

Fundamental Limitations of Performance in the Presence of Finite Capacity Feedback

Nuno C. Martins and Munther A. Dahleh

Abstract—This paper addresses a fundamental limitation of performance for feedback systems, in the presence of a communication channel. The feedback loop comprises a discrete-time, linear and time-invariant plant, a channel, an encoder and a decoder which may also embody a controller. Measurements of the plant’s output must be encoded for transmission over the channel. Information, at the other end of the channel, is decoded and used to generate a control signal, which is additively disturbed by a Gaussian and stationary stochastic process. We derive an inequality of the form $L_- \geq \sum \max\{0, \log(|\lambda_i(A)|)\} - C_{channel}$, where L_- is a measure of disturbance rejection, A is the open loop dynamic matrix and $C_{channel}$ is the Shannon capacity of the channel. Our measure L_- is non-negative and smaller L_- indicates better rejection (attenuation), while $L_- = 0$ signifies no rejection. Previous results show that $C_{channel} > \sum \max\{0, \log(|\lambda_i(A)|)\}$ is a necessary condition for stability and now we show that the extra rate $C_{channel} - \sum \max\{0, \log(|\lambda_i(A)|)\}$ determines a fundamental limitation for disturbance rejection. Additionally, we prove that, under a stationarity assumption, L_- admits a log-sensitivity integral representation. We contrast our condition with Bode’s integral formula and the water-bed effect. The new inequality shows explicitly how the capacity of the channel limits closed loop performance.

I. INTRODUCTION

Motivated by applications, such as remote feedback, control in the presence of information constraints has received considerable attention. Certainly, the exploration of such problems is exciting as they foster the interaction between the disciplines of Information Theory and Control.

So far, research in this field has, primarily, directed its attention to stabilization [18]. The basic framework is depicted in Fig 1 and comprises a plant, a channel, an encoder and a decoder, which implicitly embeds a controller. Measurements of the plant’s output must be encoded and sent through the channel. The information, received at the other end of the channel, is decoded and used to generate a control signal. It has been shown that stabilization, of a linear and time-invariant plant, requires [15], [16], [11] that $C_{channel}$, the channel’s Shannon capacity, is larger than $\sum \max\{0, \log(|\lambda_i(A)|)\}$, where A is the dynamic matrix of the state-space representation of the plant. For certain channels, the condition $C_{channel} > \sum \max\{0, \log(|\lambda_i(A)|)\}$ is sufficient for stabilization in the almost sure sense [16], but it may not suffice for moment stability [14]. In general, moment stability necessitates a more informative notion of capacity, designated as Anytime Capacity [14]. Stabilization of nonlinear systems has also been studied by [12] and [7].

Nuno C. Martins and Munther A. Dahleh are at LIDS, Dept. of EECS, Massachusetts Institute of Technology nmartins@mit.edu, dahleh@mit.edu

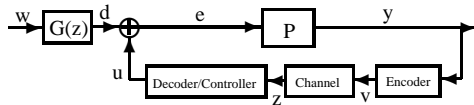


Fig. 1. Structure of the Feedback Interconnection

The work by [4] has used the integral of the log-sensitivity, as seen by the noise in an additive Gaussian channel, to establish that the optimal encoding/decoding scheme can be constructed using standard optimal control theory. Another recent area of investigation is the analysis in the presence of disturbances and uncertainty. In [9], stability in the presence of disturbances and operator theoretic uncertainty is investigated, for a particular class of channels.

Understanding the fundamental limitations of performance in a feedback system is critical for effective control design. One of the most well known trade-offs is the water-bed effect for linear feedback systems, which results from Bode’s integral formula [1]. In such classical theory, the transfer function, between the disturbance \mathbf{d} and $\tilde{\mathbf{e}} = \tilde{\mathbf{u}} + \mathbf{d}$ (see Fig 1), is denoted as sensitivity and is represented by $S(z)$. Bode’s result, for a strictly proper loop gain, is expressed as:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [\log |S(e^{j\omega})|]_- d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} [\log |S(e^{j\omega})|]_+ d\omega = \sum \max\{0, \log(|\lambda_i(A)|)\} \quad (1)$$

where $[\log |S(e^{j\omega})|]_- = \min\{0, \log |S(e^{j\omega})|\}$ and $[\log |S(e^{j\omega})|]_+ = \max\{0, \log |S(e^{j\omega})|\}$. It implies that sensitivity can’t be small at all frequencies, i.e., reduction of $\int_{-\pi}^{\pi} [\log |S(e^{j\omega})|]_- d\omega$ is achieved at the expense of increase in $\int_{-\pi}^{\pi} [\log |S(e^{j\omega})|]_+ d\omega$.

Recent publications [5], [21] have provided new versions of (1). The work by [21] has introduced a Bode-like integral inequality for non-linear systems, which is derived based on information theoretic principles.

