Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Shannon Zero Error Capacity and the Problem of Almost Sure Observability over Noisy Communication Channels

Alexey S. Matveev and Andrey V. Savkin

Abstract— **The paper addresses the state estimation problem involving communication errors and capacity constraints. Discrete-time partially observed unstable linear systems perturbed by uniformly bounded disturbances are studied. Unlike the classic theory, the sensor signals are transmitted to the estimator over a noisy digital communication link modelled as a stochastic stationary discrete memoryless channel. It is shown that the Shannon zero error capacity of the channel constitutes the border separating the cases where the plant is and respectively, is not almost surely observable.**

I. Introduction

In classical control theory, a common assumption is that data transmission between various components of the system can be performed with arbitrary high accuracy. However in many modern engineering applications, observation and control signals are sent over digital finite capacity channels. Examples concern complex dynamical processes like advanced aircraft, spacecraft, automotive, industrial and defence systems, arrays of microactuators, and power control in mobile communication. Other examples arise when a large number of mobile units is remotely controlled by a single decision maker. Bandwidths communication constraints are often major obstacles to control system design by means of the classical theory. As was shown in [22], design of control systems for platoons of underwater vehicles strongly highlights the need for control strategies that address explicitly the bandwidth limitation on communication between vehicles, which is severely restricted underwater.

The issues of stabilization and observation via limited capacity channels enjoyed much attention recently, see e.g., [1], [4], [6]–[9], [12], [14], [15], [17], [21] and the literature therein. Fundamental limitations imposed by the available communication data rate on the achievable control performance were studied e.g., in [12], [14], [15], [17], [18]. The tightest data rate bounds above which stabilization/observation of a linear plant is possible were established in [14], [15], where the focus was on the channel quantization effects, and noiseless channels with alphabets containing a finite number N of elements were considered. The stabilizability/observability criterion was given in the form $\eta < \log_2 N$, where η is the sum of the logarithms of the absolute values of the system unstable eigenvalues and $\log_2 N$ is the channel capacity.

This work was supported by the Australian Research Council

A. Matveev is with the Department of Mathematics and Mechanics, Saint Petersburg University, Universitetskii 28, Petrodvoretz, St.Petersburg, 198504, Russia almat@am1540.spb.edu

A. Savkin is with the School of Electrical Engineering and Telecommunications, The University of New South Wales and National ICT Australia Ltd., Sydney, 2052, Australia a.savkin@unsw.edu.au

Study of the interaction between control and data rate limitations of noiseless channels is a necessary step in developing the theory. However typical communication channels are noisy. Mean-square and more generally, mth moment) observability/stabilizability bounds for noisy discrete channels were addressed in [16], [21] for scalar noisy linear systems. An encoder-decoder pair for estimating the state of a scalar noisy linear system via a noisy binary symmetric channel with a perfect feedback was proposed in [21]. Conditions ensuring that the mathematical expectation of the estimation error is bounded were obtained. It was shown that these conditions improve those from [16].

Almost sure observability/stabilizability of noiseless multidimensional plants over noisy channels was addressed in [12]. It was shown that the border between the cases where such an observability/stabilizability holds and does not hold, respectively, is given by the Shannon ordinary capacity $\mathfrak{c}$ of the channel [20]. The critical feature of these results from [12] is that they concern noiseless (unperturbed) plants.

In this paper, the issue of almost sure observability is addressed for noisy both plants and channels. We consider a linear discrete-time unstable partially observed plants affected by exogenous disturbances. The sensor signals are sent to the state estimator over a finite alphabet stationary discrete memoryless channel (**DMC**). We show that taking into account the plant disturbances drastically changes the observability domain, as compared with noiseless plants. For noisy plants, the border between the cases where the system can and, respectively, cannot be observed with almost surely bounded error is constituted by the Shannon not ordinary $\mathfrak{c}$ but zero error capacity $\mathfrak{c}_0$ [19] of the channel. The conditions $\eta \leq \mathfrak{c}_0$ and $\eta < \mathfrak{c}_0$ are necessary and sufficient, respectively. As is known, $\mathfrak{c} \geq \mathfrak{c}_0$ and for many channels, $\mathfrak{c} > \mathfrak{c}_0$. For them, the observability domain does change. Moreover, we prove that an unstable linear system affected by arbitrarily and uniformly small disturbances can never be observed via **DMC** if $\mathfrak{c}_0 < \eta$. Then the estimation error is unbounded almost surely, irrespective of which algorithm of observation is employed. A similar in spirit negative fact was established for a simple scalar system in [16], where however only a special and rather small class of estimation algorithms was examined. They are confined to those employing static block encoders and decoders producing only finitely many outputs. On contrary, our results deal with all time-varying non-anticipating deterministic algorithms of estimation. In the particular case of noiseless channel, our results are in harmony with those from [14], [15]. Some further details relevant to the results of the paper can be found in [13].

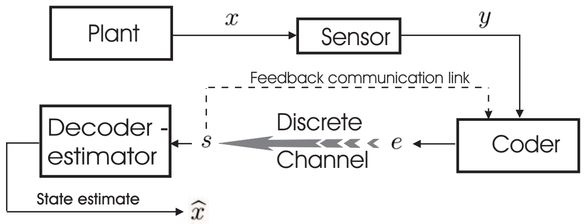

Fig. 1. Estimation via a communication channel

Recent references that describe related developments include [1]–[4], [6]–[9].

The body of the paper is organized as follows. In Section II, we pose the estimation problem. Section III recalls the notion of the zero error capacity. Section IV presents main results, whereas Section V is devoted to their proofs. Section VI offers brief conclusions. There is an Appendix containing the proof of a technical fact.

