Anytime Capacity of a Class of Markov Channels

Paolo Minero Member, IEEE and Massimo Franceschetti Senior Member, IEEE

Abstract—Several new expressions for the anytime capacity of Sahai and Mitter are presented for a time-varying rate-limited channel with noiseless output feedback. These follow from a novel parametric characterization obtained in the case of Markov time-varying rate channels, and include an explicit formula for the $r$-bit Markov erasure channel, as well as formulas for memoryless rate processes including Binomial, Poisson, and Geometric distributions. Beside the memoryless erasure channel and the additive white Gaussian noise channel with input power constraint, these are the only cases where the anytime capacity has been computed. At the basis of these results is the study of the threshold function for $m$-th moment stabilization of a scalar linear system controlled over a Markov time-varying digital feedback channel that depends on $m$ and on the channel’s parameters. This threshold is shown to be a continuous and strictly decreasing function of $m$ and to have as extreme values the Shannon capacity and the zero-error capacity as $m$ tends to zero and infinity, respectively. Its operational interpretation is that of achievable communication rate, subject to a reliability constraint.

Index Terms—Anytime capacity, entropy power inequality, Markov channels, Markov jump linear systems, quantized control.

I. INTRODUCTION

We consider the problem of moment stabilization of a dynamical system where the estimated state is transmitted for control over a time-varying, rate-limited communication channel. This set-up has been studied extensively in the context of networked control systems and discussed in several special issue journals dedicated to the topic [1]–[3]. Recently, it gained renewed attention due to its relevance for the design of cyberphysical systems [4]. A tutorial review with extensive references appears in [5], see also [6].

The notion of Shannon capacity is in general not sufficient to characterize the trade-off between the entropy rate production of the system, expressed by the growth of the state space spanned in open loop, and the communication rate required for its stabilization. A large Shannon capacity is useless for stabilization if it cannot be used in time for control. For the control signal to be effective, it must be appropriate to the current state of the system. Since decoding the wrong codeword implies applying a wrong signal and driving the system away from stability, applying an effective control signal depends on the history of whether previous codewords were decoded correctly or not. In essence, the stabilization problem is an example of interactive communication [7], where two-way communication occurs through the feedback loop between the plant and the controller. For this reason, usage of alternative capacity notions with stronger reliability constraints than a vanishing probability of error has been proposed, including the zero-error capacity [8], originally introduced by Shannon [9], and the anytime capacity of Sahai and Mitter [10], [11]–[14].

Within this general framework, we focus on the $m$-th moment stabilization of an unstable scalar system whose state is communicated over a rate-limited channel capable of supporting $R_k$ bits at each time step $k$ and evolving randomly in a Markovian fashion, see Fig. 1. The rate process is known causally to both encoder and decoder. A motivating example is given by sensors and actuators communicating over a channel whose quality varies over time due to a random fading process. For digital communication, this reflects in a time variation of the rate supported by the channel. If the channel variations are slow enough, transmitter and receiver can estimate the quality of the link by sending a training sequence to acquire channel state information, and adapt their communication scheme to the channel’s condition.

Many variations of this “bit-pipe” model have been studied in the literature [15]–[35], including the case of fixed rate channel; the erasure channel where the rate process can take value zero; and the packet loss channel, where the rate process can oscillate randomly between the two values of zero and infinity, allowing a real number with infinite precision to be transported across the channel in one time step. Connections between the rate limited and the packet loss channel have been pointed out in [28], [29], showing that results for the latter model can be recovered by appropriate limiting arguments. Our Markovian channel model includes all of
these models as special cases and our theorems recover the corresponding previous stabilization results.

Beside the bit-pipe model, stabilization over the additive white Gaussian channel has also been considered in [36]–[40] and in this case the Shannon capacity has been shown to be sufficient to express the rate needed for stabilization. Extensions to the additive colored Gaussian channel [41] show that the maximum “tolerable instability” — expressed by the sum of the logarithms of the unstable eigenvalues of the system that can be stabilized by a linear controller with a given power constraint over a stationary Gaussian channel—corresponds to the Shannon feedback capacity [42], that assumes the presence of a noiseless feedback link between the output and the input of the channel and that is subject to the same power constraint. This suggests a duality between the problems of control and communication in the presence of feedback, and indeed it has been shown that efficient feedback communication schemes can be obtained by solving a corresponding control problem [43], [44].

The novel contribution of the present paper is in the context of the bit-pipe model, and is in the introduction of a stability threshold function of the channel’s parameters and of the moment stability number $m$ that converges to the Shannon capacity for $m \to 0$, to the zero-error capacity for $m \to \infty$, and it provides a parametric characterization of the anytime capacity for the remaining values of $m$. This function yields a novel anytime capacity formula in the special case of the $r$-bit Markov erasure channel, as well as novel formulas for memoryless rate processes including Binomial, Poisson, and Geometric distributions. It also provides a general strategy to compute the anytime capacity of arbitrary memoryless channels with given rate distributions. To prove our results, we also require some extensions of the theory of Markov Jump Linear Systems (MJLS), that can be of independent value.

Coding and decoding schemes for our channel model are not an issue, since no error protection or redundancy is needed. This is always the case for erasure channels and limited rate channels with causal feedback and no noise, and it represents the main difficulty in the extension of our results to noisy channels beside erasures, and to channels with partial channel state information. On the other hand, we do not see any difficulty to extend results to vector systems by exploiting known bit-allocation techniques described in [28], [29]. Additional extensions considering disturbances of unbounded support can also be obtained using standard adaptive schemes [21], [28], [29]. We refrain from presenting these additional results to keep the length of the paper under control, and also because they would not introduce any new methods.

The rest of the paper is organized as follows. Some preliminary results on Markov Jump Linear Systems, necessary for our derivations are presented in Section II. Section III describes the system and channel model and introduces the stability threshold function, illustrating some of its properties. Section IV describes relationships with the anytime capacity, and provides some representative examples. Section V provides the formula for the anytime capacity of the Markov $r$-bit erasure channel. Section VI considers the case of memoryless bit-pipe channels of arbitrary rate distribution.

Throughout the paper, the following notation is used. Logarithms are assumed to be in base two; random variables are denoted with uppercase letters, while their realizations with lowercase letters; matrices are also denoted in uppercase letters, but we denote them with the special typeset $\mathbb{A}$.

II. MARKOV JUMP LINEAR SYSTEMS.

Consider the scalar non-homogeneous MJLS [45] with dynamics

$$z_{k+1} = a_k z_k + w_k,$$

(1)

where $z_k$ denotes the state, $w_k$ is the realization of a stochastic disturbance $W_k$, and $a_k$ is the realization of $\{A_k\}_{k \geq 0}$, a Markov rate process defined on a finite set of reals $\mathcal{A} = \{\bar{a}_1, \ldots, \bar{a}_n\}$,

(2)

and with one-step transition probability matrix $P$ having entries

$$p_{ij} = P\{A_k = \bar{a}_j | A_{k-1} = \bar{a}_i\}$$

(3)

for every $i, j \in \{1, \ldots, n\}$.