In this paper, we derive a fundamental limitation that arises when the directed information rate¹ [10], [16], denoted by $\bar{I}_{\infty}(\mathbf{v} \rightarrow \mathbf{z})$, at the channel, is upper-bounded by a constant, i.e., $\bar{I}_{\infty}(\mathbf{v} \rightarrow \mathbf{z}) \leq C_{channel}$. Our results show that the following must hold:

$$\frac{1}{2} L_- + \bar{I}_{\infty}(\mathbf{v} \rightarrow \mathbf{z}) \geq \sum \max\{0, \log(|\lambda_i(A)|)\}$$

¹This quantity is represented as $\bar{I}_{\infty}(\mathbf{v} \rightarrow \mathbf{z})$ and will be precisely defined in section II.

where L_- is a measure of disturbance rejection. Such measure satisfies $L_- \leq 0$, where $L_- = 0$ means no-rejection and small L_- attests disturbance attenuation. We show that, under stationarity assumptions, L_- becomes an integral and our condition can be expressed as:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [\log |S(e^{j\omega})|]_- d\omega + \bar{I}_\infty(\mathbf{v} \rightarrow \mathbf{z}) \geq \sum \max\{0, \log(|\lambda_i(A)|)\} \quad (2)$$

By means of an argument similar to the water-bed effect, the inequality (2) asserts that attenuation, when measured by $\int_{-\pi}^{\pi} [\log |S(e^{j\omega})|]_- d\omega$, has to be repaid by a higher information rate in the channel. Since $\bar{I}_\infty(\mathbf{v} \rightarrow \mathbf{z}) \leq C_{channel}$, we infer that the trade-off (2) creates a fundamental limitation.

Using information theoretic arguments and assuming stationarity, we also derive the Bode integral formula. Our derivations require a linear and time-invariant plant, but the encoder, the channel and the decoder/controller can be any causal operators. The paper is organized in 4 sections. Besides the introduction, section II lays down the problem formulation as well as a preview and a discussion of the results; the limitations resulting from causality are derived in section III and section IV develops a fundamental limitation that results from finite capacity feedback.

A. *The following notation is adopted:*

- Whenever it is clear from the context, we refer to a sequence $\{a(k)\}_{-\infty}^{\infty}$ of elements in \mathbb{R}^n as a . A finite segment of a sequence a is indicated as $a_{k_{min}}^{k_{max}} = \{a(k)\}_{k_{min}}^{k_{max}}$. If $k_{max} < k_{min}$ then $a_{k_{min}}^{k_{max}} = \emptyset$.
- If M is a matrix then the element in the i -th row and j -th column is indicated as $[M]_{i,j}$. Similarly, if $a \in \mathbb{R}^n$ then $[a]_i$ denotes the i -th component of the vector.
- Random variables are represented using boldface letters, such as \mathbf{a} .
- If $\mathbf{a}(k)$ is a stochastic process, then we use $a(k)$ to indicate a specific realization. Similar to the convention used for sequences, we may denote $\mathbf{a}(k)$ just as \mathbf{a} and $a(k)$ as a . A finite segment of a stochastic process is indicated as $\mathbf{a}_{k_{min}}^{k_{max}}$.
- The probability density of a random variable \mathbf{a} , if it exists, is denoted as p_a . The conditional probability, given \mathbf{b} , is indicated as $p_{a|\mathbf{b}}$.
- The expectation operator over \mathbf{a} is written as $\mathcal{E}[\mathbf{a}]$
- We write $\log_2(\cdot)$ simply as $\log(\cdot)$
- We adopt the convention $0 \log 0 = 0$
- The auto-covariance function of a given stochastic process \mathbf{a} is given by:

$$R_a(k, l) = \mathcal{E} [(\mathbf{a}(k) - \mathcal{E}[\mathbf{a}(k)])(\mathbf{a}(l) - \mathcal{E}[\mathbf{a}(l)])^T]$$

If \mathbf{a} is stationary then it's power spectral density is written as

$$\hat{F}_a(\omega) = \sum_{k=-\infty}^{\infty} R_a(k, 0) e^{-i\omega k}$$

- If \mathbf{a} is a stochastic process taking values in \mathbb{R} then we use the following covariance matrix:

$$\left[\Sigma \left(\mathbf{a}_{k_{min}}^{k_{max}} \right) \right]_{(i-k_{min}+1), (j-k_{min}+1)} = \mathcal{E} [(\mathbf{a}(i) - \mathcal{E}[\mathbf{a}(i)])(\mathbf{a}(j) - \mathcal{E}[\mathbf{a}(j)])] \quad (3)$$

where $i, j \in \{k_{min}, \dots, k_{max}\}$.

- The Singular Value Decomposition of a matrix $M = M^H \geq 0$ is indicated as $M = V_M^T \Lambda_M V_M$, where the usual ordering of singular values is assumed $[\Lambda_M]_{i+1, i+1} \leq [\Lambda_M]_{i, i}$. The singular values of M are represented in a more streamlined form as $\lambda_i(M) = [\Lambda_M]_{i, i}$. If A is a square matrix, we also represent its eigenvalues as $\lambda_i(A)$.
- If $a \in \mathbb{R}$ then we define the negative and positive parts of a as $[a]_- = \min\{a, 0\}$ and $[a]_+ = \max\{a, 0\}$, respectively.
- The following is a shorthand notation for the log-density of the eigenvalues with magnitude smaller than 1, of a covariance matrix:

$$L_-(\mathbf{a}_{k_{min}}^{k_{max}}) = \frac{\sum_{i=1}^{k_{max}-k_{min}+1} [\log(\lambda_i(\Sigma(\mathbf{a}_{k_{min}}^{k_{max}})))]_-}{k_{max} - k_{min} + 1} \quad (4)$$

Similarly, we also define L_+ , the positive counterpart of L_- , mutatis-mutandis by replacing $-$ by $+$ in (4).