II. Estimation problem

We study unstable discrete-time linear systems of the form:

$$x(t+1) = Ax(t) + \zeta(t);\ x(0) = x_0,\ y(t) = Cx(t). \quad (1)$$

Here $x \in \mathbb{R}^n$ is the state, $\zeta(t) \in \mathbb{R}^n$ is an exogenous disturbance, and $y \in \mathbb{R}^l$ is the measured output. The instability means that there is an eigenvalue λ of A with $|\lambda| \geq 1$. The initial state x_0 is random. The objective is to estimate the current state on the basis of the prior measurements.

This estimate is required at a remote location. The only way to communicate information from the sensor to this location is via a given noisy discrete channel. So to be transmitted, measurements must be first translated into a sequence of symbols e from the finite *input alphabet* $\mathcal{E}$ of the channel. This is done by a special device, referred to as *coder*. Its outputs e emitted into the channel are transformed by a noise into a sequence of channel's outputs s from a finite *output alphabet* $\mathcal{S}$. The *decoder-estimator* produces an estimate $\widehat{x}$ of the current state x on the basis of the prior messages s:

$$\widehat{x}(t) = \mathfrak{X}\left[t, s(0), s(1), \dots, s(t)\right]. \quad (2)$$

We consider two classes of coders. The first of them serves the case where there is a *feedback communication link*. Via this link, the transmission result $s(t)$ becomes known at the coder site by the time $t+1$ (see Fig. 1). The second class serves the case where no feedback is available. The coders from these classes are said to be *with* and *without a feedback* and are given by the following equations, respectively:

$$e(t) = \mathfrak{E}[t, y(0), \dots, y(t), s(0), \dots, s(t-1)] \in \mathcal{E}, \quad (3)$$

$$e(t) = \mathfrak{E}[t, y(0), \dots, y(t)] \in \mathcal{E}. \quad (4)$$

Definition 1: The coder and decoder-estimator are said to *track the state with a bounded error* if

$$\overline{\lim_{t\to\infty}} |x(t) - \widehat{x}(t)| < \infty.$$

Is it possible to construct such coder and decoder?

An answer will be given under the following assumptions.

Assumption 1: The communication channel is a stationary **DMC**: given a current input $e(t)$, the output $s(t)$ is independent of all other inputs and outputs $e(j), s(j), j \neq t$, and the conditional probability $W(s|e) := \boldsymbol{P}[s(t) = s | e(t) = e], s \in \mathcal{S}, e \in \mathcal{E}$ does not depend on time t.

This model incorporates the effect of message loss by including a special "void" symbol $\emptyset$ in the output alphabet $\mathcal{S}$: $s(t) = \emptyset \Leftrightarrow$ the message $e(t)$ is lost.

Assumption 2: The plant does not affect the channel: given an input $e(t)$, the output $s(t)$ is independent of x_0.

Assumption 3: The pair (A, C) is detectable.

We consider two classes of disturbances $\zeta(t)$ in (1). The first class consists of all deterministic bounded disturbances

$$|\zeta(t)| \leq D \qquad \forall t. \quad (5)$$

The second one interprets $\zeta(t)$ as a stochastic process. These classes differ if only since sample sequences of such a process do not necessarily constitute a set described by (5). Note that our result concerning deterministic disturbances (Theorem 1) can be formally derived from that dealing with stochastic ones (Theorem 3). We however offer an independent proof of the first result since it permits us to display the basic arguments in a more straightforward and clear fashion than in the case of stochastic disturbances.

III. Zero error capacity of the channel

The Shannon ordinary capacity of the channel is the least upper bound of rates at which information can be transmitted with as small probability of error as desired [20]. The zero error capacity is the least upper bound of rates at which it is possible to transmit information with zero probability of error. Unlike the former, the latter may depend on whether the communication feedback is available or not [19].

Channels without a feedback link. A *block code* [20] with block length m is a finite number N of the channel input code words $E^1, \dots, E^N \in \mathcal{E}^m = \{E\}, E^\nu = (e_0^\nu, \dots, e_{m-1}^\nu)$. This code is used to notify the receiver which choice of N possibilities, labeled by ν, is taken by the transmitter. The ratio $R := \frac{\log_2 N}{m}$ is the *rate* of the code. The *decoding rule* is a method to associate a unique ν with any output word of length m, that is a map $\mathfrak{D} : \mathcal{S}^m \to [1 : N]$. Such a rule is *errorless* if $\mathfrak{D}(S) = \nu$ for any ν and any output word S that occurs with a positive probability given that E^ν is sent over the channel. The *zero error capacity* $\mathfrak{c}_0 := \sup R$, where sup is over all block codes of arbitrary lengths m for which errorless decoding is possible. Two input letters e_1, e_2 are said to be *adjacent* if they may cause a common output letter s, i.e., $\exists s : W(s|e_1)W(s|e_2) > 0$, where $W(s|e)$ are the channel transition probabilities. The zero error capacity is positive $\mathfrak{c}_0 > 0$ if and only if there exists a couple of non-adjacent letters [19]. The general formula for $\mathfrak{c}_0$ is still missing [5].

