The necessity and sufficiency of anytime capacity for control over a noisy communication link

Anant Sahai

***Abstract*— We review how Shannon's classical notion of capacity is not enough to characterize a noisy communication channel if we intend to use that channel as a part of a feedback loop to stabilize an unstable system. While classical capacity is not enough, another operational sense of capacity called "anytime capacity" is shown to be necessary for the stabilization of an unstable process. In cases of sufficiently rich information patterns between the encoder and decoder, we show that adequate "anytime capacity" is also sufficient for there to exist a stabilizing controller. The sufficiency result is then generalized to cases without any explicit feedback between the observer and the controller.**

I. Introduction

It is natural to wonder if Shannon's classical capacity is the appropriate characterization for communication channels being used in distributed control systems. Schulman and others have studied interaction in the context of distributed computation.[17], [9] In that work, capacity was not a question of major interest since polynomial or constant factor slowdowns were considered acceptable. Fundamentally, this was a consequence of being able to design all the system dynamics. In automatic control contexts, we generally do not get to design the plant dynamics.

Recent work on sequential rate distortion has shown that there is a natural notion of rate, $\sum \log_2 |\lambda_i|$ where the λ_i are the unstable eigenvalues of the plant, that can be attached to an unstable discrete-time process.[22], [23], [24] This notion of rate was justified by showing how to stabilize a system using noiseless feedback link with any rate $R > \sum \log_2 |\lambda_i|$, and that it is impossible to stabilize the system using a noiseless link of lower capacity.

We had previously showed that it is possible to stabilize controlled Gauss-Markov processes over suitable power-constrained AWGN channels[11], [21] where it turned out that Shannon capacity was tight and linear observers and controllers were sufficient.[1] Next, we studied stabilization in the context of the binary erasure channel. There, we showed that Shannon capacity is not sufficient for stability and introduced the anytime capacity as a candidate figure of merit.[12] Subsequently, there was related work by Elia that used ideas from robust control to deal with communication uncertainty in a mixed continuous/discrete context.[4], [5] A more complete set of pointers to prior work can be found in [15].

A. Sahai is with the Department of Electrical Engineering and Computer Science, U.C. Berkeley, sahai@eecs.berkeley.edu. Initial portions of this work were carried out while the author was a doctoral student at MIT LIDS under the supervision of Prof. Sanjoy Mitter.

In this paper, we show that for any bounded moment sense of stability, that adequate anytime capacity is necessary for any observer/controller pair to be able to stabilize a scalar system across the noisy channel. For situations where the observer has access to the control inputs, we also show the sufficiency of adequate anytime capacity. The sufficiency result is then generalized to the case where the observer only observes the plant state.

II. Inadequacy of classical capacity

Definition 2.1: A *discrete time memoryless channel (DMC)* is a probabilistic system with an input. At every time step t, it takes an input $a_t \in \mathcal{A}$ and produces an output $b_t \in \mathcal{B}$ with probability $p(b_t|a_t)$. Conditioned on a_t, b_t is independent of any other random variable in the system that occurs at time t or earlier.

The maximum rate achievable for a given sense of reliable communication is called the associated capacity. Shannon's classical reliability requires that after a suitably large end-to-end delay N that the average probability of error on a bit is below a specified ϵ. *Shannon classical capacity* C can also be calculated in the case of memoryless channels by solving an optimization problem:

$$C = \sup_{P(A)} I(A;B)$$

where the maximization is over the input probability distribution and $I(A;B)$ represents the mutual information through the channel.[7] This is referred to as a single letter characterization of channel capacity for memoryless channels. There is another sense of reliability and its associated capacity called *zero-error capacity* which requires the probability of error to be exactly zero with sufficiently large N. It does not have a simple single-letter characterization.[19]

For memoryless channels, the presence or absence of feedback does not alter the classical Shannon capacity.[7] More surprisingly, for symmetric DMCs, the Gallager style reliability functions (error exponents) also do not change with feedback, at least in the high rate regime.[3] From a control perspective, this is our first indication that neither Shannon's capacity nor Gallager's block-coding reliability are the perfect fit for control applications.