B. Basic Facts and Definitions of Information Theory

In this section, we summarize the main definitions and facts about Information Theory which are used throughout the paper. We adopt [13], as a primary reference, because it contemplates general probabilistic spaces in a unified framework. Let $(\Omega, \mathcal{S}_\omega, \mathcal{P}_\omega)$ be a probability space along with the random variables \mathbf{a} , \mathbf{b} and \mathbf{c} , taking values in the measurable spaces $(\mathcal{A}, \mathcal{S}_a)$, $(\mathcal{B}, \mathcal{S}_b)$ and $(\mathcal{C}, \mathcal{S}_c)$. We define mutual information and conditional mutual information, between any two random variables, as:

Definition 1.1: (from [13] pp. 9) The mutual information $I : (\mathbf{a}; \mathbf{b}) \rightarrow \mathbf{R}_+ \cup \{\infty\}$, between \mathbf{a} and \mathbf{b} , is given by $I(\mathbf{a}; \mathbf{b}) = \sup \sum_{i,j} \mathcal{P}_{\mathbf{a}, \mathbf{b}}(E_i \times F_j) \log \frac{\mathcal{P}_{\mathbf{a}, \mathbf{b}}(E_i \times F_j)}{\mathcal{P}_{\mathbf{a}}(E_i) \mathcal{P}_{\mathbf{b}}(F_j)}$, where the supremum is taken over all partitions $\{E_i\}$ of \mathcal{A} and $\{F_j\}$ of \mathcal{B} .

The definition of conditional mutual information can be found in [13] (pp. 37).

Notice that, in definition 1.1, \mathcal{A} and \mathcal{B} may be different. Without loss of generality, we follow [16] as we consider probability spaces which are countable or \mathbb{R}^q , for some q . We also define the following quantities, denoted as differential entropy and conditional differential entropy, which are useful in the computation of $I(\cdot, \cdot)$ for certain cases relevant in this paper.

Definition 1.2: If \mathbf{a} is a random variable with $\mathcal{A} = \mathbb{R}^q$, finite covariance matrix Σ_a and a bounded² and measurable

²Since p_a is bounded with a finite covariance matrix Σ_a it follows that $h(\mathbf{a}) < \infty$. The fact that $h(\mathbf{a}) < \infty$ further implies that $p_a \log p_a$ is integrable. Proofs of these facts use standard analysis arguments and can be found in [8]

probability density function $p_a(\cdot)$ then we define the differential entropy of \mathbf{a} as $h(\mathbf{a}) = \int_{\mathbb{R}^q} p_a(\gamma) \log p_a(\gamma) d\gamma$. If \mathbf{b} is a random variable with $\mathcal{B} = \mathbb{R}^q$ and such that $p_{a,b}(\cdot, \cdot)$ is a bounded measurable probability density function with finite covariance then we define the conditional differential entropy of \mathbf{a} given \mathbf{b} as³:

$$h(\mathbf{a}|\mathbf{b}) = h(\mathbf{a}, \mathbf{b}) - h(\mathbf{b}) = \int_{\mathbb{R}^{q'}} \int_{\mathbb{R}^q} p_{a,b}(\gamma_a, \gamma_b) \log p_{a|b}(\gamma_a, \gamma_b) d\gamma_a d\gamma_b \quad (5)$$

If \mathcal{B} is countable and $p_{a|b}(\gamma_a, b)$ is bounded and $\log p_{a|b}$ is measurable in the measure induced in $\mathcal{A} \times \mathcal{B}$ then $h(\mathbf{a}|\mathbf{b})$ is defined as:

$$h(\mathbf{a}|\mathbf{b}) = \sum_{\gamma_b \in \mathcal{S}_b} \int_{\mathbb{R}^q} p_{a,b}(\gamma_a, \gamma_b) \log p_{a|b}(\gamma_a, \gamma_b) d\gamma_a \quad (6)$$

Likewise, the quantity $h(\mathbf{a}|\mathbf{b}, \mathbf{c})$ is defined by incorporating another sum over \mathcal{S}_c , if \mathcal{C} is discrete, or an integral if $\mathcal{C} = \mathbb{R}^{q''}$. Notice that the quantity defined in (6) may not be bounded (it can be $-\infty$) because the integrand is not necessarily integrable/summable. In the more general case, if we write $h(\mathbf{a}|\mathbf{b})$ then we assume that $p_{a|b}$ is bounded and that $\log p_{a|b}$ is integrable with respect to the probability measure induced in $\mathcal{A} \times \mathcal{B}$. A more rigorous treatment of this technicality can be found in Chapter 3 of [8].

Using Theorem 2.1.2 of [13], we know that if $\log p_a$ and $\log p_{a|b}$ are integrable with respect to the probability measure induced in $\mathcal{A} \times \mathcal{B}$ then we can compute $I(\mathbf{a}; \mathbf{b})$ as:

$$I(\mathbf{a}; \mathbf{b}) = h(\mathbf{a}) - h(\mathbf{a}|\mathbf{b}) \quad (7)$$

In this paper, if we use (7) then, implicitly, we assume that $\log p_a$ and $\log p_{a|b}$ are integrable with respect to the probability measure induced in $\mathcal{A} \times \mathcal{B}$.