We assume that the initial condition is a realization of a stochastic variable $Z_0$ that is mutually independent of the disturbance process $\{W_k\}_k$, and both have uniformly bounded $m$-th moment, i.e., $\sup_k \mathbb{E}(|W_k|^m) \leq \infty$ and $\mathbb{E}(|Z_0|^m) \leq \infty$. The system (1) is said to be weakly $m$-th moment stable if

$$\sup_k \mathbb{E}(|Z_k|^m) < \infty.$$  

(4)

Let $A \in \mathbb{Z}_+^{n \times n}$ be a diagonal matrix with diagonal entries $|\bar{a}_1|^m, \ldots, |\bar{a}_n|^m$, i.e.,

$$A = \text{diag}(|\bar{a}_1|^m, \ldots, |\bar{a}_n|^m).$$

(5)

The following theorem states the necessary and sufficient condition for $m$-th moment stability of the system (1) in terms of the the spectral radius $\rho(P^T A)$, that is the maximum among the absolute values of the eigenvalues of $P^T A$. 

0018-9286 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Theorem 1. For any $m \in \mathbb{R}^+$, the MJLS (1) is weakly $m$-th moment stable if and only if
\[ \rho(P^TA) \lesssim 1. \] (6)

The symbol \( \lesssim \) indicates that while the necessary condition holds with a weak inequality, the sufficient condition requires a strong inequality.

Proof: The sufficient condition is obtained by subsampling the original MJLS at a sampling rate of $\tau$ samples and then showing the weak stability of the subsampled system $Y_k = Z_{k\tau}$, $k = 0, 1, \ldots$. The weak stability of $\{Y_k\}_k$ is sufficient to ensure that (4) holds, because the $m$-th moment of the system in fact cannot grow unbounded in any finite time horizon due the assumptions that $|\bar{a}_1|, \ldots, |\bar{a}_n|$ are finite and that $\sup_k E(|W_k|^m) < \infty$.

By iterating (1) $\tau$ times, it can be seen that the subsampled process is a non-homogeneous MJLS with dynamics
\[ y_{k+1} = b_k y_k + c_k, \] (7)
where for every $k \geq 0$ we define $B_k = \prod_{j=k\tau}^{(k+1)\tau-1} A_k$ as the random expansion of the open-loop system during the $k$-th sampling interval and $C_k = \sum_{j=0}^{\tau-1} \prod_{i=j}^{\tau-1} A_{k\tau+i} W_{k\tau+j}$ as the total disturbance entering the system in the $k$-th sampling interval. Notice that $\sup_k E(|C_k|^m) < \infty$, since by assumption the disturbance process $\{W_k\}_k$ has bounded $m$-th moment.

By taking the $m$-th power in both sides of (7) and by applying the inequality $(x+y)^m \leq 2^m(x^m+y^m)$, which holds for any $m > 0$, it follows that
\[ E(|Y_{k+1}|^m) \leq 2^m E(|B_k|^m|Y_k|^m) + \text{const}, \] (8)
where const represents a uniform bound on the $m$-th moment of $\{C_k\}_k$. Let $v_{k,i} = E(|Y_{k+1}|^m|\{A_{k} = \bar{a}_i\})$, $i = 1, \ldots, n$. By repeatedly applying the law of total probability on both sides of (8), it can be easily seen that the vector $v_k = (v_{k,1}, \ldots, v_{k,n})^T$ satisfies the following component-wise vector inequality
\[ v_{k+1} \leq 2^m (P^TA)^T v_k + c, \] (9)
where $c$ is again a constant that only depends on the statistics of the disturbance process. If (6) holds we can choose the subsampling rate $\tau$ large enough to ensure that $\rho(2^m (P^TA)^T) < 1$ and thus that the above recursion remains bounded. Since $E(|Y_k|^m) = \sum_{i=1}^n v_{k,i}$ it follows that $\rho(P^TA) < 1$ is sufficient to ensure that $\sup_k E(|Y_k|^m) < \infty$ as desired.

The proof of the necessary condition relies on the fact that the non-homogeneous MJLS $\{Z_{k}\}_k$ is weakly stable only if the homogeneous MJLS obtained by setting $w_k = 0$ in (1) is weakly stable. Therefore, we focus on the homogeneous setting wherein $w_k = 0$ for all $k$ and let $v_{k,i} = E(|Z_{k+1}|^m|\{A_{k} = \bar{a}_i\})$ for $i = 1, \ldots, n$. By (1) and the law of total probability, the vector $v_k = (v_{k,1}, \ldots, v_{k,n})^T$ evolves over time according to the linear system
\[ v_{k+1} = P^T A v_k. \] (10)
Since $E(|Z_k|^m) = \sum_{i=1}^n v_{k,i}$, we conclude that (4) can hold only if the above linear system is stable and therefore only if $\rho(P^TA) \leq 1$, as claimed.

Theorem 1 extends the well known conditions for second moment stability given in [45] to $m$-th moment stability. A similar result appears in [46, Theorem 3.2], limited to the special case of a homogeneous MJLS driven by an i.i.d. rate process.

III. MOMENT STABILIZATION OVER BIT-PIPE MARKOV CHANNELS

Results on the stability of MJLS are used to characterize the stability of linear dynamical systems where the estimated state is sent to the controller over a digital communication link whose state is described by a Markov process, as depicted in Fig. 1.

![Feedback loop model](image)

Fig. 1. Feedback loop model. The estimated state is quantized, encoded and sent to a decoder over a digital channel of state $R_k$ that evolves according to a Markov process of discrete time step $k$.

A. System model

Consider the scalar dynamical system
\[ x_{k+1} = \lambda x_k + u_k + v_k, \]
\[ y_k = x_k + w_k, \] (11)
where $k \in \mathbb{N}$, and $|\lambda| \geq 1$. The variable $x_k$ represents the state of the system, $u_k$ the control input, $v_k$ is the realization of an additive stochastic disturbance $V_k$, $y_k$ is the sensor measurement, and $w_k$ is the realization of the random measurement noise $W_k$. The initial condition $x_0$ is the realization of a stochastic variable $X_0$. Both disturbance and noise are independent of each other and of the initial condition. They are also independent of the channel state, as defined below. $X_0$, $V_k$ and $W_k$, $k \geq 0$, all have uniformly bounded $m$th moment, i.e., $E(|X_0|^m) \leq \infty$, $\sup_k E(|V_k|^m) \leq \infty$ and

\[ E(|W_k|^m) \leq \infty \]
The disturbance has a continuous probability density function \( f \) of finite differential entropy \( h(V_k) := - \int f(x) \log_2 f(x) \, dx \), so there exists a constant \( \beta > 0 \) such that \( e^{2h(V_k)} > \beta \) for all \( k \).

### B. Channel Model

The state observer is connected to the actuator through a noiseless digital communication link that at each time \( k \) allows transmission without errors of \( r_k \) bits. The rate process \( \{R_k\}_{k \geq 0} \) is modeled as a homogeneous positive-recurrent Markov process defined on the finite set

\[
R = \{\bar{r}_1, \ldots, \bar{r}_n\},
\]

for some integer numbers \( 0 \leq \bar{r}_1 < \cdots < \bar{r}_n \), and with one-step transition probability matrix \( P \) having entries

\[
p_{ij} = P\{R_k = \bar{r}_j | R_{k-1} = \bar{r}_i\}
\]

for every \( i, j \in \{1, \ldots, n\} \). In the sequel, we define \( R \in \mathbb{Z}^{n \times n}_+ \) as the diagonal matrix with diagonal entries \( \bar{r}_1, \ldots, \bar{r}_n \), i.e.,

\[
R = \text{diag}(\bar{r}_1, \ldots, \bar{r}_n).
\]

Encoder and decoder are supposed to have causal knowledge of the rate process. More precisely, for every \( k \geq 0 \), the coder is a function \( s_k = s_k(y_0, \ldots, y_k) \) that maps all past and present measurements into the set \( \{1, \ldots, 2^n\} \), the digital link is the identity function on this set \( \{1, \ldots, 2^n\} \), and the control \( u_k \) is a function \( \hat{x}_k(s_0, \ldots, s_k) \) that maps all past and present symbols sent over the digital link into \( R \).