A. The control problem

$$X_{t+1} = \lambda X_t + U_t + W_t, \quad t \geq 0 \tag{1}$$

where $\{X_t\}$ is an $I\!R$-valued state process. $\{U_t\}$ is an $I\!R$-valued control process and $\{W_t\}$ is a bounded

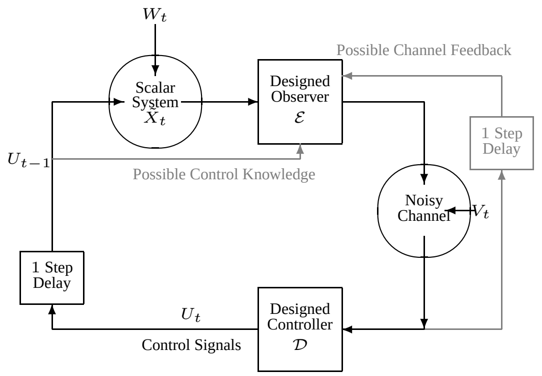

Fig. 1. Control over a noisy communication channel.

noise/disturbance process s.t. $\|W_t\| \leq \frac{\Omega}{2}$. For convenience, we also assume a known initial condition $X_0 = 0$.

To make things interesting, $\lambda > 1$ so the open-loop system is exponentially unstable. The distributed nature to the problem is illustrated in figure 1. We require an observer/encoder system $\mathcal{E}$ to observe X_t and causally generate inputs a_t to the channel. We also require a decoder/controller system $\mathcal{D}$ to observe channel outputs b_t and causally generate control signals U_t. We allow both $\mathcal{E}, \mathcal{D}$ to have arbitrary memory and to be nonlinear in general.

Definition 2.2: A dynamic system with time-varying state X_t is *η-stable* if there exists a constant K s.t. $E[|X_t|^\eta] \leq K$ for all t.

This definition requires the η-th moment to be bounded.[1] Because of the stochastic nature of the channel and the driving noise, it is too much to require that the state is bounded almost surely. There will usually be rare sequences of channel noise that will cause the state to grow outside any boundary. If the system is allowed to run for a long time, then it becomes increasingly certain that a run of bad luck on the channel will occur sometime. Trying to hold the η-th moment within bounds is a way of keeping large deviations rare. The larger η is, the more strongly we penalize very large deviations.

For the scalar case, the intrinsic rate $\log_2 \lambda$. This means that it is impossible to stabilize the system if the feedback channel's Shannon classical capacity $C < \log_2 \lambda$.

B. The real erasure channel

The packet erasure channel models situations where errors can be reliably detected at the receiver. In the model, the packet does not make it through with probability δ, but otherwise it makes it through correctly.

Definition 2.3: The *real packet erasure channel* has $\mathcal{A} = \mathcal{B} = \mathbb{R}$ and $p(x|x) = 1 - \delta$ while $p(0|x) = \delta$.

[1]To avoid getting bogged down in notation, we are ignoring the issue of not having a probability distribution for the plant disturbance.[13] and [12] deal with this explicitly by invoking a hypothetical observer which considers the distribution of an interval of uncertainty that contains the state. Alternatively, you can consider us as requiring the η-th moment to be bounded with a uniform bound over all possible disturbance sequences.

This model has also been explored in the context of Kalman filtering with lossy observations.[6]

For our counterexample, set $\eta = 2$ and $\lambda = \frac{3}{2}$. Let $\delta = \frac{1}{2}$ so that there is a 50% chance of any real number being erased. For the driving noise/disturbance W_t, assume that it is zero-mean and i.i.d. with variance σ^2.

The classical Shannon capacity[2] of the real erasure channel is $\infty > \log_2 \frac{3}{2}$ since when the real number is received correctly, it can communicate an infinite number of bits correctly. Thus the basic necessary condition for stabilizability is met.

In this case, the optimal distributed control is obvious — set $a_t = X_t$ as the transmission and use $U_t = -\lambda b_t$ as the control. With every successful reception, the system state is reset to the initial condition of zero. For an arbitrary time t, the time since it was last reset is distributed like a $\frac{1}{2}$-geometric random variable[3]. Thus the second moment is:

$$\begin{aligned}
E[|X_{t+1}|^2] &> \sum_{i=0}^{t} \frac{1}{2}(\frac{1}{2})^i E[(\sum_{j=0}^{i}(\frac{3}{2})^j W_{t-j})^2] \\
&= \sum_{i=0}^{t} \frac{1}{2}(\frac{1}{2})^i \sum_{j=0}^{i}\sum_{k=0}^{i}(\frac{3}{2})^{j+k} E[W_{t-j}W_{t-k}] \\
&= \sum_{i=0}^{t} (\frac{1}{2})^{i+1} \sum_{j=0}^{i}(\frac{9}{4})^j \sigma^2 \\
&= \frac{4\sigma^2}{5} \sum_{i=0}^{t} \left((\frac{9}{8})^{i+1} - (\frac{1}{2})^{i+1} \right)
\end{aligned}$$