To pave the way to channels with feedback, note that in the absence of feedback, encoding by block codes is equivalent to encoding ν via *block functions*, i.e., rules of the form

$$e(t) = \mathfrak{E}_*[t, e(0), \dots, e(t-1), \nu], \quad t = 1, \dots, m-1, \qquad e(0) = \mathfrak{E}_*[0, \nu]. \quad (6)$$

Channels with a complete feedback. In this case, the block function (6) with block length m takes the form [19]

$$e(t) = \mathfrak{E}_*[t, e(0), \dots, e(t-1), s(0), \dots, s(t-1), \nu], \quad t = 1, \dots, m-1, \qquad e(0) = \mathfrak{E}_*[0, \nu]. \quad (7)$$

It is still used to encode messages labeled by ν for transmission over the channel. The other particulars in the definition of the zero error capacity remain unchanged; the corresponding capacity is denoted by $\mathfrak{c}_{0F}$. The zero error capacity may, in some cases, be grater with feedback than without [19]. It is known that $\mathfrak{c}_{0F} = 0$ if all pairs of input letters are adjacent. Otherwise, $2^{-\mathfrak{c}_{0F}} = \min \max_{s \in \mathcal{S}} \sum_{e \in \mathcal{E}_s} \boldsymbol{P}(e)$ [19]. Here $\min$ is over all probability distributions on the input channel alphabet $\mathcal{E}$, and $\mathcal{E}_s$ is the set of all input symbols that cause the output letter s with a nonzero probability.

IV. Main results

A. Deterministic disturbances

Theorem 1: Let Assumptions 1 and 2 hold and the noise do occur in the plant: $D > 0$ in (5). Denote by $\lambda_1, \dots, \lambda_n$ the eigenvalues of the system (1) repeating in accordance with their algebraic multiplicities, and put

$$\eta(A) := \sum_{\lambda_j : |\lambda_j| \geq 1} \log_2 |\lambda_j|.$$

If there exists a coder-decoder pair without (with) a feedback that with a nonzero probability, tracks the state with a bounded error for any disturbance satisfying (5), then $\mathfrak{c}_0 \geq \eta(A)$ (respectively, $\mathfrak{c}_{0F} \geq \eta(A)$).

The proof of this theorem is given in Section V.

The conclusion of the theorem holds for any $D > 0$. So the level of the plant noise may be arbitrarily small.

The zero error capacity of many channels is 0. Theorem 1 implies that for them and strictly unstable plants $\eta(A) > 0$, the estimation error is almost surely unbounded, even if the disturbance is arbitrarily uniformly small and irrespective of which estimation scheme is employed. This in particular holds for the binary symmetric channel with cross-over probability $0 < p < 1$ and erasure channel with arbitrary alphabet (of size ≥ 2) and positive erasure probability.

The next theorem demonstrates that the necessary conditions given by Theorem 1 are "almost" sufficient.

Theorem 2: Suppose that Assumptions 1—3 hold and the noise bound D from (5) is known.

If $\mathfrak{c}_0 > \eta(A)$ ($\mathfrak{c}_{0F} > \eta(A)$), then there exists a coder-decoder pair without (with) a feedback that with probability 1, tracks the state with uniformly bounded error:

$$\overline{\lim_{t \to \infty}} \sup_{\{\zeta(\cdot)\}} |x(t) - \widehat{x}(t)| < \infty, \quad (8)$$

where $\sup$ is over all disturbances satisfying (5).

In fact, this theorem is known since by increasing the sampling period, this case is reduced to that of the noiseless channel with capacity $c > \eta(A)$.

B. Stochastic disturbances

Now we assume that in (1), $\{\zeta(t)\}$ is a stochastic process satisfying the following assumptions.

Assumption 4: The random vectors $\zeta(t)$ are identically distributed according to a probability density $p(\zeta)$, mutually independent, and independent of x_0.

The major points of this paper concern the case where the disturbances $\zeta(t)$ are uniformly and arbitrarily small: the support of $p(\zeta)$ is a subset of a small ball centered at 0.

Assumption 5: The system (1) does not affect the channel: given an input $e(t)$, the output $s(t)$ is independent of not only the initial state x_0 but also disturbances $\zeta(\theta)$.

Theorem 3: Let Assumptions 1, 2, 4, 5 hold.

If $\eta(A) > \mathfrak{c}_{0F}$ (or $\eta(A) > \mathfrak{c}_0$), then for any coder with a feedback (3) (without a feedback (4)) and decoder-estimator (2), the estimation error is a.s. unbounded:

$$\overline{\lim_{t \to \infty}} |x(t) - \widehat{x}(t)| = \infty \qquad \text{a.s.} \quad (9)$$

The proof of this theorem is available upon request.

V. Proof of Theorem 1

To prove Theorem 1, it suffices to justify the following.

Proposition 1: Let $\eta(A) > \mathfrak{c}_0$ (or $\eta(A) > \mathfrak{c}_{0F}$). Consider any coder and decoder without (with) a feedback. Then with probability 1, there exists an admissible (i.e., satisfying (5)) disturbance $\{\zeta(t)\}$ for which the error is unbounded:

$$\overline{\lim_{t \to \infty}} |x(t) - \widehat{x}(t)| = \infty. \quad (10)$$

We start the proof of this claim by noting that attention can be switched to the uniform observability.

Lemma 1: Suppose that for any coder and decoder without (with) a feedback and initial state distribution satisfying Assumption 2, the error is not kept uniformly bounded:

$$\overline{\lim_{t \to \infty}} \sup_{\{\zeta(\cdot)\}} |x(t) - \widehat{x}(t)| = \infty \qquad \text{a.s.} \quad (11)$$

Then the conclusion of Proposition 1 holds.

The proof of this technical observation is placed in Appendix.