This channel model corresponds to a finite state channel where the Markov state is available causally at both the encoder and the decoder. A channel with state is defined by a triple \((\mathcal{F} \times \mathcal{S}, p(g|f, s), \mathcal{G})\) consisting of an input set \( \mathcal{F} \), a state set \( \mathcal{S} \), an output set \( \mathcal{G} \), and a conditional probability assignment \( p(g|f, s) \) for every \( f \in \mathcal{F}, s \in \mathcal{S}, \text{and } g \in \mathcal{G} \). This channel is memoryless if the output \( g_k \) at time \( k \) is conditionally independent of everything else given \( (f_k, s_k) \). The state sequence is Markov if \( S_0, S_1, \ldots \) forms a Markov chain. According to these definitions, our channel model is a discrete-memoryless channel with Markov state \((\mathcal{F} \times \mathcal{S}, p(g|f, s), \mathcal{G})\) with \( \mathcal{F} = \mathcal{G} = \{1, \ldots, \bar{r}_n\}, \mathcal{S} = \{\bar{r}_1, \ldots, \bar{r}_n\} \),

\[
p(g|f, s) = \begin{cases} 1 & f = g \text{ and } f \leq s \\ 0, & \text{otherwise} \end{cases}
\]

and state transition probabilities

\[
p(s_{k+1} = \bar{r}_j | s_k = \bar{r}_i) = p_{ij}.
\]

The Shannon capacity of this channel is [47]

\[
C = \sum_{i=1}^{n} \pi_i \bar{r}_i,
\]

where \( (\pi_1, \ldots, \pi_n) \) denotes the unique stationary distribution of \( P \).

The zero-error capacity of this channel is [9]

\[
C_0 = \bar{r}_1.
\]

Next, we show that the capacities in (17) and (18) are the limiting values of a stability threshold function indicating the channel’s rate-reliability constraint required to achieve a given level of stabilization. As \( m \to \infty \) and the system is highly stable, then the stability threshold function tends to the zero-error capacity that has a hard reliability constraint of providing no decoding error. Conversely, as \( m \to 0 \) and the system’s stability level decreases, then the stability threshold function tends to the Shannon capacity that has a soft reliability constraint of vanishing probability of error.

### C. Stability threshold function

The system \((11)\) is \( m \)-th moment stable if

\[
\sup_k \mathbb{E}[|X_k|^m] < \infty,
\]

where the expectation is taken with respect to the random initial condition \( x_0 \), the additive disturbance \( \nu_k \), and the rate process \( R_k \).

The following Theorem establishes the equivalence between the \( m \)-th moment stability of \((11)\) and the weak moment stability of a suitably defined MJLS.

**Theorem 2.** It is possible to stabilize the scalar system \((11)\) in \( m \)-th moment sense over the rate-limited Markov channel if and only if the scalar non-homogeneous MJLS with dynamics

\[
z_{k+1} = |\lambda| \frac{1}{2mr_e} z_{k} + \text{const,}
\]

for some suitably defined constant \( \text{const} \) is weakly \( m \)-th moment stable, i.e., if and only if

\[
\log |\lambda| \lesssim -\frac{1}{m} \log \rho(P^{\top}2^{-mR}) \equiv R(m).
\]

The proof is given in the appendix assuming state feedback. In the case of output feedback the proof is more technical but the extension is immediate from standard techniques [29]. We also assume the disturbance has bounded support. This assumption is made to compare results to the ones on the anytime capacity that only apply to plants with bounded disturbance [10].
The extension to unbounded disturbance can be easily obtained using standard adaptive encoding schemes described in [21], [28], [29].

We mention several properties of the threshold function $R(m)$, whose proofs are given in the appendix.

**Proposition 1.** The following holds:

1) **Monotonicity:** $R(m)$ is continuous and strictly decreasing for $m > 0$.

2) **Convergence to the Shannon capacity:**

$$
\lim_{m \to 0} R(m) = \sum_{i=1}^{n} \pi_i \tilde{r}_i = C. \quad (22)
$$

3) **Convergence to the Zero Error capacity:**

$$
R(m) \sim \tilde{r}_1 - \frac{1}{m} \log p_{11}, \quad \text{as } m \to \infty, \quad (23)
$$

and hence

$$
\lim_{m \to \infty} R(m) = \tilde{r}_1 = C_0. \quad (24)
$$

4) **The function $mR(m)$ is nonnegative, strictly increasing, and strictly concave.** If $\tilde{r}_1 \neq 0$, then $mR(m)$ grows unbounded as $m \to \infty$. If instead $\tilde{r}_1 = 0$, then

$$
\lim_{m \to \infty} mR(m) = -\log p_{11}. \quad (25)
$$

**IV. ANYTIME CAPACITY OF BIT-PIPE MARKOV CHANNELS**

We now relate the stability threshold function $R(m)$ to the anytime capacity. For the given Markov channel, our results show that $R(m)$ provides a parametric representation of the anytime capacity in terms of system’s stability level $m$.

The anytime capacity is defined in the following context [12]. Consider a system for information transmission over a noisy channel that evolves in discrete time steps, allows the decoding time to be infinite, and improves the reliability of the estimated message as time progresses. More precisely, at each step $k$ a new message $m_k$ of $r$ bits is generated that must be sent over the channel. The coder and decoder at the two ends of the communication link are equipped with infinite memory and keep track of the messages and their corresponding estimates, respectively. At each time step, a single bit is sent by the encoder to the decoder over the channel. Upon reception of the new bit, the decoder updates the estimates for all messages up to time $k$. It follows that at time $k$ the estimates

$$
\hat{m}_0[k], \hat{m}_1[k], \ldots, \hat{m}_k[k]
$$

are constructed, given all the bits received up to time $k$. The estimate for any message $m_i$ is then continuously refined for all $k \geq i$. A reliability level $\alpha$ is achieved if the probability that there exists at least a message in the past whose estimate is incorrect decreases $\alpha$-exponentially with the number of bits received, namely for all $d \leq k$

$$
P\{(\hat{M}_0[k], \ldots, \hat{M}_d[k]) \neq (M_0, \ldots, M_d)\} = O(2^{-\alpha(k-d)}), \quad (26)
$$
as $k \to \infty$. It follows that the described communication system is characterized by a rate-reliability pair $(r, \alpha)$. The work in [12] has shown a correspondence between the ability to construct such a communication system over a given channel, and the ability to stabilize a scalar linear system of the type (11) over the same channel.

To precisely state this result in the context of our Markov channel, we define the $\alpha$-anytime capacity $C_A(\alpha)$ to be the supremum of the rate $r$ that can be achieved with reliability $\alpha$. Sahai and Mitter have proven the following result:

**Theorem 3** (Sahai, Mitter [12]). *The necessary and sufficient condition for $m$-th moment stabilization of a scalar system with bounded disturbances and in the presence of channel output feedback over a Markov channel is*

$$
\log |\lambda| \leq C_A(m \log |\lambda|). \quad (27)
$$

The anytime capacity is an intermediate notion between the zero-error capacity and the Shannon capacity. The zero-error capacity requires decoding without error. The Shannon capacity allows the decoding error to tend to zero as the length of the code used for transmission increases. While only a critical value of the zero-error capacity can guarantee the almost sure stability of the system [8], Sahai and Mitter show that for scalar systems subject to bounded disturbances a critical value of the anytime capacity can guarantee the ability to stabilize the system in the weaker $m$-th moment sense.