The following is a list of properties used in the sections III and IV. The proof of such properties may be found in [13] and, in some cases, in [2]: **(P1)**: $I(\mathbf{a}; \mathbf{b}) = I(\mathbf{b}; \mathbf{a}) \geq 0$ and $I(\mathbf{a}; \mathbf{b}|\mathbf{c}) \geq 0$; **(P2) Kolmogorov's formula**⁴ **(equation 3.6.6 in [13])**:

$$I((\mathbf{a}, \mathbf{b}); \mathbf{c}|\mathbf{d}) = I(\mathbf{b}; \mathbf{c}|\mathbf{d}) + I(\mathbf{a}; \mathbf{c}|\mathbf{b}, \mathbf{d})$$

(P3): Theorem 3.7.1 in [13]: If f and g are measurable functions in the appropriate probability spaces then $I(f(\mathbf{a}); g(\mathbf{b})|\mathbf{c}) \leq I(\mathbf{a}; \mathbf{b}|\mathbf{c})$ and equality holds if f and g are invertible⁵; **(P4)**: From property (P3), we conclude that $I(\mathbf{a}; (\mathbf{b}, \mathbf{c})|\mathbf{d}) = I(\mathbf{a}; (\mathbf{b} - \mathbf{c}, \mathbf{c})|\mathbf{d})$. Using (P2), such equality also leads to $I(\mathbf{a}; \mathbf{b}|\mathbf{c}, \mathbf{d}) = I(\mathbf{a}; \mathbf{b} - \mathbf{c}|\mathbf{c}, \mathbf{d})$; **(P5)**: By means of (P1) and (7), we infer that $h(\mathbf{a}) \geq h(\mathbf{a}|\mathbf{b})$, where equality holds if and only if \mathbf{a} and \mathbf{b} are independent. Likewise, we can use properties (P1)-(P2) to

³Notice that the equalities bellow hold because all the integrands are integrable

⁴Notice that equation 3.6.3 in [13] has a typographic mistake. On the left hand side of the equality, the correct is $I(\xi, \zeta)$

⁵In [13] equality is guaranteed for everywhere dense f and g . Everytime we say that a function is invertible in this context we are implicitly assuming that it is everywhere dense.

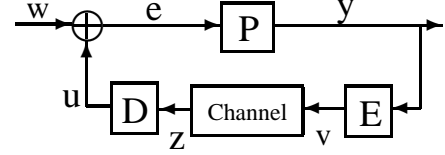


Fig. 2. Simplified Structure of the Feedback Interconnection using $e = G^{-1}\tilde{e}$ and $u = G^{-1}\tilde{u}$.

state that $I(\mathbf{a}; (\mathbf{b}, \mathbf{c})) \geq I(\mathbf{a}; \mathbf{b})$, which can be used with (7) to derive $h(\mathbf{a}|\mathbf{b}) \geq h(\mathbf{a}|\mathbf{b}, \mathbf{c})$; **(P6)**: Using a change of variables in the integrals of definition 1.2, we reckon that if $f: \mathcal{B} \rightarrow \mathcal{A}$ is any given function then $h(\mathbf{a}|\mathbf{b}) = h(\mathbf{a} - f(\mathbf{b})|\mathbf{b})$; **(P7) [2]**: If \mathbf{a} has a finite covariance matrix Σ_a then $h(\mathbf{a}) \leq \frac{1}{2} \log((2\pi e)^n \det(\Sigma_a))$.

In order to simplify our notation, we also define the following quantities:

Definition 1.3: Let \mathbf{a} and \mathbf{b} be stochastic processes. The following are useful limit information rates:

$$\bar{I}_\infty(\mathbf{a}; \mathbf{b}) = \limsup_{k \rightarrow \infty} \frac{I(\mathbf{a}_1^k; \mathbf{b}_1^k)}{k}$$

$$\bar{I}_\infty(\mathbf{a} \rightarrow \mathbf{b}) = \limsup_{k \rightarrow \infty} \frac{I(\mathbf{a}_1^k \rightarrow \mathbf{b}_1^k)}{k}$$

where directed mutual information is defined as [10], [16]:

$$I(\mathbf{a}_1^k \rightarrow \mathbf{b}_1^k) = \sum_{i=1}^k I(\mathbf{a}_1^i; \mathbf{b}(i)|\mathbf{b}_1^{i-1})$$

II. PROBLEM FORMULATION AND DISCUSSION OF RESULTS

Consider the feedback interconnection depicted in Figure 1. In such information pattern [17], measurements of the state of the plant have to be encoded and sent over a communication channel. The transmitted information is used, at the decoder/controller, to generate the control signal \mathbf{u} . In order to make the paper more comprehensible, we proceed with the equivalent block diagram of Fig 2.

A. General Assumptions

In the present formulation, the following assumptions are made: the process \mathbf{w} , with $w(k) \in \mathbb{R}$, is an i.i.d., zero mean, unit variance, white and Gaussian process; \mathbf{e} is a scalar ($e(k) \in \mathbb{R}$) stochastic process for which $\mathbf{e}_{k_{min}}^{k_{max}}$ has a probability density function, for every finite k_{min}, k_{max} ; $G(z)$ is an all-pole stable filter of the form $G(z) = \frac{\alpha}{1 - \sum_{m=1}^p a_m z^{-m}}$ for some integer $p \geq 1$ and constants a_i and $\alpha > 0$. Given n , P is a single input plant with state $\mathbf{x}(k) \in \mathbb{R}^n$, which satisfies the following state-space equation:

$$\mathbf{x}(k+1) = \begin{bmatrix} x_u(k+1) \\ x_s(k+1) \end{bmatrix} = \begin{bmatrix} A_u & 0 \\ 0 & A_s \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} b_u \\ b_s \end{bmatrix} \mathbf{e}(k) \quad (8)$$

$$\mathbf{y}(k) = C\mathbf{x}(k), |\lambda_i(A_u)| \geq 1, |\lambda_i(A_s)| < 1 \text{ and } k \geq 0$$

The state partitions \mathbf{x}_u and \mathbf{x}_s represent the unstable and stable open-loop dynamics, respectively. In addition, if $A \neq A_s$ then $\mathbf{x}_u(1)$ is a random variable with a probability density function and $|h(\mathbf{x}_u(1))| < \infty$.