This diverges as $t \to \infty$ since $\frac{9}{8} > 1$. Notice that the root of the problem is that $(\frac{3}{2})^2(\frac{1}{2}) > 1$. The system is getting away from us faster than the noisy channel is able to give us reliability. This causes the second moment to explode. Notice that the first moment[4] *is* bounded for all t since $(\frac{3}{2})(\frac{1}{2}) < 1$. So, *the adequacy of the channel depends on which moment we want to be bounded.* No single-number characterization like classical capacity can give us the figure-of-merit that we need to evaluate the channel for control applications.

III. Anytime capacity and its necessity

Intuitively, the estimates of the bits at the decoder must become increasingly reliable with time. For simplicity of notation, let M_i be the R bit message that the channel encoder took in at time i. As such, we can think of the anytime decoder as providing estimates $\hat{M}_i(t)$, the best estimate for message i at time t. If we are considering a delay d, the probability of error[5] we are interested in is $P(\hat{M}_{t-d}(t) \neq M_{t-d})$.

[2]Because a long string of erasures is always possible, the Shannon zero-error capacity of this channel is 0 as long as $\delta > 0$.

[3]This is the process that models the time till first head in a sequence of fair coin tosses.

[4]$E[|X_t|]$ rather than $E[X_t]$ since the latter is zero.

[5]In what follows, we will often consider single bit messages for simplicity of exposition.

A. Anytime capacity

Definition 3.1: The *α-anytime capacity* $C_{\text{anytime}}(\alpha)$ of a channel is the least upper bound of the rates at which the channel can be used to transmit data so that there exists a uniform constant K such that for all d and all times t we have[6]

$$P(\hat{M}_{t-d}(t) \neq M_{t-d}(t)) \leq K2^{-\alpha d}$$

The probability is taken over the channel and any randomness that we deem the encoder and decoder to have access to. Due to the fast convergence of an exponential, we could just as well have required that the probability of error for *all messages sent upto time* $t-d$ is bounded by $K'2^{-\alpha d}$.

The requirement for exponential decay in the probability of error with delay is reminiscent of the reliability function $E(R)$ of a channel given in [7]. There is one crucial difference, however. In the standard study of error exponents, we let both the encoder and decoder vary with blocklength N and hence delay. In our definition of α-anytime capacity, we require the encoding to be fixed and the decoder has to work at all delays.

This additional requirement is why we call it "anytime" capacity. We can query the decoding process for a given bit at any time and we require the answer to be increasingly meaningful the longer we wait. The anytime α specifies the exponential rate at which we want the answers to improve. The above sense of reliable transmission lies between that represented by zero-error capacity and Shannon capacity. It should be clear that $\forall \alpha, C_0 \leq C_{\text{anytime}}(\alpha) \leq C$.

By using a random coding argument over infinite tree codes, it is possible to show the existence of anytime codes without using feedback for all rates less than the Shannon capacity. We can show that

$$C_{\text{anytime}}(E_r(R)) \geq R$$

where $E_r(R)$ is Gallager's random coding error exponent calculated in bits and powers of two.[13] Since feedback plays an essential role in control, it will turn out that we are interested in the anytime capacity when feedback *is* available to the encoder.

B. Necessity of anytime capacity

Assume that we have an observer/controller pair that can stabilize the unstable system. We will constructively show the necessity of anytime capacity by using that pair to construct an anytime encoder and decoder for a channel with noiseless feedback. The key is to consider the simulated plant state as the sum of the states of two different unstable LTI systems. The first, with state denoted $\tilde{X}_t$, is driven entirely by the controls and starts in state 0.

$$\tilde{X}_{t+1} = \lambda \tilde{X}_t + U_t \tag{2}$$

These controls are available at the encoder due to the presence of noiseless feedback.[7] The other, with state denoted $\check{X}_t$, is driven entirely by a simulated driving noise that is generated from the data stream we want to communicate.

$$\check{X}_{t+1} = \lambda \check{X}_t + W_t \tag{3}$$

The sum $X_t = (\tilde{X}_t + \check{X}_t)$ is then fed to the observer which uses them to generate inputs for the noisy channel.