By combining Theorem 2 and Theorem 3, we obtain the following result on the equivalence of the reliability function $R(m)$ and the anytime capacity $C_A(\alpha)$ on the boundary of the stability region, where $\alpha = mR(m)$. With this choice of $\alpha$, both the reliability function and the anytime capacity can be arbitrarily close to $\log |\lambda|$ and still guarantee stability. This shows that the operational definition of anytime capacity in terms of a hypothetical communication system satisfying the exponential constraint (26) is not needed to characterize stability in this case. The simpler notion of reliability function is enough for our purposes. On the other hand, by carefully choosing $m$ as a function of $\alpha$, the reliability
function can also be used to compute the corresponding value of the anytime capacity.

**Theorem 4.** For Markov bit-pipe channels as defined in Section III-B, the following holds:

1) **Parametric characterization of the anytime capacity:** For every $m > 0$, the anytime capacity $C_A$ satisfies

$$C_A(mR(m)) = R(m),$$

i.e., for every $\alpha \geq 0$, there exists a unique $m(\alpha)$ such that

$$m(\alpha)R(m(\alpha)) = \alpha$$

and

$$C_A(\alpha) = R(m(\alpha)) = \frac{\alpha}{m(\alpha)}.$$  

2) $C_A(\alpha)$ is a strictly decreasing function function of $\alpha > 0$.

3) **Convergence to the Shannon capacity:**

$$\lim_{\alpha \to 0} C_A(\alpha) = \sum_{i=1}^{n} \pi_i \bar{r}_i = C,$$  

4) **Convergence to the Zero Error capacity:** If $\bar{r}_1 = 0$, then for every $\alpha \geq \log(1/p_{11})$

$$C_A(\alpha) = 0 = C_0.$$  

Conversely, if $\bar{r}_1 \neq 0$, then $C_A(\alpha)$ has unbounded support and

$$C_A(\alpha) \sim \bar{r}_1 \frac{\alpha}{\alpha - \log(1/p_{11})}, \quad \text{as } \alpha \to \infty,$$

hence

$$\lim_{\alpha \to \infty} C_A(\alpha) = \bar{r}_1 = C_0.$$  

**Proof:** By Theorem 2 and Theorem 3, at the boundaries of the stability region we must have that

$$\log |\lambda| = R(m).$$

and

$$\log |\lambda| = C(m \log |\lambda|).$$

Therefore, (28) follows by combining (35) and (36). Next, by Proposition 1, the function $\phi(m) = mR(m)$ is increasing and strictly concave, thus invertible. It follows that for every $\alpha \geq 0$, there exists a unique $m := m(\alpha)$ such that

$$mR(m) = \alpha.$$  

Substituting this equality into (28), it follows that $C_A(\alpha) = R(m(\alpha))$, as claimed. The remaining properties are immediate consequences of Proposition 1.

By definition, $C_A(\alpha)$ is non increasing in $\alpha$. Since both $mR(m)$ and $R(m)$ are both monotonic function of $m$, however, it follows that $C_A(\alpha)$ must be strictly decreasing in $\alpha$, thus establishing 2). Property 3) follows from (28) combined with the fact that $mR(m) \to 0$ and $R(m) \to C$ as $m \to 0$. Similarly, property 4) follow directly by combining (28) with properties 3) and 5) in Proposition 1.

We now give some representative examples of the stability threshold function, visually showing its extremal properties and its relationship to the anytime capacity.

**Example IV.1.** Let $n = 4, \mathcal{R} = \{1, 3, 4, 5\}$ and $P$ be is a $4 \times 4$ circulant matrix with first row equal to $\frac{1}{16}(1, 13, 1, 1)$, namely

$$P = \frac{1}{16} \begin{bmatrix} 1 & 13 & 1 & 1 \\ 13 & 1 & 1 & 1 \\ 1 & 1 & 1 & 13 \\ 1 & 1 & 13 & 1 \end{bmatrix}.$$  

In this case it is easy to compute $C = \frac{1}{4}(1 + 3 + 4 + 5) = \frac{13}{4}$ and $C_0 = 1$. Figure 2 plots the stability threshold function $R(m)$ (together with its asymptotic approximation) and the anytime capacity $C_A(\alpha)$. Both curves have the same shape and they are in fact related by an affine transformation as $m$ grows. Furthermore, both curves have unbounded support and tend to one at infinity. There is a change of convexity for small values of $m$ and $\alpha$, as indication that $R(m)$ and $C_A(\alpha)$ are generally not convex functions. In contrast, the function $\phi(m) = mR(m)$, reported in red in the top plot of Figure 2, is strictly convex and increasing.

**Example IV.2.** Let $\mathcal{R} = \{0, 3, 4, 5\}$ and $P$ is as in (38). The only difference with the previous example is that $\bar{r}_1$ is now 0 instead than 1. In this case it is easy to compute $C = \frac{1}{4}(0 + 3 + 4 + 5) = 3$ and $C_0 = 0$. Figure 3 plots the stability threshold function $R(m)$ (together with its asymptotic approximation) and the anytime capacity $C_A(\alpha)$. In this case, while $R(m)$ has unbounded support, $C_A(\alpha)$ is zero for all $\alpha \geq -\log p_{11} = \log 16 = 4$. This occurs because the function $\phi(m) = mR(m)$ saturates as $m \to \infty$, tending to the limiting value $-\log p_{11} = 4$.

When viewed together, the two examples above show that for some channels, like the one in Example IV.2, a communication system with an arbitrary rate-reliability pair $(r, \alpha)$ cannot be constructed, because the anytime capacity may have bounded support and tend abruptly to zero. However, in order to achieve $m$-th moment stabilization it is sufficient to consider the parametric
function $R(m) = C_A(mR(m))$, and construct a communication system whose reliability level depends on the desired stabilization level. It follows that we do not need to compute the whole anytime capacity if we are interested only in moment stabilization, and we may be content with determining the threshold function $R(m)$ corresponding to its parametric representation. The extremal properties of $R(m)$ determine the support of the anytime capacity corresponding to the achievable reliability level $\alpha$. If $R(m) = O(1/m)$ then the anytime capacity has support bounded by the pre-constant of the asymptotic order. On the other hand, if $R(m)$ decreases at most sub-linearly to zero, or it tends to a constant zero-error capacity, then the anytime capacity has unbounded support and any reliability level $\alpha$ can be achieved.

V. THE MARKOV $\tau$-BIT ERASURE CHANNEL

We use the stability threshold function to characterize the anytime capacity of the Markov $\tau$-bit erasure channel.

The Markov $\tau$-bit erasure channel corresponds to a two-state Markov process where $n = 2 \mathcal{R} = \{0, \bar{r}\}$, $p_{12} = q$, and $p_{21} = p$, where $0 < p, q < 1$. In this case,

$$p_{T,2}^{mR} = \left(1 - q\right) \frac{1 - 2^{-m}}{2^{mR}} + \left(q\right) \frac{1 - 2^{-mR}}{2^{mR}} \left(1 - p\right),$$

(39)

and we have the following result.