Let the assumptions of Proposition 1 be valid. To prove (11), we shall argue by contradiction. Suppose that (11) is not true, i.e., (8) holds with a positive probability for some coder and decoder. By sacrificing a small probability, the error can be made uniformly bounded: for some nonrandom $t_*, b_* < \infty$, with a nonzero probability,

$$\sup_{\{\zeta(\cdot)\}} |x(t) - \widehat{x}(t)| < b_* \qquad \text{for all} \quad t \geq t_*. \quad (12)$$

The idea is to show that then there is an errorless block code hidden within the observer. We start with preliminaries.

By the assumptions of Proposition 1, $\eta(A) > 0$. So $\sigma_\oplus := \{\lambda \in \sigma(A) : |\lambda| > 1\} \neq \emptyset$, where $\sigma(A)$ is the spectrum of A. Let $L_\oplus$ denote the invariant subspace of A related to $\sigma_\oplus$, and $A_\oplus$ the operator A acting in $L_\oplus$. It is easy to see that

$\eta(A) = \log_2 |\det A_\oplus|$. We also recall that a set $V \subset \mathbb{R}^n$ is called *b-separated* if $|v_1 - v_2| > b$ for any $v_1 \neq v_2 \in V$.

Lemma 2: For any $b > 0$ and $m = 1, 2, \ldots$, there exist

$$N \geq \left(b^{-1} D\right)^{\dim L_\oplus} 2^{(m-2)\eta(A)} \quad (13)$$

admissible disturbances $\Xi_\nu = \{\zeta_\nu(t)\}_{t=0}^{m-2}, \nu = 1, \ldots, N$ that drive the system (1) from the zero initial state to the states $x_\nu(m-1) \overset{\Xi_\nu}{\longleftarrow} 0$ forming a b-separated set.

Proof: We look for Ξ_ν among disturbances of the form: $\zeta_\nu(0) \in L_\oplus, \zeta_\nu(t) = 0$ if $t \geq 1$. For them, $x_\nu(m-1) = A_\oplus^{m-2}\zeta_\nu(0)$. Among the subsets of the ball $B_0^D \subset L_\oplus$ transformed by $A_\oplus^{m-2}$ into b-separated sets, we pick one $Z = \{\zeta_1(0), \ldots, \zeta_N(0)\}$ with the maximal cardinality N. Then $A_\oplus^{m-2} B_0^D \subset \bigcup_{\nu=1}^N B^b_{x_\nu(m-1)}$ (all balls are in $L_\oplus$): otherwise, one more point can be put in Z. Hence

$$|\det A_\oplus|^{m-2} \,\mathbf{mes}\, B_0^D = \mathbf{mes}\left[A_\oplus^{m-2} B_0^D\right] \leq \sum_{\nu=1}^{N} \mathbf{mes}\left[B^b_{\boldsymbol{x}_\nu(m-2)}\right] = N \,\mathbf{mes}\, B_0^b,$$

$$N \geq |\det A_\oplus|^{m-2} \frac{\mathbf{mes}\, B_0^D}{\mathbf{mes}\, B_0^b} = 2^{(m-2)\eta(A)} \left(\frac{D}{b}\right)^{\dim L_\oplus} . \bullet$$

To construct an errorless code with block length $m \geq 2$, we take t_*, b_* from (12) and $\Xi_1, \ldots, \Xi_N$ from Lemma 2, where we use $b > 2b_*$. At first, we introduce infinitely many random codes. Then we pick a unique deterministic code as a realization of one of them. To this end, for any $i \geq 1$, we consider the process $x(t), y(t), \widehat{x}(t), e(t), s(t)$ generated by the zero disturbance until $t = t_*^i - 1, t_*^i := t_* + im$. Then we introduce N continuations of this disturbance on $[t_*^i, t_*^{i+1} - 2]$ by $\Xi_1, \ldots, \Xi_N$, respectively. Each of them Ξ_ν gives rise to a continuation of the process on the interval $[t_*^i, t_*^{i+1} - 1]$; this continuation is marked by ${}^i_\nu$. Now we introduce the block code (6) (or (7)) that encodes ν by acting just as the coder (4) (respectively, (3)) does under the continuation Ξ_ν:

$$\mathfrak{E}_*^i\Big[t, e_0, \ldots, e_{t-1}, \boxed{s_0, \ldots, s_{t-1},}\ \nu\Big] := \mathfrak{E}\Big[t + t_*^i, y(0), \ldots, y(t_*^i), y_\nu^i(t_*^i + 1), \ldots, y_\nu^i(t_*^i + t), \boxed{s(0), \ldots, s(t_*^i - 1), s_0, \ldots, s_{t-1}}\Big], t \leq m - 1. \quad (14)$$

(For $t = 0$, all arguments of the form $s_\theta, e_\theta, y_\nu^i(\theta)$ are dropped.) The dashed expressions are omitted if there is no communication feedback. The decoding rule $\mathfrak{D}^i : \mathcal{S}^m \to [1 : N]$ is fabricated on the basis of the decoder (2):

$$\mathfrak{D}^i[s_0, \ldots, s_{m-1}] := \begin{cases} \nu & \text{if } \mathfrak{A} \text{ is true} \\ 1 & \text{otherwise} \end{cases}, \text{ where}$$

$$\mathfrak{A} \equiv \Big\{\text{the ball in } \mathbb{R}^n \text{ with the radius } b_* \text{ centered at } \mathfrak{X}\big[\theta_i, \{s(t)\}_{t=0}^{t_*^i - 1}, \{s_j\}_{j=0}^{m-1}\big] \text{ contains } x_\nu^i(\theta_i) \text{ and does not contain } x_{\nu'}^i(\theta_i) \text{ with any } \nu' \neq \nu\Big\}. \quad (15)$$

where $\theta_i := t_*^{i+1} - 1$. Thus a random coding-decoding pair $\mathfrak{P}_{cd}^i := [\mathfrak{E}_*^i(\cdot), \mathfrak{D}^i(\cdot)]$ with block length m is constructed.