In this paper, we will also refer to Channels which are stochastic operators conforming to the following definition:

Definition 2.1: (Memory-less Channel) Let \mathcal{V} and \mathcal{Z} be given input and output alphabets, along with a white stochastic process, denoted as \mathbf{c} , with alphabet \mathcal{C} . Consider $f: \mathcal{V} \times \mathcal{C} \rightarrow \mathcal{Z}$ such that the following maps are invertible:

$$\begin{aligned} g_1(v(k), c(k)) &= (v(k), f(v(k), c(k))) \\ g_2(v(k), c(k)) &= (f(v(k), c(k)), c(k)) \end{aligned}$$

The pair (f, \mathbf{c}) defines a memory-less channel. The following are examples of memory-less channels:

- **Additive white Gaussian channel:** $\mathcal{V} = \mathcal{Z} = \mathcal{C} = \mathbb{R}$, \mathbf{c} is an i.i.d. white Gaussian sequence with unit variance and $f(c, v) = c + v$.
- **Binary symmetric channel, with error probability p_e :** $\mathcal{V} = \mathcal{Z} = \mathcal{C} = \mathbb{Z}_2 = \{0, 1\}$, \mathbf{c} is an i.i.d sequence satisfying $\mathcal{P}(\mathbf{c}(k) = 1) = p_e$ and $f(c, v) = c +_{\text{mod}2} v$

B. Assumptions about the Encoder and the Decoder

We also assume the following about the encoder and the decoder:

- (A1) the encoder and the decoder are causal operators defined in the appropriate spaces, i.e., $E: \mathcal{Y}^\infty \rightarrow \mathcal{V}^\infty$, $D: \mathcal{Z}^\infty \rightarrow \mathcal{U}^\infty$ where $\mathbf{v}(k) = f_k^e(\mathbf{y}_{-\infty}^k)$ and $\mathbf{u}(k) = f_k^d(\mathbf{z}_{-\infty}^k)$ for some functions f_k^e and f_k^d .
- (A2) additionally, the decoder satisfies the following finite memory condition:

$$\forall k > \alpha, \mathbf{u}_{1+\alpha}^k = \tilde{f}_k^d(\mathbf{u}_1^\alpha, \mathbf{z}_1^k) \quad (9)$$

for some $\alpha \in \mathbb{N}_+$ and a sequence of functions $\tilde{f}_k^d: \mathcal{U}^\alpha \times \mathcal{Z}^k \rightarrow \mathcal{U}^{k-\alpha-1}$.

- (A3) (Fading memory condition) For technical reasons, we assume that the following condition holds:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{u}_1^\alpha; \mathbf{x}(1), \mathbf{w}_1^k | \mathbf{z}_1^k) = 0$$

where α is the smallest constant for which (A2) holds. If $\alpha = 0$ then we adopt the convention that (A3) is satisfied. A particular instance of $\alpha = 0$ is if the decoder is a dynamical system with zero initial conditions. Several aspects of this assumption are clarified below. More details can be found in [8].

C. Further Remarks about (A2)

Notice that a synchronous block decoder, with delay α , falls into this category. In addition, any dynamic system, of the form $\mathbf{u}(k) = f(\mathbf{u}_{k-\alpha}^{k-1}, \mathbf{z}_{k-\alpha}^k)$, will satisfy (9). We emphasize that this representation does not presuppose a full-information system. For example, if $\mathbf{y}_c(k)$ is the output of an observable n-th order linear and time-invariant system, with input $\mathbf{z}(k)$, then it is possible to represent its input-output behavior in the form $\mathbf{y}_c(k) = f(\mathbf{y}_{c, k-n}^{k-1}, \mathbf{z}_{k-n}^k)$.

1) *Assumption (A3) when \mathcal{U} is countable:* If \mathcal{U} is countable then we can use (P1)-(P2) to conclude that:

$$I(\mathbf{u}_1^\alpha; \mathbf{w}_1^k | \mathbf{z}_1^k) \leq H(\mathbf{u}_1^\alpha) \quad (10)$$

As such, if $H(\mathbf{u}_1^\alpha) < \infty$ holds then $I(\mathbf{u}_1^\alpha; \mathbf{w}_1^k | \mathbf{z}_1^k) < \infty$ is satisfied. If \mathcal{U} has \aleph_U elements, such quantity is upper-bounded [2] as $H(\mathbf{u}_1^\alpha) \leq \alpha \log(\aleph_U)$. The confinement to finite control alphabets is expected if the channel, itself, is discrete or in the presence of quantizers. Finite \mathcal{U} further encompasses digital controllers, as they constitute dynamic systems evolving on a finite precision algebra.

D. Problem Statement and Discussion of Results

We investigate the fundamental limitations of the eigenvalue distribution of $\Sigma(\mathbf{e}_{k_{\min}}^{k_{\max}})$. In order to simplify the exposé, we state our results in terms of $L_-(\mathbf{e}_{k_{\min}}^{k_{\max}})$ and $L_+(\mathbf{e}_{k_{\min}}^{k_{\max}})$.

In section III we reach a fundamental limitation which is a consequence of causality alone. The result is presented in theorem 3.3, which states that if the feedback system in Fig 2 is stable then the following must hold:

$$\frac{1}{2} \liminf_{k \rightarrow \infty} (L_-(\mathbf{e}_1^k) + L_+(\mathbf{e}_1^k)) \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (11)$$

The inequality in (11) demonstrates that not all of the eigenvalues, of $\Sigma(\mathbf{e}_1^k)$, can be made small and that the reduction of some necessarily imply the increase of others. That is comparable to the water-bed effect, associated to the classic Bode integral limitation. Such comparison is not coincidental and is explored in section III-A.