Theorem 5. The anytime capacity of the Markov $\tau$-bit erasure channel is

$$C_A(\alpha) = \frac{\alpha \overline{r}}{\alpha + \log_2 \left(1 - p \frac{2^{mR} (1 - p - q)}{1 - (1 - q) 2^{mR}}\right)},$$

(40)

if $0 \leq \alpha < - \log_2 (1 - q)$, and 0 otherwise.

Proof: By (29) and the definition of $R(m)$, for every $\alpha$ there exists an $m(\alpha)$ such that

$$\rho(P_{T,2}^{m(\alpha)R}) = 2^{-\alpha}$$

(41)

In the case of a Markov erasure channel, a simple calculation shows that

$$\rho(P_{T,2}^{mR}) = \frac{\text{tr}}{2} + \frac{1}{2} \sqrt{\text{tr}^2 - 4 \text{det}},$$

(42)
where \( \text{tr} \) and \( \det \) denote the trace and determinant of (39), respectively. By combining (41) and (42) and squaring both sides of the resulting equation, it follows that \( m(\alpha) \) must satisfy
\[
2^{-\alpha} \text{tr} - 4 \det = 2^{-\alpha + 1},
\]
where \( \text{tr} = (1 - q) + 2^{-m(\alpha)}(1 - p) \) and \( \det = 2^{-m(\alpha)}(1 - p - q) \). Solving (43) yields
\[
m(\alpha) = \frac{1}{r} \left( \alpha + \log_2 \left( \frac{1 - p - 2^\alpha (1 - p - q)}{1 - (1 - q)2^\alpha} \right) \right)
\]
for \( 0 \leq \alpha < -\log_2(1 - q) \). Finally, substituting (44) into (30) yields (40).

**A. Special cases**

Several special cases recover previous results in the literature. By (40) it follows that the anytime capacity of the binary erasure channel (BEC) with Markov erasures and with noiseless channel output feedback is
\[
C_A(\alpha) = \frac{\alpha}{\alpha + \log_2 \left( \frac{1-p-2^\alpha(1-p-q)}{1-(1-q)2^\alpha} \right)}.
\]

By letting \( q = 1-p \), the erasure process becomes i.i.d. and we recover the anytime capacity of the memoryless BEC with erasure probability \( p \) derived by Sahai [10, page 129] (in parametric form) and by Xu [14, Theorem 1.3] (in non-parametric form)
\[
C_A(\alpha) = \frac{\alpha}{\alpha + \log_2 \left( \frac{1-p}{1-p2^\alpha} \right)}.
\]

By (40), letting \( \alpha \to 0 \), we have that
\[
\lim_{\alpha \to 0} C_A(\alpha) = \frac{q}{p + q} = C,
\]
This recovers the Shannon capacity of an \( \bar{r} \)-bit erasure channel with Markov erasures and with noiseless channel output feedback.

In the case \( n = 2 \), \( \bar{r}_1 = 0 \), \( \bar{r}_2 = r \), and an i.i.d. rate process with \( P(R_k = 0) = p_1 \) and \( P(R_k = r) = p_2 \) for all \( k \), then the stability condition becomes
\[
|\lambda|^m (p_1 + p_2 2^{-mr}) < 1,
\]
which provides a converse to the achievable scheme of Yüksel and Meyn [48, Theorem 3.3].

If we further let \( r \to \infty \), then the stability condition \( p_1 > 1/|\lambda|^m \) only depends on the erasure rate of the channel. In this case, our condition generalizes the packet loss model result in [30].

**VI. MEMORYLESS BIT-PIPE CHANNELS**

We show that the stability threshold function can be used to determine the anytime capacity of an arbitrary memoryless bit-pipe channel of given rate distribution and provide three relevant examples of this computation. Aside the memoryless erasure channel and the additive white Gaussian noise channel with input power constraint, these are the only cases where the anytime capacity has been computed.

Consider the special case of an i.i.d. rate process \( R_k \) where \( R_k \sim R \) has probability mass function \( p_i = P\{R = \bar{r}_i\}, \bar{r}_i \in \mathcal{R} \). For \( t \) real, let \( M_R(t) = \mathbb{E}(e^{tR}) \) denote the moment generating function of \( R \) and let \( M_R^{-1}(y) \) denotes the inverse of the \( M_R(t) \), if it exists, i.e., \( M_R^{-1}(y) = t \) if and only if \( M_R(t) = y \).

We have the following result.

**Theorem 6.** The anytime capacity of a memoryless bit-pipe channel with rate distribution \( R \) is
\[
C_A(\alpha) = \frac{\ln 2^{-\alpha}}{M_R^{-1}(2^{-\alpha})}
\]
for \( \alpha < -\log_2 p_{11} \) if \( \bar{r}_0 = 0 \), or for any \( \alpha > 0 \) if \( \bar{r}_0 \neq 0 \).

**Proof:** In the special case of an i.i.d. rate process,
\[
P^T 2^{-mR} = (p_1, \ldots, p_n)^T (2^{-m\bar{r}_1}, \ldots, 2^{-m\bar{r}_n})
\]
is a rank-one matrix whose only nonzero eigenvalue is \( \sum_{i=1}^{n} p_i 2^{-m\bar{r}_i} = \mathbb{E}(2^{-mR}) \). Therefore, in this case
\[
R(m) = -\frac{1}{m} \log \mathbb{E}(2^{-mR}) = -\frac{1}{m} \log M_R(\ln 2). \quad (49)
\]

By combining (29) and (49), we find that \( \alpha = -\log M_R(\ln 2) \) and so
\[
m(\alpha) = -\frac{1}{\ln 2} M_R^{-1}(2^{-\alpha}), \quad (50)
\]
where \( M_R^{-1}(y) \) exists in light of property 5) in Proposition 1. Thus, by (30), (48) follows.

Theorem 6 shows that in the case the channel is memoryless, the anytime capacity can be evaluated by computing the inverse of the moment generating function of \( R \), as illustrated in the next three examples.

**Example VI.1.** Suppose that \( R \) is a binomial random variable with parameters \( k \) and \( 1 - p \). Then,
\[
M_R(t) = (p + (1 - p)e^t)^k
\]
and
\[
M_R^{-1}(y) = \ln \frac{y^{1/k} - p}{1 - p}, \quad y < p^k,
\]
and thus by (48)

\[ C_A(\alpha) = \frac{\alpha}{\alpha/k + \log_2 \left( \frac{1-p}{1-p^2\alpha/k} \right)} \]  

(53)

for \( \alpha < -k \log p \). Notice that (53) recovers (46) in the special case \( k = 1 \) in which \( R \) is Bernoulli with parameter \( 1-p \).

**Example VI.2.** Suppose that \( R \) is a Poisson random variable with parameter \( \lambda \). Then,

\[ M_R(t) = e^{\lambda(e^t-1)} \]  

(54)

and

\[ M_R^{-1}(y) = \ln(1 + 1/\lambda \ln y), \quad y > 0, \]  

(55)

and thus by (48)

\[ C_A(\alpha) = -\frac{\alpha}{\log(1 - \alpha/\lambda \ln 2)} \]  

(56)

for \( \alpha < -\lambda/\ln 2 \).

Notice in both examples \( \bar{r}_1 = 0 \), so the anytime capacity has bounded support. In the next example \( \bar{r}_1 = 1 \) and thus the anytime capacity is defined for all \( \alpha > 0 \).