Further it will be convenient to think about the channel as a sequence of mutually independent (and independent of x_0) and identically distributed random maps $G_t : \mathcal{E} \to \mathcal{S}$ such that $\boldsymbol{P}[G_t(e) = s] = W(s|e)\ \forall t, e, s$, where $W(\cdot|\cdot)$ is taken from Assumption 1. Then $s(t) = G_t[e(t)]$. We also introduce the block maps $G_i^m(E) := \left\{G_{t_*^i + t}(e_t)\right\}_{t=0}^{m-1}$ acting on the input words $E = \{e_t\}_{t=0}^{m-1}$. When one uses a coding-decoding pair $\mathfrak{P}_{cd} = [\mathfrak{E}_*(\cdot), \mathfrak{D}(\cdot)]$ with block length m to transmit a message ν during the interval $[t_*^i : t_*^{i+1} - 1]$, the result ν_{tr} depends on both this pair and G_i^m. With a slight abuse of notation, we write this $\nu_{tr} = \mathfrak{P}_{cd}[\nu, G_i^m]$.

Now we are in a position to state the key property of $\mathfrak{P}_{cd}^i$.

Lemma 3: Whenever (12) holds,

$$\nu = \mathfrak{P}_{cd}^i[\nu, G_i^m] \quad \forall \nu \text{ and } i. \quad (16)$$

Proof: By (14), $\mathfrak{E}_*^i$ encodes ν into $\{e_\nu^i(t_*^i + t)\}_{t=0}^{m-1}$, and

$$G_i^m\left[\{e_\nu^i(t_*^i + t)\}_{t=0}^{m-1}\right] = \{s_\nu^i(t_*^i + t)\}_{t=0}^{m-1} \overset{\text{def}}{=} S_\nu^i,$$
$$\widehat{x}_\nu^i(\theta_i) = \mathfrak{X}\left[\theta_i, \{s(t)\}_{t=0}^{t_*^i - 1}, S_\nu^i\right], \theta_i := t_*^{i+1} - 1. \quad (17)$$

Since the state at $t = t_*^i$ is common for all ν, the set $\{x_\nu^i(\theta_i)\}_{\nu=1}^N$ is a displacement of the set $\{x_\nu(m-1)\}_{\nu=1}^N$ from Lemma 2. Since the latter is b-separated, so is the former. It follows that any ball of radius $b_* < b/2$ contains no more than one point of the form $x_{\nu'}^i(\theta_i)$. At the same time, (12) ensures that the ball centered at $\widehat{x}_\nu^i(\theta_i)$ contains $x_\nu^i(\theta_i)$. This and (15), (17) imply (16). $\bullet$

By Lemma 3, the decoding rule (15) does not make error for a particular realization of the random channel map G^m. At the same time, a rule is errorless if it does not make errors for every realization assumed with a nonzero probability. Now we are going to show that the variety of samples assumed by the random codes $\mathfrak{P}_{cd}^i, i = 1, 2, \ldots$ contains an errorless pair.

We note first that there are only finitely many coding-decoding pairs with a given block length m and the number N of messages. So in any sample sequence of the stochastic process $\{\mathfrak{P}_{cd}^i\}$, some particular pair is encountered infinitely many times. Moreover, let us observe all such sequences corresponding to elementary events for which (12) holds. By the same argument, there is a pair $\mathfrak{P}_{cd}$ such that with a nonzero probability, both (12) is true and $\mathfrak{P}_{cd}$ is encountered in $\{\mathfrak{P}_{cd}^i\}_{i=1}^\infty$ infinitely many times.

Lemma 4: The pair $\mathfrak{P}_{cd}$ is errorless.

To prove this lemma, we need the following fact.

Theorem 4 ([11, §32, p.53]): Suppose that $\mathcal{F}_i$ is a flow of non decreasing σ-algebras in a probability space, the random variable ξ_i is $\mathcal{F}_i$-measurable, and $b_i \uparrow \infty, b_i > 0$. Suppose also that $\boldsymbol{E}\big|\xi_i - \boldsymbol{E}(\xi_i|\mathcal{F}_{i-1})\big| < \infty$ and

$$\sum_{i=1}^\infty b_i^{-2} \boldsymbol{E}\Big\{\big[\xi_i - \boldsymbol{E}(\xi_i|\mathcal{F}_{i-1})\big]^2\Big\} < \infty. \quad (18)$$

Then with probability 1,

$$b_r^{-1}\sum_{i=1}^{r}\left[\xi_i-\boldsymbol{E}(\xi_i|\mathcal{F}_{i-1})\right]\xrightarrow{r\to\infty}0. \qquad (19)$$

Proof of Lemma 4: We put $i_0 := 0$ and $i_{k+1} := \min\{i : i > i_k, \mathfrak{P}_{cd}^i = \mathfrak{P}_{cd}\}$ for $i = 0,1,\ldots$, where $\min\emptyset \overset{\text{def}}{=} +\infty$. It is easy to see that $\{i_k = l\} \in \mathfrak{F}_l\ \forall l < \infty$, where $\mathfrak{F}_i$ is the σ-algebra generated by $x_0, G_0, \ldots, G_{t_*^i-1}$. Thus i_k is a Markov time [10] with respect to the filtration $\mathfrak{F}_i$. By the choice of $\mathfrak{P}_{cd}$ and Lemma 3 (where $i := i_k$), we see that

$$i_k<\infty \quad \text{and} \quad \nu=\mathfrak{P}_{cd}[\nu,G_{i_k}^m] \qquad \forall\nu,k \qquad (20)$$

with a nonzero probability. So it suffices to show that for almost all events where (20) holds, the sequence $\{G_{i_k}^m\}$ runs over all realizations $\boldsymbol{G}_*^m$ of the random map G^m assumed with nonzero probabilities.