In the fundamental limitation expressed in (11), the characteristics of the channel do not play a role. It remains the question of whether the “shaping” of the eigenvalues of $\Sigma(\mathbf{e}_1^k)$ depends on the information flow in the feedback loop. The answer is given in theorem 4.3 which states that:

$$\frac{1}{2} \liminf_{k \rightarrow \infty} L_-(\mathbf{e}_1^k) + \bar{I}_\infty(\mathbf{v} \rightarrow \mathbf{z}) \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (12)$$

As a consequence of (12), we find that reduction of the eigenvalues of $\Sigma(\mathbf{e}_1^k)$, for values bellow unity, must come at the expense of information flow in the channel, as quantified by $\bar{I}_\infty(\mathbf{v} \rightarrow \mathbf{z})$.

Under stationary assumptions, corollaries 3.4 and 4.4 show that the inequalities (11) and (12) can be expressed as:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [\log(S(\omega))]_- d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} [\log(S(\omega))]_+ d\omega \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (13)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [\log(S(\omega))]_- d\omega + \bar{I}_\infty(\mathbf{v} \rightarrow \mathbf{z}) \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (14)$$

where $S(\omega) = \sqrt{\hat{F}_e(\omega)} = \sqrt{\frac{\hat{F}_e(\omega)}{\hat{F}_d(\omega)}}$.

The inequalities (13) and (14) must be satisfied by any stable and causal loop of the form depicted in Fig 1 or Fig 2.

III. FUNDAMENTAL LIMITATIONS CREATED BY CAUSALITY

In this section, we derive a fundamental limitation that results from causality. Such digression is also used to present some of the preliminary results, which will be used in section IV. Our technique follows the one by [21], with the exception of the way we tackle initial conditions and unstable modes of the plant. More specifically, theorem 3.3 states a fundamental limitation that explicitly incorporates the eigenvalues of A . At the end of the section, we specialize the result, under stationarity assumptions, and derive the Bode-Integral formula in Corollary 3.4.

The following lemma shows that the difference, between the entropy rate of \mathbf{e} and the entropy rate of \mathbf{w} , is lower-bounded by the mutual information between the initial state and \mathbf{e} .

Lemma 3.1: If $\mathbf{x}(k)$ is the solution of the state-space equation (8) then the following holds:

$$\liminf_{k \rightarrow \infty} \frac{h(\mathbf{e}_1^k)}{k} \geq \liminf_{k \rightarrow \infty} \frac{I(\mathbf{e}_1^k; \mathbf{x}(1))}{k} + h(\mathbf{w}(1)) \quad (15)$$

Proof: We start by noticing that since the plant is strictly causal and $\mathbf{x}(1)$ is independent of \mathbf{w} , we get that:

$$\begin{aligned} h(\mathbf{w}(k)) &= h(\mathbf{w}(k)|\mathbf{x}(1), \mathbf{u}_1^k, \mathbf{w}_1^{k-1}) = \\ &h(\mathbf{e}(k)|\mathbf{x}(1), \mathbf{u}_1^k, \mathbf{e}_1^{k-1}) \leq h(\mathbf{e}(k)|\mathbf{x}(1), \mathbf{e}_1^{k-1}) \end{aligned} \quad (16)$$

where we used properties (P5) and (P6). Since $h(\mathbf{w}(k))$ does not depend on k , we use (16) and the law of iterated differential entropy ⁶ to derive:

$$\sum_{i=1}^k h(\mathbf{e}(i)|\mathbf{x}(1), \mathbf{e}_1^{i-1}) = h(\mathbf{e}_1^k|\mathbf{x}(1)) \geq h(\mathbf{w}(1))$$

The proof is concluded once we notice, from (7), that $h(\mathbf{e}_1^k|\mathbf{x}(1)) = h(\mathbf{e}_1^k) - I(\mathbf{e}_1^k; \mathbf{x}(1))$. \square

The following lemma, corroborates the results by [15], [16], [14], [20], [11], and unveils that stability implies that \mathbf{e} must carry a bit-rate, of information about the initial state, of at least $\sum_i \max\{0, \log(|\lambda_i(A)|)\}$.

Lemma 3.2: Let $\mathbf{x}(k)$ be the solution of the state-space equation (8). If the system is stable, i.e., $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ holds then the following is satisfied:

$$\liminf_{k \rightarrow \infty} \frac{I(\mathbf{e}_1^k; \mathbf{x}(1))}{k} \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (17)$$

⁶Notice that the law of iterated differential entropy, equation (9.33) in [2], holds here because we assume that \mathbf{e} and $\mathbf{x}(1)$ have a joint probability density function.

Proof: If $A = A_s$ then we just use $I(\mathbf{e}_1^k; \mathbf{x}(1)) \geq 0$. If $A \neq A_s$ then we consider the following homogeneous system:

$$\mathbf{x}_e(k+1) = A_u \mathbf{x}_e(k) + b_u \mathbf{e}(k), \quad x_e(1) = 0 \quad (18)$$

and define the estimate $\hat{\mathbf{x}}(k) = A_u^{-k} \mathbf{x}_e(k)$. Since $\mathbf{x}_u(k) = \mathbf{x}_e(k) + A_u^k \mathbf{x}_u(1) = A_u^k (\hat{\mathbf{x}}(k) - \mathbf{x}_u(1))$, we know that:

$$\begin{aligned} k \log(|\det(A_u A_u^T)|) + \log(\det(R_{\mathbf{x}_{error}}(k))) = \\ \log(\det(R_{\mathbf{x}_u}(k, k))) < \beta < \infty \end{aligned} \quad (19)$$

where $\mathbf{x}_{error}(k) = \hat{\mathbf{x}}(k) - \mathbf{x}_u(1)$. Since $\hat{\mathbf{x}}(k)$ is a function of \mathbf{e}_1^k , we have that:

$$I(\mathbf{x}(1); \mathbf{e}_1^k) \geq I(\mathbf{x}_u(1); \mathbf{e}_1^k) \geq h(\mathbf{x}_u(1)) - h(\hat{\mathbf{x}}(k) - \mathbf{x}_u(1)) \quad (20)$$