**Example VI.3.** Suppose that \( R \) is a geometric random variable with parameter \( p \). Then,

\[ M_R(t) = \frac{pe^t}{1 - (1-p)e^t} \]  

(57)

for \( t < -\ln(1-t) \) and

\[ M_R^{-1}(y) = \ln \frac{y}{p + y(1-p)}, \quad y > 0, \]  

(58)

for \( y > 0 \), and thus

\[ C_A(\alpha) = \frac{\alpha}{\log((1-p) + p2^\alpha)}, \quad \alpha > 0. \]  

(59)

Plots of the anytime capacity for different memoryless bit-pipe channels are given in Fig. 4. Plots for the Binomial, Poisson, and Geometric distributions correspond to the results in Examples VI.1, VI.2, and VI.3. The plot for the uniform distribution is obtained by numerically inverting the moment generating function of the rate process.

**VII. CONCLUSION**

We presented a parametric characterization of the anytime capacity in terms of the threshold function for \( m \)-th moment stabilization of a scalar linear system controlled over a Markov time-varying digital feedback channel that depends on \( m \) and on the channel’s parameters. This parametrization allows an explicit computation of the anytime capacity of the \( r \)-bit Markov erasure channel, and of Binomial, Poisson, and Geometric memoryless rate processes. It also provides a general strategy to compute the anytime capacity of arbitrary memoryless bit-pipe channels of given rate distribution, based on the inversion of the moment generating function of the process. The operational interpretation of the stability threshold function is that of achievable communication rate, subject to a reliability constraint, and it has as extreme values the Shannon capacity, corresponding to a soft reliability constraint, and the zero-error capacity, corresponding to hard reliability. It follows that a more stringent stability requirement for the system, expressed in terms of higher moment, translates in a harder reliability constraint for communication.

Our derivation relies on some extensions of the theory of MJLS. We provide the generalization of the necessary and sufficient conditions for second moment stability of MJLS to \( m \)-th moment stability. A commonly encountered notion of stochastic stability in the MJLS literature [45] is the one of strong moment stability, according to which the MJLS (1) is stable if there exists a finite \( \alpha_m \in \mathbb{R} \) such that \( E[|Z|^m] \to \alpha_m \), as \( m \to \infty \). Clearly, if a MJLS is strongly stable, then it is also weakly stable according to our definition. It can be verified that our proof of the necessary condition in Theorem 1 extends to the case of strong moment stability, thereby establishing the necessity of (6) for the strong moment stability of (1).

In contrast, the subsampling technique used in the direct part of Theorem 1 is not suitable to prove strong moment stability because it cannot prevent undesirable fluctuations in the system dynamics between any two sampling periods. In [49] we presented a different proof technique based on the binomial expansion of \((a_kz_k + w_k)^m\) which
can be used to prove that (6) is a sufficient condition for the strong moment stability of (1). This approach however requires \( m \) to be an integer number, as opposed to the subsampling technique which can be applied to any \( m > 0 \).

Finally, we see is no difficulty, sic et simpliciter, but at the expense of a more technical treatment, to extend our results to vector systems using bit-allocation techniques outlined in [28], [29], and to disturbances of unbounded support using standard adaptive schemes [21], [28], [29].

**Appendix A**

**Auxiliary Results**

We state a series of auxiliary results that are needed in the rest of the paper. We begin with a maximum entropy result which generalizes the well-know result that the second moment of a continuous random variable \( X \) is lower bounded by its entropy power \( 2^{\text{h}(X)} \) to the case of the \( L^m \)-norm of a continuous random vector.

**Lemma 1.** Let \( X \) be a continuous \( n \)-dimensional vector-valued random variable with \( \mathbb{E}[||X||_m^n] < \infty \), where \( m \) is a positive real number. Let \( ||X||_m = (\sum_i |X_i|^m)^{1/m} \) denote the \( L^m \)-norm of \( X \). Then,

\[
\mathbb{E}[||X||_m^n] \geq \frac{n}{c_m} e^{\frac{m}{n} \text{h}(X)} \tag{60}
\]

where \( c_m = 2^{\frac{m}{1-m}} e^{\left( \Gamma(\frac{1}{m}) \right)^m} \) and \( \Gamma(x) = \int_0^\infty e^{-tx} x^{\alpha-1} dt \). Equality holds if and only if \( X_1, \ldots, X_n \) are i.i.d. \( X \), where

\[
f_X(x) = \frac{\exp(-\frac{|x|^m}{2m(1-m)/m \Gamma(\frac{1}{m})})}{2m^{(1-m)/m} \Gamma(\frac{1}{m})}, \quad x \in \mathbb{R}. \tag{61}
\]

**Proof:** First, notice that by the maximum entropy theorem [50, Theorem 12.1.1], the unique density \( f_X(x) \) that maximizes the differential entropy \( \text{h}(X) \) over all probability densities \( f \) having \( m \)-th moment equal to \( \mathbb{E}[|X|^m] \) is given in (61). Notice that if \( X \sim f_X(x) \), then \( \text{h}(X) = \frac{1}{m} \log[c_m \mathbb{E}[|X|^m]] \). Next, by using the fact that conditioning reduces the entropy,

\[
\text{h}(X) \leq \sum_{i=1}^n \text{h}(X_i)
\]

\[
\leq \sum_{i=1}^n \frac{1}{m} \log[c_m \mathbb{E}[|X_i|^m]]
\]

\[
\leq \frac{n}{m} \log \left[ \frac{c_m}{n} \sum_{i=1}^n \mathbb{E}[|X_i|^m] \right]
\]

\[
= \frac{n}{m} \log \left[ \frac{c_m}{n} \mathbb{E}[||X||_m^m] \right]
\]

where the second inequality follows from the maximum entropy theorem applied to each random variable \( X_i \), while the last inequality follows from Jensen’s inequality.

It should be remarked that inequality (60) with \( n = 1 \) is a special case of a result proved in [51] relating the \( m \)-th moment and the Renyi entropy of a random variable.

Next, we state a known result on the log-convexity of the spectral radius of a nonnegative matrix which is needed to prove some of the properties of \( R(m) \).

**Lemma 2** (Friedland, Theorem 4.2 [52]). Let \( D_n \) be the set of \( n \times n \) real-valued diagonal matrices. Let \( A \) be a fixed \( n \times n \) non-negative matrix having a positive spectral radius. Define \( \phi : D_n \rightarrow \mathbb{R} \) by \( \phi(D) := \log \rho(e^{D}A) \).

Then \( \phi(D) \) is a convex functional on \( D_n \). Specifically: for every \( D_1, D_2 \in D_n \) and \( \alpha \in (0, 1) \),

\[
\phi(\alpha D_1 + (1 - \alpha) D_2) \leq \alpha \phi(D_1) + (1 - \alpha) \phi(D_2). \tag{62}
\]

Moreover, if \( A \) is irreducible and the diagonal entries of \( A \) are positive (or \( A \) is fully indecomposable) then equality holds in (62) if and only if

\[
D_1 - D_2 = cI \tag{63}
\]

for some \( c \in \mathbb{R} \), where \( I \) is the identity matrix.

Finally, we show a result on the asymptotic behavior of \( P^T 2^{-mR} \) as \( m \to \infty \).