For such a realization $\boldsymbol{G}_*^m$, we apply Theorem 4 to

$$b_k := k, \mathcal{F}_k := \mathfrak{F}_{i_k+1}, \xi_k := I_{\{G_{i_k}^m=\boldsymbol{G}_*^m\}\wedge\{i_k<\infty\}}. \qquad (21)$$

Here I_E is the indicator of the random event E and the σ-algebra $\mathfrak{F}_{i_k+1}$ consists of all random events S such that $S\wedge\{\tau_k = l\} \in \mathfrak{F}_{l+1}$ for all $l < \infty$. As is known, $\mathcal{F}_k \subset \mathcal{F}_{k+1}$ since $i_k\uparrow$ are Markov times [10, Lemma 1.5, Ch.1]. Note that

$$\xi_k=\sum_{l=1}^{\infty} I_{\{G_l^m=\boldsymbol{G}_*^m\}\wedge\{i_k=l\}}=\chi_{i_k}, \qquad (22)$$

where $\chi_l := I_{\{G_l^m=\boldsymbol{G}_*^m\}}$ if $l < \infty$ and $\chi_\infty := 0$. Since χ_l is $\mathfrak{F}_{l+1}$-measurable for $l < \infty$, the variable ξ_k is $(\mathfrak{F}_{i_k+1} = \mathcal{F}_k)$-measurable [10, Lemma 1.8, Ch.1]. Since $0 \le \xi_k \le 1 \Rightarrow$ (19), Theorem 4 ensures (19), where

$$\begin{aligned}
\boldsymbol{E}[\xi_k|\mathcal{F}_{k-1}] &= \boldsymbol{E}[\xi_k|\mathfrak{F}_{i_{k-1}+1}] \overset{\text{Lemma 1.9 [10, Ch.1]}}{=\!=\!=} \sum_{j=1}^{\infty}\boldsymbol{E}[\xi_k|\mathfrak{F}_{j+1}]I_{i_{k-1}=j}+\xi_k I_{i_{k-1}=\infty}\\
&\overset{\text{a)}}{=} \sum_{j=1}^{\infty}\boldsymbol{E}[\xi_k|\mathfrak{F}_{j+1}]I_{i_{k-1}=j}\\
&\overset{(22)}{=} \sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\boldsymbol{E}[I_{G_l^m=\boldsymbol{G}_*^m\wedge i_k=l}|\mathfrak{F}_{j+1}]I_{i_{k-1}=j}\\
&\overset{\text{b)}}{=} \sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\boldsymbol{E}[I_{G_l^m=\boldsymbol{G}_*^m\wedge i_k=l}\times I_{i_{k-1}=j}|\mathfrak{F}_{j+1}]\\
&\overset{i_k>i_{k-1}}{=\!=\!=} \sum_{j=1}^{\infty}\sum_{l=j+1}^{\infty}\boldsymbol{P}[G_l^m=\boldsymbol{G}_*^m\wedge i_k=l|\mathfrak{F}_{j+1}]I_{i_{k-1}=j}\\
&\overset{\text{c)}}{=} \sum_{j=1}^{\infty}\sum_{l=j+1}^{\infty}\underbrace{\boldsymbol{P}[G_l^m=\boldsymbol{G}_*^m]}_{\alpha(\boldsymbol{G}_*^m)}\boldsymbol{P}[i_k=l|\mathfrak{F}_{j+1}]I_{i_{k-1}=j}\\
&\overset{i_k>i_{k-1}}{=\!=\!=} \alpha(\boldsymbol{G}_*^m)\sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\boldsymbol{P}[i_k=l|\mathfrak{F}_{j+1}]I_{i_{k-1}=j}\\
&= \alpha(\boldsymbol{G}_*^m)\Bigg\{\sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\boldsymbol{P}[i_k=l|\mathfrak{F}_{j+1}]I_{i_{k-1}=j} + \underbrace{I_{i_k<\infty}I_{i_{k-1}=\infty}}_{=0}\Bigg\}\\
&\overset{\text{Lemma 1.9 [10, Ch.1]}}{=\!=\!=} \alpha(\boldsymbol{G}_*^m)\boldsymbol{P}\big[i_k<\infty\big|\mathfrak{F}_{i_{k-1}+1}\big]\\
&= \alpha(\boldsymbol{G}_*^m)\boldsymbol{P}\big[i_k<\infty\big|\mathcal{F}_{k-1}\big].
\end{aligned}$$

Here a) holds since $i_{k-1} = \infty \Rightarrow i_k = \infty \overset{(21)}{\Rightarrow} \xi_k = 0$; b) holds since $\{i_{k-1} = j\} \in \mathfrak{F}_j \subset \mathfrak{F}_{j+1}$, and c) holds since the random map G_l^m is independent of $\mathfrak{F}_l$, where $\mathfrak{F}_l \ni \{i_k = l\}$ and $\mathfrak{F}_l \supset \mathfrak{F}_{j+1}$ for $l \ge j+1$. Thus by (19),

$$\frac{1}{r}\sum_{k=1}^{r}\xi_k-\alpha(\boldsymbol{G}_*^m)\frac{1}{r}\sum_{k=1}^{r}\boldsymbol{P}\big[i_k<\infty\big|\mathcal{F}_{k-1}\big]\xrightarrow{r\to\infty}0 \quad \text{a.s.}$$