The decoder runs its own copy of the $\tilde{X}_t$ process based on the controls U_t coming from the controller which uses the outputs from the noisy channel. The fact that the original observer/controller pair stabilized the original system implies that $|X_t| = |\check{X} - (-\tilde{X}_t)|$ is small and hence $-\tilde{X}_t$ stays close to $\check{X}_t$. All we need to do now is extract our encoded bits from $(-\tilde{X}_t)$ and bound the probability of error.

1) Encoding data bits into the state: $\check{X}_t$ is the part of X_t that only has the driving noise as its input. We can expand the recurrence relation for it as:

$$\begin{aligned} \check{X}_t &= \lambda \check{X}_{t-1} + W_{t-1} \\ &= \lambda^{t-1} \sum_{j=0}^{t-1} \lambda^{-j} W_j \end{aligned}$$

This looks like the representation of a fractional number in base λ which is then multiplied by λ^{t-1}. We will exploit this in our encoding by choosing the bounded disturbance sequence so that we have

$$\check{X}_t = \gamma \lambda^t \sum_{k=0}^{\lfloor Rt \rfloor} (2+\epsilon_1)^{-k} B_k \tag{4}$$

where B_k is the k-th bit[8] of data that the anytime encoder has to send and $\lfloor Rt \rfloor$ is just the total count of bits that are ready to be sent by time t. γ is a constant chosen so as to meet the bound on the simulated disturbance W.

To see that (4) is always possible to achieve by appropriate choice of bounded disturbance, use induction. (4) clearly holds for $t=0$. Now assume that it holds for time t. To see that it holds at time $t+1$, we know that:

$$\begin{aligned} \check{X}_{t+1} &= \lambda \check{X}_t + W_t \\ &= \gamma \lambda^{t+1} \Big(\sum_{k=0}^{\lfloor Rt \rfloor} (2+\epsilon_1)^{-k} B_k\Big) + W_t \end{aligned}$$

So setting

$$W_t = \gamma \lambda^{t+1} \sum_{k=\lfloor Rt \rfloor + 1}^{\lfloor R(t+1) \rfloor} (2+\epsilon_1)^{-k} B_k \tag{5}$$

[6]We could alternatively have bounded the probability of error by $2^{-\alpha(d-\log_2 K)}$ and interpreted $\log_2 K$ as a minimum required delay.

[7]The encoder can have a copy of the controller and feed it the noiseless feedback to get the controls.

[8]For the next section, we want the disturbances to be balanced around zero and so we choose to represent the bit as +1 or -1 rather than the usual 0 or 1.

gives the desired result. To see that this can be made to meet the bound, we manipulate $W_t =$

$$\gamma\lambda^{t+1}(2+\epsilon_1)^{-\lfloor Rt\rfloor}\sum_{j=1}^{\lfloor R(t+1)\rfloor-\lfloor Rt\rfloor}(2+\epsilon_1)^{-j}B_{\lfloor Rt\rfloor+j}$$

$$= \gamma\lambda\frac{(2+\epsilon_1)^{Rt-(\lfloor Rt\rfloor)}}{\lambda^{-t(1-R\frac{\log_2(2+\epsilon_1)}{\log_2\lambda})}}\sum_{j=1}^{\lfloor R(t+1)\rfloor-\lfloor Rt\rfloor}(2+\epsilon_1)^{-j}B_{\lfloor Rt\rfloor+j}$$

To keep this bounded, we choose

$$\epsilon_1 = 2^{\frac{\log_2\lambda}{R}} - 2 \tag{6}$$

which is strictly positive only if $R < \log_2\lambda$. So by choosing

$$\gamma = \frac{\Omega}{2\lambda^{1+\frac{1}{R}}} \tag{7}$$

we are guaranteed to stay within the specified bound on the driving noise/disturbance.

2) Extracting data bits from the state estimate: At the decoder, we have $-\tilde{X}_t = \check{X}_t - X_t$ which is close to $\check{X}$ since X_t is small. To see how to extract bits from $-\tilde{X}_t$, we first consider how to recursively extract those bits from $\check{X}_t$.