**Lemma 3.** The following equality holds

\[
\lim_{m \to 0} (P^T 2^{-mR})^{\frac{1}{m}} = \lim_{m \to 0} \left( \begin{array}{cccc}
\pi_1 2^{-m\bar{r}_1} & \cdots & \pi_1 2^{-m\bar{r}_n} \\
\vdots & \ddots & \vdots \\
\pi_n 2^{-m\bar{r}_1} & \cdots & \pi_n 2^{-m\bar{r}_n}
\end{array} \right)^{\frac{1}{m}}
\]

**Proof:** Let \( 1/m = k \). By the monotonicity property of \( R(m) \), it is sufficient to prove the claim for \( k \) integer. Therefore, let

\[
A := \begin{bmatrix}
p_{11} 2^{-\bar{r}_1/k} & \cdots & p_{1n} 2^{-\bar{r}_n/k} \\
\vdots & \ddots & \vdots \\
p_{n1} 2^{-\bar{r}_1/k} & \cdots & p_{nn} 2^{-\bar{r}_n/k}
\end{bmatrix}
\]

and

\[
B := \begin{bmatrix}
\pi_1 2^{-\bar{r}_1/k} & \cdots & \pi_n 2^{-\bar{r}_1/k} \\
\vdots & \ddots & \vdots \\
\pi_1 2^{-\bar{r}_n/k} & \cdots & \pi_n 2^{-\bar{r}_n/k}
\end{bmatrix}
\]

To establish the claim it suffices to show that

\[
\lim_{k \to \infty} A^k = \lim_{k \to \infty} B^k.
\]

To do so, we prove that

\[
\lim_{k \to \infty} [A^k]_{i,j} = \lim_{k \to \infty} [B^k]_{i,j}.
\]

Note that

\[
[A^k]_{i,j} = \sum_{l_1, \ldots, l_{k-1}} (p_{i,l_1} 2^{-l_1/k}) \cdots (p_{l_{k-1},l_j} 2^{-l_{k-1}/k})
\]
\[ \sum_{l_1, \ldots, l_{k-1}} (p_{l_1} \cdots p_{l_{k-1}}) 2^{-(i+\cdots+l_{k-1})/k} \]

and

\[
[B^k]_{i,j} = \sum_{l_1, \ldots, l_{k-1}} (\pi_{l_1} 2^{-i/k}) \cdots (\pi_{l_j} 2^{-l_{k-1}/k})
\]

Then, denoting the stationary distribution of \( P \) as \( \Pi = \lim_{k \to \infty} P^k \)
\[
\lim_{k \to \infty} ([A^k]_{i,j} - [B^k]_{i,j}) = \lim_{k \to \infty} \sum_{l_1, \ldots, l_{k-1}} (p_{l_1} \cdots p_{l_{k-1}} - \pi_{l_1} \cdots \pi_{l_j}) 2^{-(i+\cdots+l_{k-1})/k}
\]
\[
\leq \lim_{k \to \infty} \sum_{l_1, \ldots, l_{k-1}} (p_{l_1} \cdots p_{l_{k-1}} - \pi_{l_1} \cdots \pi_{l_j})
\]
\[
= \left( \lim_{k \to \infty} \sum_{l_1, \ldots, l_{k-1}} p_{l_1} \cdots p_{l_{k-1}} \right) - \left( \lim_{k \to \infty} \sum_{l_1, \ldots, l_{k-1}} \pi_{l_1} \cdots \pi_{l_j} \right)
\]
\[
= \left( \lim_{k \to \infty} [P^k]_{i,j} \right) - \lim_{k \to \infty} \left[ \Pi^k \right]_{i,j}
\]
\[
= 0.
\]

Appendix B

Proof of Theorem 2

Necessity. To establish the necessary condition, we prove that for every \( k = 0, 1, \ldots \),
\[ c_m \mathbb{E} [\|X_k\|^m] \geq \mathbb{E} [\|Z_k\|^m], \]
(64)
where \( c_m \) is a constant defined in Lemma 1 and \( \{ z_k \} \) is a homogeneous MJLS with dynamics
\[ z_{k+1} = \frac{|\lambda|}{2r_k} z_k \]
and \( z_0 = e^{h(x_0)} \).

Let \( S^k = \{ S_0, \ldots, S_k \} \) denote the symbols transmitted over the digital link up to time \( k \). By the law of total expectation and Lemma 1,
\[ \mathbb{E} (|X_{k+1}|^m) \geq \frac{1}{c_m} \mathbb{E}^k (e^{mh(x_{k+1}^i|S^k=s^k)}) \]
(65)

It follows that the \( m \)-th moment of the state is lower bounded by the average entropy power of \( X_k \) conditional on \( S^k \). From the translation invariance property of the differential entropy, the conditional version of entropy power inequality [53], and Assumption of finite differential entropy of the disturbance, it follows that
\[ E_{S^k} (e^{mh(x_{k+1}^i|S^k=s^k)}) = E_{S^k} (e^{mh(\lambda x_k + \hat{x} + z_k|S^k=s^k)}) \geq E_{S^k} (e^{2h(x_k|S^k=s^k)}) \]
\[ \geq E^k_{S^k} (e^{mh(\lambda x_k|S^k=s^k)}) \]
\[ = |\lambda|^m E^k_{S^k} (e^{mh(x_k|S^k=s^k)}) \] (67)

We can further lower bound (67) as in Lemma 1 in [28], and obtain that for every \( k \geq 0 \)
\[ E_{S^k} (e^{mh(x_k|S^k=s^k)}) \geq \frac{1}{2mR_k} e^{mh(x_k|S^{k-1}=s^{k-1})}, \]
(68)
with \( S_{-1} = 0 \). By the tower rule of conditional expectation, it then follows that
\[ E_{S^k} (e^{mh(x_k|S^k=s^k)}) \geq E_{S^{k-1},R_k} \left( \frac{1}{2mR_k} e^{mh(x_k|S^{k-1}=s^{k-1})} \right). \]
(69)

Combining (69) and (67) gives
\[ E_{S^k} (e^{mh(x_k|S^k=s^k)}) \geq E_{S^{k-1},R_k} \left( \frac{|\lambda|^m}{2mR_k} E_{S^{k-1},R_k} (e^{mh(x_k|S^{k-1}=s^{k-1})}) \right). \]
(70)

Following similar steps and using the Markov chain \( S^{k-1} \to (S^{k-2}, R_{k-1}) \to R_k \), we obtain
\[ E_{S^{k-1},R_k} (e^{mh(x_k|S^{k-1}=s^{k-1})}) \geq |\lambda|^m E_{S^{k-1},R_k} (e^{mh(x_{k-1}|S^{k-1}=s^{k-1})}) \]
\[ \geq E_{S^{k-2},R_{k-1},R_k} \left( \frac{|\lambda|^m}{2mR_{k-1}} e^{mh(x_{k-1}|S^{k-2}=s^{k-2})} \right) \]
\[ = E_{R_{k-1},R_k} \left( \frac{|\lambda|^m}{2mR_{k-1}} E_{S^{k-2}|R_{k-1}} (e^{mh(x_{k-1}|S^{k-2}=s^{k-2})}) \right). \]
(71)

Substituting (71) into (70) and re-iterating \( k \) times, it follows that \( E_{S^k} (e^{mh(x_k|S^k=s^k)}) \) is lower bounded by
\[ E_{R_{k-1},R_k} \left( \frac{|\lambda|^m}{2m(R_{k-1} + R_k)} \right) \times E_{S^{k-2}|R_{k-1}} (e^{mh(x_{k-1}|S^{k-2}=s^{k-2})}) \]
\[ \geq E_{R_{k-2},\ldots,R_k} \left( \frac{|\lambda|^mk}{2m(R_{k-2} + \cdots + R_k)} e^{mh(x_{k-1}|S_{0}=s_{0})} \right) \]
(72)
\[ = E \left( \frac{|\lambda|^m(k+1)}{2m(R_{k-1} + \cdots + R_k)} \right) e^{mh(x_0)}, \]
(73)
where (72) uses the fact that the initial condition of the state \( x_0 \) is independent of the rate process \( R_k \). Let \( \{ Z_k \} \) be a non-homogeneous MJLS with dynamics
\[ z_{k+1} = \frac{|\lambda|}{2r_k} z_k, \]
(74)
with \( z_0 = e^{h(x_0)} \). By taking the expectation on both sides of (74) and iterating \( k \) times, it is easy to see that the right hand side of (73) is equal to the \( m \)th moment of \( Z_{k+1} \). Hence, combining (66)–(73) we conclude that \( E(|X_k|^m) > \frac{1}{e^m} E(|Z_k|^m) \), which is the claim.