But, from (P7) we know that $\limsup_{k \rightarrow \infty} \frac{h(\hat{\mathbf{x}}(k) - \mathbf{x}_u(1))}{k} \leq \limsup_{k \rightarrow \infty} \frac{\log(\det(R_{\mathbf{x}_{error}}(k)))}{2k}$. As a consequence, we can use (19) to get $\limsup_{k \rightarrow \infty} \frac{h(\hat{\mathbf{x}}(k) - \mathbf{x}_u(1))}{k} \leq -\log(|\det(A_u)|)$. The proof follows by direct substitution \square .

Using the results in the previous lemmas, we derive theorem 3.3. It states that causality and stability imply that the log-sum of the eigenvalues of $\Sigma(\mathbf{e}_1^k)$ are, in the limit, lower bounded by the unstable eigenvalues of the plant.

Theorem 3.3: (Causality fundamental limitation) Let $\mathbf{x}(k)$ be the solution of the state-space equation (8). If the system is stable, i.e., $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ holds then the following is satisfied:

$$\liminf_{k \rightarrow \infty} (L_-(\mathbf{e}_1^k) + L_+(\mathbf{e}_1^k)) \geq 2 \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (21)$$

Proof: From lemmas 3.1 and 3.2 we know that:

$$\begin{aligned} \liminf_{k \rightarrow \infty} \frac{h(\mathbf{e}_1^k)}{k} - h(\mathbf{w}(1)) \geq \liminf_{k \rightarrow \infty} \frac{I(\mathbf{e}_1^k; \mathbf{x}(1))}{k} \geq \\ \sum_i \max\{0, \log(|\lambda_i(A)|)\} \end{aligned} \quad (22)$$

Using the fact (P7), we conclude that $h(\mathbf{e}_1^k) - h(\mathbf{w}(1)) \leq \frac{1}{2} \log(\det(\Sigma(\mathbf{e}_1^k)))$ which, together with (22), leads to the final result \square

A. Deriving Bode's Integral Formula

Under stationarity assumptions, theorem 3.3 is at the base of Bode-integral formula. A precise description of such property is in the subsequent Corollary.

Corollary 3.4: Let $\mathbf{x}(k)$ be the solution of the state-space equation (8). If the system is stable, i.e., $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ holds and \mathbf{e} is a stationary process, where $0 < m < \hat{F}_e(\omega) < M < \infty$ is Lebesgue integrable, then the following is satisfied:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(S(\omega)) d\omega \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (23)$$

where $S(\omega) = \sqrt{\hat{F}_e(\omega)} = \sqrt{\frac{\hat{F}_e(\omega)}{\hat{F}_d(\omega)}} = \sqrt{\frac{\hat{F}_e(\omega)}{|G(e^{j\omega})|^2}}$. The processes $\tilde{\mathbf{e}}$ and $\tilde{\mathbf{d}}$ are the ones depicted in Fig 1.

Proof: The proof follows from Theorems 3.3 and 5.2, which implies that if \mathbf{e} is stationary, then $\lim_{k \rightarrow \infty} L(\mathbf{e}_1^k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(\hat{F}_e(\omega)) d\omega$. \square

IV. FUNDAMENTAL LIMITATIONS CREATED BY FINITE CAPACITY FEEDBACK

In this section, we examine the fundamental limitations, in the eigenvalues of $\Sigma(\mathbf{e}_1^k)$, that originate from the constraint $\bar{I}_\infty(\mathbf{v} \rightarrow \mathbf{z}) \leq C_{channel}$. The main inequality, involving the channel directed information rate and the eigenvalues of A , is given in theorem 4.3.

Subsequently, we provide a lemma which unveils how the information, traveling in the feedback loop, is allocated. By inspecting the proof of such lemma, we identify that the feedback mutual information, dictated by $I((\mathbf{x}(1), \mathbf{w}_1^k); \mathbf{u}_1^k)$, must account for two terms. The first is due to stabilization information $I(\mathbf{x}(1); \mathbf{e}_1^k)$ and the second quantifies the interaction between the control signal and the disturbance $I(\mathbf{u}_1^k; \mathbf{w}_1^k)$. In addition, the case study in [8] shows that the inequality, presented in the lemma below, is not conservative.

Lemma 4.1: (Fundamental Lemma of the Information Flux) Let $\mathbf{x}(k)$ be the solution of the state-space equation (8). If the system is stable, i.e., $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ holds then the following is satisfied:

$$\bar{I}_\infty(\mathbf{v} \rightarrow \mathbf{z}) \geq \liminf_{k \rightarrow \infty} \frac{I(\mathbf{x}(1); \mathbf{e}_1^k)}{k} + \bar{I}_\infty(\mathbf{u}; \mathbf{w}) \quad (24)$$