Starting with the first bit, we notice that the set of all possible $\check{X}_t$ that start with $B_0 = +1$ is separated from the set of all possible $\check{X}_t$ that start with $B_0 = -1$ by a gap of

$$\gamma\lambda^t\left((1-\sum_{k=1}^{\lfloor Rt\rfloor}(2+\epsilon_1)^{-k}) - (-1+\sum_{k=1}^{\lfloor Rt\rfloor}(2+\epsilon_1)^{-k})\right)$$

$$> \gamma\lambda^t 2(1-\sum_{k=1}^{\infty}(2+\epsilon_1)^{-k})$$

$$= \lambda^t\frac{2\epsilon_1\gamma}{1+\epsilon_1}$$

Notice that this is a positive number that is growing exponentially in t. If the first $i-1$ bits are the same, then we can scale both sides by $(2+\epsilon_1)^i = \lambda^{\frac{i}{R}}$ to get the same expressions above and so by induction, it quickly follows that the minimum gap between the encoded state corresponding to two sets of bits that first differ in bit position i is given by

$$gap_i(t) > \lambda^{t-\frac{i}{R}}\frac{2\gamma\epsilon_1}{1+\epsilon_1} \tag{8}$$

Of course, the equation (8) is only valid for bits $i \leq \lfloor Rt\rfloor$ since the other bits have not been encoded yet. Because the gaps are all positive, the above shows that it is always possible to perfectly extract our data bits from $\check{X}_t$ by using an iterative procedure:[9]

1) Initialize threshold $T_0 = 0$ and counter $i = 0$.
2) Compare input I_t to T_i. If $I_t \geq T_i$, set $\hat{B}_i(t) = 1$. If $I_t < T_i$, set $\hat{B}_i(t) = -1$.
3) Increment counter i and update threshold $T_i = \gamma\lambda^t\sum_{k=0}^{i-1}(2+\epsilon_1)^{-k}\hat{B}_k$
4) Goto 2 as long as $i \leq \lfloor Rt\rfloor$

Since the gaps given by (8) are always positive, the procedure works perfectly if we use input $I_t = \check{X}_t$. At the decoder, we will use $I_t = -\tilde{X}_t$ and this leads to:

Lemma 3.1: Given a channel with access to noiseless feedback, for all rates $R < \log_2\lambda$, it is possible to encode bits into the simulated scalar plant so that the uncontrolled process behaves like (4) by using disturbances given in (5) and the formulas in (6) and (7). At the output end of the noisy channel, we can extract estimates $\hat{B}_i(t)$ for the i-th bit sent. The error event $\{\omega|\hat{B}_j(t) \neq B_j(t)\} \subseteq$

$$\{\omega|\exists i \leq j, \hat{B}_i(t) \neq B_i(t)\} \subseteq \{\omega||X_t| \geq \frac{\lambda^{t-\frac{j}{R}}\gamma\epsilon_1}{1+\epsilon_1}\} \tag{9}$$

Here ω denotes members of the underlying sample space.[10] (9) is easy to verify by looking at the complementary event $\{\omega||X_t| < \frac{\lambda^{t-\frac{j}{R}}\gamma\epsilon_1}{1+\epsilon_1}\}$ and using (8).

Theorem 3.2: For a given noisy channel, bound Ω, and $\eta > 0$, if there exists an observer $\mathcal{O}$ and controller $\mathcal{C}$ for the unstable scalar system that achieves $E[|X_t|^\eta] < K$ for all bounded driving noise $-\frac{\Omega}{2} \leq W_t \leq \frac{\Omega}{2}$, then $C_{\text{anytime}}(\eta\log_2\lambda) \geq \log_2\lambda$ bits per channel use for the noisy channel considered with noiseless feedback.

Proof: Using Markov's inequality we have:

$$\begin{aligned} P(|X_t| > m) &= P(|X_t|^\eta > m^\eta) \\ &\leq E[|X_t|^\eta]m^{-\eta} \\ &< Km^{-\eta} \end{aligned}$$

Using this with lemma 3.1, we get:

$$\begin{aligned} P(\hat{B}_i(t) \neq B_i(t)) &\leq P(|X_t| \geq \frac{\lambda^{t-\frac{i}{R}}\gamma\epsilon_1}{1+\epsilon_1}) \\ &< K(\frac{1}{\gamma}+\frac{1}{\gamma\epsilon_1})^\eta\lambda^{-\eta(t-\frac{i}{R})} \\ &= (K(\frac{1}{\gamma}+\frac{1}{\gamma\epsilon_1})^\eta)2^{-(\eta\log_2\lambda)(t-\frac{i}{R})} \end{aligned}$$

Notice that $t - \frac{i}{R}$ is the delay between the time that bit i was ready to send and the decoding time. □

IV. The Sufficiency of Anytime Capacity

To show the sufficiency of anytime capacity in characterizing a noisy channel, the choice of information pattern[25] can be critical.[21] We first establish the result when both channel outputs and controls are known to the observer.