Applying Theorem 4 to $\mathcal{F}_k$, $b_k := k, \xi_k := I_{i_k<\infty}$ yields

$$\frac{1}{r}\sum_{k=1}^{r}\boldsymbol{P}\big[i_k<\infty\big|\mathcal{F}_{k-1}\big]-\frac{1}{r}\sum_{k=1}^{r}I_{i_k<\infty}\xrightarrow{r\to\infty}0 \quad \text{a.s.}$$

It follows that

$$\frac{1}{r}\sum_{k=1}^{r}\xi_k-\alpha(\boldsymbol{G}_*^m)\underbrace{\frac{1}{r}\sum_{k=1}^{r}I_{i_k<\infty}}\xrightarrow{r\to\infty}0 \quad \text{a.s.}$$

Here $\alpha(\boldsymbol{G}_*^m) = \boldsymbol{P}\big[G_t^m = \boldsymbol{G}_*^m\big] > 0$ by the choice of $\boldsymbol{G}_*^m$, and the underbraced expression equals 1 whenever (20) holds. By (21), this implies that the sample sequence $G_{i_k}^m$ runs through $\boldsymbol{G}_*^m$ infinitely many times for almost all events where (20) holds. It remains to invoke that $\boldsymbol{G}_*^m$ is an arbitrary realization assumed by G^m with a positive probability.

Proof of Proposition 1. By Lemma 1, it suffices to justify (11). Let (11) fail to be true for some coder-decoder pair without (with) a feedback. By the foregoing, then for any $m \ge 2$, there exists a zero error block code of length m without (with) a feedback for which (13) holds. So

$$\begin{aligned}
&\mathfrak{c}_0(\text{respectively}, \mathfrak{c}_{0F}) \ge m^{-1}\log_2 N\\
&\overset{(13)}{\ge} m^{-1}(\log_2 D-\log_2 b)\dim L_\oplus+\left(1-2m^{-1}\right)\eta(A).
\end{aligned}$$

Letting $m \to \infty$ yields $\mathfrak{c}_0 \ge \eta(A)$ ($\mathfrak{c}_{0F} \ge \eta(A)$), in violation of the hypotheses of Proposition 1. The contradiction obtained proves that (11) does hold.

Proof of Theorem 1. This theorem follows from Proposition 1.

VI. Conclusions

We studied observability of linear discrete-time unstable systems over noisy discrete memoryless channels. The system is affected by exogenous disturbances. We followed the natural approach aimed at observability along any (or almost any) trajectory (in other words, at keeping the estimation error bounded almost surely.) Is it possible to achieve this objective, provided that the plant disturbances are uniformly bounded and arbitrarily small? We showed that generally speaking, the answer is in the affirmative. However the capability of a noisy channel to ensure the affirmative answer is, in some sense, identical to its capability to transmit information with the zero probability of error. The standard measure of the latter capability is the so-called zero-error capacity of the channel [19]. Unfortunately, this capacity equals zero for most channels of practical interest [5]. The results of this paper mean that for such channels, the trajectory-wise observability cannot be achieved by any means. Specifically, for any non-anticipating observer, the error is unbounded with probability 1: uniformly bounded and arbitrarily small plant disturbances cause, sooner or later, arbitrarily large estimation errors. Then only weaker forms of observability (such as observability in probability or m-th moment one) appear to be relevant. However they may ensure only that unavoidable large errors do not occur systematically and frequently.

Appendix: Proof of Lemma 1

Consider arbitrary coder and decoder without (with) a feedback and pick $\{b_i\}_{i=1}^{\infty}, b_i\uparrow\infty$. Thanks to (11), there exists a random time τ_1 and random admissible disturbance $\zeta(t), t = 0, \ldots, \tau_1 - 1$ such that $|x(\tau_1) - \widehat{x}(\tau_1)| > b_1$ a.s. Let us take the least such a time τ_1. Then τ_1 is the stopping time [10], i.e., for every k, the event $\{\tau_1 = k\}$ is in the σ-algebra generated by $x_0, G_0, \ldots, G_k$. The above disturbance extended $\zeta(t) := 0$ on $t = \tau_1$ can be chosen as a function of these variables and τ_1. Now we consider the tail $t \geq \tau_1 + 1$ of the process $x(t), y(t), \widehat{x}(t), e(t), s(t), G_t$ in the (regular [10]) probability space obtained by conditioning over $\tau_1 = k, x_0 = \boldsymbol{x}, G_0 = \widetilde{\boldsymbol{G}}_0, \ldots, G_k = \widetilde{\boldsymbol{G}}_k$. The data $\boldsymbol{d} = [k, \boldsymbol{x}, \widetilde{\boldsymbol{G}}_0, \ldots, \widetilde{\boldsymbol{G}}_k]$ uniquely determines the state $x(\tau_1 + 1) = \boldsymbol{x}_{k+1}$ and $y(0) = \boldsymbol{y}_0, \ldots, y(k) = \boldsymbol{y}_k, s(0) = \boldsymbol{s}_0, \ldots, s(k) = \boldsymbol{s}_k$. (We recall that for $t = 0, \ldots, k$, the disturbance was chosen as a function of $\boldsymbol{d}$.) Hence the tail starts at $t = k + 1$ at $x(k + 1) = \boldsymbol{x}_{k+1}$ and is governed by the coder and decoder that are obtained from putting $y(0) = \boldsymbol{y}_0, \ldots, y(k) = \boldsymbol{y}_k$ and $s(0) = \boldsymbol{s}_0, \ldots, s(k) = \boldsymbol{s}_k$ into (3) (or (4)) and (2), respectively. Since the above conditioning does not alter the transition probabilities of the channel, (11) still holds for the tail at hand. So by repeating the starting arguments of the proof, we see that there exist random time $\Delta\tau_2^{\boldsymbol{d}}$ and disturbance $\zeta^{\boldsymbol{d}}(t), t = k + 1, \ldots, k + \Delta\tau_2^{\boldsymbol{d}} - 1$ for which $|x(k + \Delta\tau_2^{\boldsymbol{d}}) - \widehat{x}(k + \Delta\tau_2^{\boldsymbol{d}})| > b_2$. Now we put $\tau_2 := \tau_1 + \Delta\tau_2^{\tau_1, x_0, G_0, \ldots, G_{\tau_1}}$ and continue the disturbance $\zeta(t) := \zeta^{\tau_1, x_0, G_0, \ldots, G_{\tau_1}}(t)$ on $t = \tau_1 + 1, \ldots, \tau_2 - 1$. After this, we have $|x(\tau_2) - \widehat{x}(\tau_2)| > b_2$ a.s. By continuing likewise, we construct a sequence $\{\tau_i\}$ of random times and a random admissible disturbance such that $|x(\tau_i) - \widehat{x}(\tau_i)| > b_i \,\forall i$ a.s., which implies (10).