**Sufficiency.** Let \( \omega \) denote a uniform bound on the measure of the support of the initial condition, the noise, and the disturbance. Then, we claim that the plant dynamics can be bounded as follows

\[
|x_k| \leq z_k, \quad k = 0, 1, \ldots, \tag{75}
\]

where \( Z_k \) is a homogeneous MJLS with dynamics

\[
z_{k+1} = \frac{\lambda}{2R^n} z_k + \omega, \quad k = 0, 1, \ldots
\]

and \( z_0 = \omega \). To see this, consider the following inductive proof. By assumption, \( |x_0| \leq \omega = z_0 \). Assume that the claim holds for all times up to \( k \), so \( |x_k| \leq z_k \). Suppose that at time \( k \) the uncertainty set \([-z_k, z_k]\) is quantized using a \( r_k \)-bit uniform quantizer, and that the encoder communicates to the decoder the interval containing the state. Then, the decoder approximates the state by the centroid \( \hat{x}_k \) of this interval and sends to the plant the control input \( u_k = -\lambda \hat{x}_k \). By construction \( |x_k - \hat{x}_k| \leq z_k/2^k \), thus

\[
|x_{k+1}| = |\lambda(x_k - \hat{x}_k) + v_k| \\
\leq |\lambda||x_k - \hat{x}_k| + \omega \\
\leq |\lambda|/2^k z_k + \omega \\
= z_{k+1}
\]

(76)
i.e., the claim holds at time \( k + 1 \) as well. It follows that \( x_k \) is \( m \)-th moment stable if the homogeneous MJLS \( Z_k \) is weakly \( m \)-th moment stable, i.e., if and only if (21) holds.

**APPENDIX C**

**PROOF OF PROPOSITION 1**

The monotonicity property is an immediate consequence of the log-convexity of the spectral radius of a nonnegative matrix stated in Lemma 2. Let \( D_1 = -mR \ln 2 \), \( D_2 = 0_{n \times n} \) and \( \alpha = \frac{\alpha}{m} \). Notice that \( 2^{-mR} = e^{D_1} \), \( \log(PT^{\alpha D_1}) = 0 \), and \( PT^{2-mR} = PT^{\alpha D_1} \). Then, for every \( n < m \), by Lemma 2

\[
-nR(m) = \frac{n}{m} \log \rho(PT^{2-mR}) \\
= \alpha \log \rho(PT^{\alpha D_1}) + (1 - \alpha) \log \rho(PT^{D_2}) \\
> \log \rho(PT^{\alpha D_1} + (1 - \alpha)D_2) \\
= \log \rho(PT^{2-mR}) \\
= -nR(n).
\]

(77)

By dividing both sides by \(-n\), it then follows that \( R(n) > R(m) \), establishing that \( R(m) \) is a strictly decreasing function of \( m \).

To establish the convergence of \( R(m) \) to the Shannon capacity as \( m \to 0 \), observe that

\[
\lim_{m \to 0} R(m) \\
= \lim_{m \to 0} \log \rho((PT^{2-mR})^{1/m}) \\
= \log \rho(\rho(PT^{2-mR})^{1/m}) \\
= \log \rho_P(\rho(PT^{2-mR})^{1/m}) \\
\sim \Gamma - \frac{1}{m} \log p_{11},
\]

where (a) follows from Lemma 3 and (b) follows from l'Hôpital’s rule.

The convergence of \( R(m) \) to the zero-error capacity as \( m \to \infty \) can be proved using the fact that, as \( m \to \infty \)

\[
R(m) = -\frac{1}{m} \log \rho(PT^{2-mR}) \\
= -\log \rho(PT^{2-mR})^{1/m} \\
\sim \Gamma - \frac{1}{m} \log p_{11},
\]

where the last equation follows from the fact that for any column vector \( v = (v_1, \ldots, v_n)^T \), \( \rho\left[\begin{array}{c}v \\ 0_{n \times (n-1)}\end{array}\right] = v_1 \). From the above asymptotic approximation it immediately follows that if \( \Gamma = 0 \), then

\[
\lim_{m \to \infty} mR(m) = -\log p_{11}.
\]

(78)

Next, the monotonicity property of \( -mR(m) = \log \rho(PT^{2-mR}) \) follows from the fact that all the entries of the matrix \( PT^{2-mR} \) are monotonically decreasing in \( m \). Finally, the strict concavity of \( mR(m) \) is again a direct consequence of the log-convexity of the spectral radius of a nonnegative matrix stated in Lemma 2.
REFERENCES


Paolo Minero (M’11) received the Laurea degree (with highest honors) in electrical engineering from the Politecnico di Torino, Turino, Italy, in 2003, the M.S. degree in electrical engineering from the University of California at Berkeley in 2006, and the Ph.D. degree in electrical engineering from the University of California at San Diego in 2010. He is currently a Staff Engineer in the modem system engineering group at Qualcomm Inc, San Diego, CA, where he is engaged in the design and development of both 4G and 5G wireless technology. Prior to joining Qualcomm, he was an Assistant Professor in the Department of Electrical Engineering at the University of Notre Dame, IN, conducting research on information theory and control. Dr. Minero received the U.S. Vodafone Fellowship in 2004 and 2005 and the Shannon Memorial Fellowship in 2008.

Massimo Franceschetti (M’98-SM’11) received the Laurea degree (with highest honors) in computer engineering from the University of Naples, Naples, Italy, in 1997, the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, CA, in 1999, and 2003, respectively. He is Professor of Electrical and Computer Engineering at the University of California at San Diego (UCSD). Before joining UCSD, he was a postdoctoral scholar at the University of California at Berkeley for two years. He has held visiting positions at the Vrije Universiteit Amsterdam, the École Polytechnique Fédérale de Lausanne, and the University of Trento. His research interests are in physical and information-based foundations of communication and control systems. He is co-author of the book “Random Networks for Communication” published by Cambridge University Press. Dr. Franceschetti served as Associate Editor for Communication Networks of the IEEE Transactions on Information Theory (2009 – 2012) and as Guest Associate Editor of the IEEE Journal on Selected Areas in Communications (2008, 2009). He is currently serving as Associate Editor of the IEEE Transactions on Control of Network Systems and the IEEE Transactions on Network Science and Engineering. He was awarded the C. H. Wilts Prize in 2003 for best doctoral thesis in electrical engineering at Caltech; the S.A. Schelkunoff Award in 2005 for best paper in the IEEE Transactions on Antennas and Propagation, a National Science Foundation (NSF) CAREER award in 2006, an Office of Naval Research (ONR) Young Investigator Award in 2007, the IEEE Communications Society Best Tutorial Paper Award in 2010, and the IEEE Control theory society Ruberti young researcher award in 2012.