymbol{A}_-$ is stable; then the following argument can be easily adopted for the case $k < n$.

Therefore, the convergence of $\boldsymbol{\Sigma}_{t+1} = \text{Ric}(\mathbb{A}, \mathbb{C}, \boldsymbol{\Sigma}_t)$ with initial condition $\boldsymbol{\Sigma}_0$ is equivalent to the convergence of $\widetilde{\boldsymbol{\Sigma}}_{t+1} = \text{Ric}(\widetilde{\mathbb{A}}, \widetilde{\mathbb{C}}, \widetilde{\boldsymbol{\Sigma}}_t)$ with initial condition $\widetilde{\boldsymbol{\Sigma}}_0$. By [42], $\widetilde{\boldsymbol{\Sigma}}_t$ would converge if

(A.18)
$$\left| \left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_m \end{bmatrix} - \widetilde{\boldsymbol{\Sigma}}_0 \begin{bmatrix} \boldsymbol{I}_{n+1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_{22} \end{bmatrix} \right) \right| \neq 0,$$

where $\boldsymbol{X}_{22}$ is the negative semi-definite matrix to the discrete-time Lyapunov equation

(A.19)
$$\boldsymbol{X}_{22} = \boldsymbol{F}\boldsymbol{X}_{22}\boldsymbol{F}' - (\underline{C} + \boldsymbol{\psi}'\underline{H})(\underline{C} + \boldsymbol{\psi}'\underline{H})'.$$

Then direct computation shows that $\widetilde{\boldsymbol{\Sigma}}_t$ indeed converges to a limit $\widetilde{\boldsymbol{\Sigma}}_\infty$, and the limit is a positive semi-definite solution to the Riccati equation

$\widetilde{\boldsymbol{\Sigma}}_\infty = \mathrm{Ric}(\widetilde{\mathbf{A}}, \widetilde{\mathbb{C}}, \widetilde{\boldsymbol{\Sigma}}_\infty)$. By [29], this solution is the unique stabilizing solution because $(\widetilde{\mathbf{A}}, \widetilde{\mathbb{C}}')$ is observable (since $(\mathbf{A}, \mathbb{C}')$ is observable) and $\widetilde{\mathbf{A}}$ does not have any eigenvalues on the unit circle, and it can be computed as $\widetilde{\boldsymbol{\Sigma}}_\infty = \mathrm{diag}(\widetilde{\boldsymbol{\Sigma}}_{11}, \mathbf{0})$ where $\widetilde{\boldsymbol{\Sigma}}_{11}$ is the positive-definite solution to a reduced-order Riccati equation $\widetilde{\boldsymbol{\Sigma}}_{11} = \mathrm{Ric}(\boldsymbol{A}, (\underline{C}' + \underline{H}'\boldsymbol{\psi}), \widetilde{\boldsymbol{\Sigma}}_{11})$ and has rank equal to the number of anti-stable eigenvalues of $\widetilde{\mathbf{A}}$. Thus, $\boldsymbol{\Sigma}_t$ converges to $\boldsymbol{\Sigma}_\infty$ with rank equal to the number of anti-stable eigenvalues of $\widetilde{\mathbf{A}}$. The the results follow.

A.5. Proof of Proposition 7

Consider the coordinate transformations used in the proof of Proposition 5 that transform $\mathbf{A}$, $\mathbb{C}$, and $\boldsymbol{\Sigma}$ into $\widetilde{\boldsymbol{\Sigma}}$ into $\check{\mathbf{A}}$, $\check{\mathbb{C}}$, and $\check{\boldsymbol{\Sigma}}$. Note that the block in $\check{\boldsymbol{\Sigma}}_\infty$ (i.e. the solution to the Riccati equation defined by $\check{\mathbf{A}}$ and $\check{\mathbb{C}}$) associated with the $\boldsymbol{A}_-$ block is zero. By Proposition 6, in the new coordinates the rate and power due to the $\boldsymbol{A}_-$ block are both zero, and hence in the original coordinates the rate and power due to the stable eigenvalues of $\boldsymbol{A}$ are both zero. Then we remove the dimensions corresponding to $\boldsymbol{A}_-$ in $\check{\mathbf{A}}$, $\check{\mathbb{C}}$, and $\check{\boldsymbol{\Sigma}}$. It is easy to check that this leads to a pair of reduced order $(\boldsymbol{A}_k, \underline{C}_k)$ with $\boldsymbol{A}_k$ anti-stable and satisfying $R_{\infty,n}(\boldsymbol{A}, \underline{C}) = R_{\infty,k}(\boldsymbol{A}_k, \underline{C}_k)$ and $P_{\infty,n}(\boldsymbol{A}, \underline{C}) = P_{\infty,k}(\boldsymbol{A}_k, \underline{C}_k)$.

Acknowledgments

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**(2):172–182, Apr. 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**(2):183–189, Apr. 1966.

[3] J. K. Omura, *Optimum linear transmission of analog data for channels with feedback*, IEEE Trans. Inform. Theory, **14**(1):38–43, Jan. 1968.

[4] N. Elia, *When Bode meets Shannon: Control-oriented feedback communication schemes.* IEEE Trans. Automat. Contr., **49**(9):1477–1488, Sept. 2004.

[5] A. Sahai and S. Mitter, *The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link: Parts I: Scalar systems.* IEEE Trans. Inform. Theory, **52**(8):3369–3395, Aug. 2006.

[6] S. Tatikonda and S. Mitter, *The capacity of channels with feedback.* IEEE Trans. Inform. Theory, **55**(1):323–349, Jan. 2009.