References

[1] R.W. Brockett and D. Liberzon, Quantized Feedback Stabilization of Linear Systems, *IEEE Transactions on Automatic Control*, 45, pages 1279–1289, 2000.

[2] J. Hespanha, A. Ortega, and L. Vasudevan, "Towards the Control of Linear Systems with Minimum bit-rate", In *Proceedings of the International Symp. on the Mathematical Theory of Networks and Systems*, 2002. available at http://www.ece.ucsb.edu/ hespanha/published.html.

[3] H. Ishii and T. Basar, "Quadratic Stabilization of Sampled Data Systems with Quantization", In *Proceedings of the 15th IFAC World Congress*, Barselona, 2002.

[4] H. Ishii and T. Basar, "Remote Control of LTI Systems over Networks with State Quantization", In *Proceedings of the 41st IEEE Conference on Decision and Control*, pages 830–835, Las Vegas, USA, 2002.

[5] J. Körner and A. Orlitsky, Zero-error Information Theory, *IEEE Transactions on the Information Theory*, 44(6), pages 2207–2229, 1998.

[6] D. Liberzon, "A Note on Stabilization of Linear Systems Using Coding and Limited Communication", In *Proceedings of the 41st Conference on Decision and Control*, pages 836–841, Las Vegas, Nevada, December 2002.

[7] D. Liberzon, Hybrid Feedback Stabilization of Systems with Quantized Signals, *Automatica*, 39, pages 1543–1554, 2003.

[8] D. Liberzon, On Stabilization of Linear Systems with Limited Information, *IEEE Transactions on Automatic Control*, 48(2), pages 304–307, 2003.

[9] D. Liberzon, "Stabilizing a Nonlinear System with Limited Information Feedback", In *Proceedings of the 42nd Conference on Decision and Control*, pages 182–186, Maui, Hawaii, USA, December 2003.

[10] R.S. Liptser and A.N. Shiryayev, *Statistics of Random Processes I. General Theory*, Springer, NY, 1977.

[11] M. Loéve, *Probability Theory*, volume 2, Springer, New York, 4th edition, 1978.

[12] A.S. Matveev and A.V. Savkin, "An Analogue of Shannon Information Theory for Networked Control Systems: Part 1, State Estimation via a Noisy Discrete Channel; part 2, Stabilization via a Noisy Discrete Channel", In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 4485–4490,4491–4496, Atlantis, Bahamas, 2004.

[13] A.S. Matveev and A.V. Savkin, Comments on "Control Over Noisy Channels" and Relevant Negative Results, *IEEE Transactions on Automatic Control*, November 2005 (to appear).

[14] G.N. Nair and R.J. Evans, Stabilization with Data-Rate-Limited Feedback: Tightest Attainable Bounds, *Systems and Control Letters*, 41(1), pages 49–56, 2000.

[15] G.N. Nair and R.J. Evans, Exponential Stabilisability of Finite-Dimensional Linear Systems with Limited Data Rate, *Automatica*, 39, pages 585–593, 2003.

[16] A. Sahai, *Anytime Information Theory*, PhD thesis, Massachusetts Institute of Technology, February 2001.

[17] A.V. Savkin, Analysis and Synthesis of Networked Control Systems: Topological Entropy, Observability, Robustness, and Optimal Control, *Proceedings of the XVI IFAC World Congress, Prague*, July 2005..

[18] A.V. Savkin and I.R.Petersen, "Set-Valued State Estimation via a Limited Capacity Communication Channel", In *IEEE Transactions on Automatic Control*, 48, pages 676–680, 2003.

[19] C.E. Shannon, The Zero Error Capacity of a Noisy Channel, *IEEE Transactions on Information Theory*, IT-2, pages 8–19, 1956.

[20] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

[21] T. Simsek, R. Jain, and P. Varaiya, Scalar Estimation and Control with Noisy Binary Observation, *IEEE Transactions on Automatic Control*, 49(9), pages 1598–1603, 2004.

[22] D.J. Stiwell and B.E. Bishop, Platoons of Underwater Vehicles, *IEEE Control Syst. Magazine*, 20(6), pages 45–52, 2000.