A. Observer

The observer is constructed to keep the state uncertainty inside a box of size Δ by using bits at the rate R. It will do this by controlling a virtual process $\bar{X}_t$ governed by:

$$\bar{X}_{t+1} = \lambda\bar{X}_t + W_t + \bar{U}_t \tag{10}$$

To simulate $\bar{X}_{t+1}$ the observer computes $W_t = X_{t+1} - \lambda X_t - U_t$. The virtual control $\bar{U}_t$ takes on one of

[9] This is a minor twist on the procedure followed by serial A/D converters.

[10] If the bits to be sent are deterministic, this is the sample space giving channel noise realizations.

$2^{\lfloor R(t+1) \rfloor - \lfloor Rt \rfloor}$ values. For simplicity of exposition, we will ignore the integer effects and consider it to be one of 2^R values[11]. Suppose that $\bar{X}_t$ is known to lie within $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$. Then $\lambda \bar{X}_t$ will lie within $[-\frac{\lambda\Delta}{2}, \frac{\lambda\Delta}{2}]$. By choosing 2^R control values uniformly spaced within that set, we can guarantee that $\lambda \bar{X}_t + U_t$ will lie within $[-\frac{\lambda\Delta}{2^{R+1}}, \frac{\lambda\Delta}{2^{R+1}}]$. Finally, the state will be disturbed and $\bar{X}_{t+1}$ will be known to lie within $[-\frac{\lambda\Delta}{2^{R+1}} - \frac{\Omega}{2}, \frac{\lambda\Delta}{2^{R+1}} + \frac{\Omega}{2}]$.

To get a handle on the minimum Δ required as a function of R, we solve for the steady state where $\frac{\lambda\Delta}{2^R} + \Omega = \Delta$. This occurs[12] when $\Delta = \frac{\Omega}{1 - \lambda 2^{-R}}$ for $R > \log_2 \lambda$.

It is clear that the virtual controls $\bar{U}_t$ can be encoded causally using R bits per unit time. These bits are sent to the anytime encoder for transport over the noisy channel.[13]

B. Controller

The controller uses the current estimates from the anytime decoder to apply a control that tries to make the true state X_t stay close to the virtual state $\bar{X}_t$. It does this by having a pair of internal models. The first models the unstable system driven only by the actual controls. This is referred to as $\tilde{X}_t$ in (2) above. The second is its best estimate $\hat{X}_t$, based on the current bit estimates from the anytime decoder, of where the unstable system should be driven by only the virtual controls $\bar{U}_t$.

$$\hat{X}_{t+1}(t) = \sum_{i=0}^{t} \lambda^i \hat{U}_{t-i}(t) \tag{11}$$

Notice that this is not given in recursive form since all of the past estimates for the virtual controls are subject to re-estimation at the current t. We then apply a control U_t designed to make $\tilde{X}_{t+1} = \hat{X}_{t+1}(t)$.

$$U_t = \hat{X}_{t+1}(t) - \lambda \tilde{X}_t \tag{12}$$

C. Performance

With controls given by (12), it is clear that the true state X_t can be written as:

$$\begin{aligned} X_t &= \check{X}_t + \tilde{X}_t = \check{X}_t + \hat{X}_t(t-1) \\ &= \sum_{i=0}^{t-1} \lambda^i (W_{t-i} + \hat{U}_{t-i}(t-1)) \end{aligned}$$

Notice that the actual state X_t differs from the virtual state $\bar{X}_t$ only due to errors in virtual control estimation due to channel noise. If there are no errors in the prefix of $\hat{U}_{t-j}$ for $j \geq d$, and arbitrarily bad errors for $j < d$, then we could start at $\bar{X}_{t-d}$ and see how much the errors could have propagated since then:

$$X_t = \lambda^d \bar{X}_{t-d} + \sum_{i=0}^{d-1} \lambda^i (W_{t-i} + \hat{U}_{t-i}(t-1))$$

[11]For the details of how to deal with fractional R, please see the causal source code discussion in [13]

[12]In reality, our uncertainty approaches this from below since we start at the known initial condition 0

[13]Because the actual controls are being subtracted out, that the output of the observer is completely independent of the channel noise.