[7] X. Feng, K. A. Loparo and Y. Fang, *Optimal state estimation for stochastic systems: An information theoretic approach.* IEEE Trans. Automat. Contr., **42**(6):771–785, June 1997.

[8] M. M. Seron, J. H. Braslavsky and G. C. Goodwin, *Fundamental Limitations in Filtering and Control.* Springer, 1997.

[9] G. D. Forney, *On the role of MMSE estimation in approaching the information theoretic limits of linear Gaussian channels: Shannon meets Wiener.* Proc. 41st Annual Allerton Conference on Communication, Control, and Computing, Oct. 2003.

[10] G. Zang and P. A. Iglesias, *Nonlinear extension of Bode's integral based on an information theoretic interpretation.* Syst. Control Lett, **50**:11–19, 2003.

[11] D. Guo, S. Shamai and S. Verdu, *Mutual information and minimum mean-square error in Gaussian channels.* IEEE Trans. Inform. Theory, **51**(4):1261–1282, Apr. 2005.

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

[13] N. C. Martins and M. A. Dahleh, *Feedback control in the presence of noisy channels: Bode-like fundamental limitations of performance.* IEEE Trans. Automat. Contr., **53**(7):1604–1615, Aug. 2008.

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

[15] L. H. Ozarow, Random coding for additive Gaussian channels with feedback. *IEEE Trans. Inform. Theory*, **36**(1):17–22, Jan. 1988.

[16] S. Ihara, *Information Theory for Continuous Systems.* World Scientific, 1993.

[17] O. Shayevitz and M. Feder, *Optimal feedback communication via posterior matching.* IEEE Trans. Inform. Theory, **57**(3):1186–1222, March 2011.

[18] Y. Kim, *Feedback capacity of stationary Gaussian channels.* IEEE Trans. Inform. Theory, **56**(1):57–85, Jan. 2010.

[19] J. Massey, *Causality, feedback, and directed information.* Proc. IEEE International Symposium on Information Theory and Applications (ISITA), pages 303–305, 1990.

[20] E. Ordentlich, *A class of optimal coding schemes for moving average additive Gaussian noise channels with feedback.* Proc. IEEE International Symposium on Information Theory (ISIT), page 467, June 1994.

[21] Y. Kim, *Feedback capacity of the first-order moving average Gaussian channel.* IEEE Trans. Inform. Theory, **52**(7):3063–3079, July 2006.

[22] S. Yang, A. Kavcic and S. Tatikonda, *On the feedback capacity of power-constrained Gaussian noise channels with memory.* IEEE Trans. Inform. Theory, **53**(3):929–954, Mar. 2007.

[23] A. El Gamal and Y. Kim, *Network Information Theory.* Cambridge University Press, 2011.

[24] J. Liu, N. Elia and S. Tatikonda, *Capacity-achieving feedback scheme for Markov channels with channel state information.* Submitted to IEEE Trans. Inform. Theory, Aug. 2005.

[25] S. Tatikonda and S. Mitter, *Control under communication constraints.* IEEE Trans. Automat. Contr., **49**(7):1056–1068, July 2004.

[26] R. Middleton, A. Rojas, J. Freudenberg and J. Braslavsky, *Feedback stabilization over a first-order moving average Gaussian noise channel.* IEEE Trans. Automat. Contr., **54**(1):163–167, 2009.

[27] E.Ardestanizadeh, P. Minero and M. Franceschetti, *LQG control approach to Gaussian broadcast channels with feedback.* IEEE Trans. Inform. Theory, **58**(8):5267–5278, Aug. 2012.

[28] E.Ardestanizadeh and M. Franceschetti, *Control-theoretic approach to communication with feedback.* IEEE Trans. Automat. Contr., **57**(10):2576–2587, Oct. 2012.

[29] T. Kailath, A. Sayed and B. Hassibi, *Linear Estimation.* Prentice Hall, 2000, Englewood Cliffs, NJ, 2000.

[30] S. K. Mitter and N. J. Newton, *A variational approach to nonlinear estimation.* SIAM Journal on Control and Optimization, **42**:1813–1833, 2003.

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

[32] J. Liu and N. Elia, *Writing on dirty paper with feedback.* Communications in Information and Systems, **5**(4):401–422, 2005.

[33] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley and Sons, 1968.

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

[35] C. T. Chen, *Linear Systems Theory and Design.* Oxford University Press, 1999.

[36] H. L. Weidemann and E. B. Stear, *Entropy analysis of feedback control systems.* IEEE Trans. Inform. Theory, **16**(3):264–270, May 1970.

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

[38] S. K. Mitter and N. J. Newton, *The duality between estimation and control.* In E. Rofman J. L. Menaldi and A. Sulem, editors, *Optimal Control and Partial Differential Equations.* 2000.

[39] P. A. Iglesias, *Tradeoffs in linear time-varying systems: an analogue of Bode's sensitivity integral.* Automatica, **37**:1541–1550, 2001.

[40] J. P. M. Schalkwijk, *Center of gravity information feedback.* IEEE Trans. Inform. Theory, **IT-14**:324–331, 1968.

[41] J. Liu, *Fundamental Limits in Gaussian Channels with Feedback: Confluence of Communication, Estimation, and Control.* PhD thesis, Iowa State University, Ames, IA, Apr. 2006.

[42] K. Gallivan, X. Rao and P. Van Dooren, *Singular Riccati equations stabilizing large-scale systems.* Lin. Alg. Appl., **415**:359–372, 2006.

Huawei, Rolling Meadows, IL 60008, USA
E-mail address: `jialing.liu@huawei.com`

Department of Electrical and Computer Engineering
Iowa State University, Ames, IA 50011, USA
E-mail address: `nelia@iastate.edu`