Comparing this with $\bar{X}_t$, and noticing that the maximum difference between the two inputs is $2\lambda\Delta$ gives us:

$$\begin{aligned} |X_t - \bar{X}_t| &= |\sum_{i=0}^{d-1} \lambda^i (\bar{U}_{t-i} - \hat{U}_{t-i}(t-1))| \\ &\leq 2\Delta\lambda^d \sum_{i=0}^{d-1} \lambda^{-i} \\ &< \lambda^d \frac{2\Delta}{1 - \lambda^{-1}} \end{aligned}$$

Since $|\bar{X}_t| \leq \frac{\Delta}{2}$, if we know that there were no errors in the prefix of estimated virtual controls up through d time steps ago, that

$$|X_t| < \lambda^d \frac{3\Delta}{1 - \lambda^y - 1} \tag{13}$$

Specializing to the case of α-anytime capacity, we see that we get:

$$\begin{aligned} P(|X_t| \geq M) &\leq K''' 2^{-\alpha \frac{\log_2 M}{\log_2 \lambda}} \\ &= K''' M^{-\frac{\alpha}{\log_2 \lambda}} \end{aligned}$$

which gives a power-law bound on the tail. From this we see that if we want to hold the η-th moment bounded,

$$\begin{aligned} E[|X_t|^\eta] &= \int_0^\infty P(|X_t|^\eta \geq m) dm \\ &\leq 1 + K''' \int_1^\infty m^{-\frac{\alpha}{\eta \log_2 \lambda}} dm \end{aligned}$$

As long as $\alpha > \eta \log_2 \lambda$, the integral above converges and hence the controlled process has a bounded η-moment. This result is summarized in the following theorem:

Theorem 4.1: It is possible to control an unstable scalar process driven by a bounded disturbance over a noisy channel so that the η-moment of X_t stays finite for all time if the channel with feedback has $C_{\text{anytime}}(\eta \log_2 \lambda + \epsilon) > \log_2 \lambda$ for some $\epsilon > 0$ and the observer is allowed to observe the noisy channel outputs and the state exactly.

V. Relaxing feedback

In this section, we relax the assumption that the observer can observe the outputs of the noisy channel directly. Instead, we require the observer to only see the states X_t. As such, at time t the observer can see the control signal, and thus indirectly the channel output, only as it is corrupted by the process disturbance since $U_{t-1} + W_{t-1} = X_t - \lambda X_{t-1}$.

If we want noiseless information to flow from the controller to the observer, we can ask what is the instantaneous zero-error capacity of the effective channel *through the plant.* In addition, there remains the problem of the dual-nature of the control signal — it is simultaneously being asked to stabilize the plant as well as to feedback information about the channel outputs. It turns out that the ability of the controller to move the plant provides enough feedback to the encoder in the case of finite channel output alphabets.

Theorem 5.1: It is possible to control an unstable scalar process driven by a bounded disturbance over a noisy finite-alphabet channel so that the η-moment of X_t stays finite for all time if the channel with feedback has $C_{\text{anytime}}(\eta \log_2 \lambda + \epsilon) \geq \log_2 \lambda$ for some $\epsilon > 0$ if the observer is allowed to observe the state X_t exactly.

Proof: Let S be the size of the channel output alphabet. We use a modified control signal $U_t'' = Q(U_t) + F(Z_t) - \lambda F(Z_{t-1})$. Here Q quantizes its input to the nearest integer multiple of $S\Omega$. F maps the channel output into the range $(-\frac{S\Omega}{2}, +\frac{S\Omega}{2})$ as follows:

$$F(y) = y\Omega - \frac{\Omega}{2} - \frac{S\Omega}{2}$$

Notice that $\min_{i \neq j} |F(j) - F(i)| = \Omega$ and so under a bounded disturbance no bigger than $\frac{\Omega}{2}$ in either direction we can always recover the channel output Y perfectly. By induction, the observer can remove $\lambda F(Z_{t-1})$ and will then be able to extract both $Q(U_t)$ and $F(Z_t)$ noiselessly from $U_t'' + W_t$. The net impact on the control of the state will be finite — no more than $\lambda S \Omega$ in either direction — and so will not effect the qualitative tail probabilities. □

VI. Conclusions

At this point, it is interesting to consider three implications of the proof of Theorem 3.2. For a DMC define $\epsilon_m = \min_{i,j} p(i,j)$. So with or without feedback, the probability of error after d time steps is lower bounded by ϵ_m^d. Thus generically, the probability of error can drop no more than exponentially in d. It turns out that this rules out the "risk sensitive" sense of stability in which we require $P(|X_t| > x)$ to decrease exponentially. In fact in the context of theorem 3.2, this implies that there is an η beyond which all moments must be infinite! *If any unstable process is controlled over a DMC, then the resulting state can have at best a power-law bound on its distribution.*

We further observe that if all we want is for the state to have some finite moment $\eta > 0$, then we are fine with any DMC with Shannon capacity larger than $\log_2 \lambda$. This is because the limit as $\alpha \to 0$ is the Shannon capacity.[13] The important thing to notice even in this case is that the constructions for necessity given above must use an anytime decoder for the channel, even though α might be very small.

Finally, if we want the state to stay inside a particular box almost surely or if we have a finite bound on the actuator U_t, then the same proof technique can be used to show that Shannon's zero-error capacity is necessary for stabilization. These implications, figures, expanded proofs for the results here, as well as the generalizations of these results to both the continuous and vector case, are explored in [15].

VII. Acknowledgments

I would like to thank the anonymous reviewers for their useful suggestions. If I had more room, I would have included some more illustrations here.

References

[1] R. Bansal and T. Basar, "Simultaneous Design of Measurement and Control Strategies for Stochastic Systems with Feedback." Automatica, Volume 25, No. 5, pp 679-694, 1989.

[2] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley, New York,1991.

[3] R. L. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback." Problemy Kibernetiki, Vol 8, pp 161-168, 1962.

[4] N. Elia, "When Bode meets Shannon: control-oriented feedback communication schemes." *Submitted to Transactions on Automatic Control*, June 2003

[5] N. Elia. "Stabilization of systems over analog memoryless channels." *Submitted to Systems and Control Letters*, May 2003

[6] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, S. Sastry, "Kalman filtering with intermittent observations," Preprint 2003.

[7] R.G. Gallager, *Information Theory and Reliable Communication*. New York, NY: John Wiley and Sons, 1971.

[8] G. Nair and R. Evans, "State Estimation with a Finite Data Rate." Forthcoming paper, 1998

[9] S. Rajagopalan, "A Coding Theorem for Distributed Computation." PhD Dissertation, University of California at Berkeley, 1994

[10] A. Sahai, "Information and Control." Unpublished Area Exam, Massachusetts Institute of Technology, December 1997

[11] A. Sahai, S. Tatikonda, S. Mitter, "Control of LQG Systems Under Communication Constraints." Proceedings of the 1999 American Control Conference, 1999.

[12] A. Sahai, "Evaluating Channels for Control: Capacity Reconsidered," Proceedings of the 2000 American Control Conference. ACC (IEEE Cat. No.00CH36334). American Autom. Control Council. Part vol.4, 2000, pp.2358-62 vol.4. Danvers, MA, USA

[13] A. Sahai, "Any-time Information Theory," PhD Dissertation, MIT February 2001.

[14] A. Sahai and S. K. Mitter, "A Fundamental Need for Differentiated "Quality of Service" Over Communication Links: An Information Theoretic Approach.", 2000 Allerton Conference on Communication, Control, and Computing.

[15] A. Sahai and S. K. Mitter, "The necessity and sufficiency of anytime capacity for control over a noisy communication link: Parts I and II", In preparation.

[16] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback – I: No bandwidth constraint." IEEE Trans. on Information Theory, Vol 12, pp 172-182, April 1966

[17] L. Schulman, "Coding for Interactive Communication." IEEE Trans. on Information Theory, Vol 42, pp. 1745-1756, November 1996.

[18] C. E. Shannon, "A Mathematical Theory of Communication." Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July and October 1948.

[19] C. Shannon, "The Zero Error Capacity of a Noisy Channel." IEEE Trans. on Information Theory, Vol 2, pp. S8-S19, September 1956.

[20] S. Tatikonda, A. Sahai, S. Mitter, "Control of LQG Systems Under Communication Constraints." Proceedings of the 37th IEEE Conference on Decision and Control, 1998.

[21] S. Tatikonda, A. Sahai, S. Mitter, "Stochastic Linear Control Over a Communication Channel," accepted at IEEE Transactions on Automatic Control, May 2003

[22] S. Tatikonda, "Control Under Communication Constraints." PhD Dissertation, 2000.

[23] S. Tatikonda, S. Mitter, "Control Under Communication Constraints," accepted at IEEE Transactions on Automatic Control, March 2002

[24] S. Tatikonda, S. Mitter, "Control Over Noisy Channels," accepted at IEEE Transactions on Automatic Control, March 2002

[25] H. Witsenhausen, "Separation of Estimation and Control for Discrete Time Systems." Proceedings of the IEEE, Volume 59, No. 11, November 1971.