Proof: We start by using (P2) to write $I((\mathbf{x}(1), \mathbf{w}_1^k); \mathbf{u}_1^k) = I(\mathbf{x}(1); \mathbf{u}_1^k | \mathbf{w}_1^k) + I(\mathbf{u}_1^k; \mathbf{w}_1^k)$ which can be rewritten, by means of (P3), as:

$$I((\mathbf{x}(1), \mathbf{w}_1^k); \mathbf{u}_1^k) = I(\mathbf{x}(1); \mathbf{e}_1^k | \mathbf{w}_1^k) + I(\mathbf{u}_1^k; \mathbf{w}_1^k) \quad (25)$$

On the other hand, using (P2) we get

$$I(\mathbf{x}(1); \mathbf{e}_1^k | \mathbf{w}_1^k) = I(\mathbf{x}(1); \mathbf{e}_1^k) - I(\mathbf{x}(1); \mathbf{w}_1^k) + I(\mathbf{x}(1); \mathbf{w}_1^k | \mathbf{e}_1^k) \quad (26)$$

Since \mathbf{w} is independent from $\mathbf{x}(1)$, the second term, on the right-hand side of (26), vanishes and we resort to (P1) to get $I(\mathbf{x}(1); \mathbf{e}_1^k | \mathbf{w}_1^k) \geq I(\mathbf{x}(1); \mathbf{e}_1^k)$. Consequently, we substitute the aforementioned inequality in (25) and obtain the following:

$$I((\mathbf{x}(1), \mathbf{w}_1^k); \mathbf{u}_1^k) \geq I(\mathbf{x}(1); \mathbf{e}_1^k) + I(\mathbf{u}_1^k; \mathbf{w}_1^k) \quad (27)$$

The present lemma is proven once we resort to theorem 5.1 \square

The following lemma suggests that attenuation can happen only if the channel sends information about the disturbance.

Lemma 4.2: If $g(k)$ is a sequence such that $\frac{g(k)}{k} \xrightarrow{k \rightarrow \infty} 0$ then the following holds:

$$\frac{1}{k} I(\mathbf{u}_1^k; \mathbf{w}_1^k) \geq -\frac{1}{2} L_-(\mathbf{e}_{g(k)}^k) \quad (28)$$

Proof: Let the following be the singular value decomposition of $\Sigma(\mathbf{e}_{g(k)}^k)$:

$$\Sigma(\mathbf{e}_{g(k)}^k) = \begin{bmatrix} V_+ \\ V_- \end{bmatrix}^T \begin{bmatrix} \Lambda_+ & 0 \\ 0 & \Lambda_- \end{bmatrix} \begin{bmatrix} V_+ \\ V_- \end{bmatrix} \quad (29)$$

where $[\Lambda_-]_{ii} < 1$ and $[\Lambda_+]_{ii} \geq 1$.

We establish the following relation⁷:

$$I(\mathbf{w}_1^k; \mathbf{u}_1^k) \geq I(\mathbf{w}_{g(k)}^k; \mathbf{u}_{g(k)}^k) \geq I(V_- \mathbf{w}_{g(k)}^k; V_- \mathbf{u}_{g(k)}^k) = h(V_- \mathbf{w}_{g(k)}^k) - h(V_- \mathbf{w}_{g(k)}^k | V_- \mathbf{u}_{g(k)}^k) \geq h(V_- \mathbf{w}_{g(k)}^k) - h(V_- \mathbf{e}_{g(k)}^k) \quad (30)$$

where we have used (7), (P6) and (P5). Moreover, since \mathbf{w} is Gaussian with unit variance then (P7) guarantees that $h(V_- \mathbf{w}_{g(k)}^k) - h(V_- \mathbf{e}_{g(k)}^k) \geq -\frac{1}{2} \log(\det(V_- \Sigma(\mathbf{e}_{g(k)}^k) V_-^T)) = -\frac{k}{2} L_-(\mathbf{e}_{g(k)}^k)$. \square

Subsequently, we provide the theorem which states the main inequality in the paper. It reflects a trade-off between disturbance attenuation, as measured by $L_-(\mathbf{e}_1^k)$, and the directed information rate through the channel, expressed by $\bar{I}_\infty(\mathbf{v} \rightarrow \mathbf{z})$.

Theorem 4.3: (Main theorem) Let $\mathbf{x}(k)$ be the solution of the state-space equation (8) and $g(k)$ be an arbitrary sequence satisfying $\frac{g(k)}{k} \xrightarrow{k \rightarrow \infty} 0$. If the system is stable, i.e., $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ holds then the following is satisfied:

$$\bar{I}_\infty(\mathbf{v} \rightarrow \mathbf{z}) + \frac{1}{2} \liminf_{k \rightarrow \infty} L_-(\mathbf{e}_{g(k)}^k) \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (31)$$

Proof: The result follows by direct substitution of Lemmas 3.2 and 4.2 into Lemma 4.1. \square

The corollary below is an immediate consequence of theorem 4.3 and shows that if $C_{channel}$ is too close to the critical stabilization rate, given by $\sum_i \max\{0, \log(|\lambda_i(A)|)\}$, then disturbance rejection is not possible.

Corollary 4.4: Let $\mathbf{x}(k)$ be the solution of the state-space equation (8) and $g(k)$ be an arbitrary sequence satisfying $\frac{g(k)}{k} \xrightarrow{k \rightarrow \infty} 0$. If the system is stable, i.e., $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ holds then the following is satisfied:

$$\frac{1}{2} \liminf_{k \rightarrow \infty} L_-(\mathbf{e}_{g(k)}^k) + C_{channel} \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (32)$$

Proof: Follows from theorem 4.3 and the fact that $I(\mathbf{v}_1^k \rightarrow \mathbf{z}_1^k) \leq k C_{channel}$. \square

⁷Notice that we have used an abuse of notation in equation (30). We write $V_- \mathbf{e}_{g(k)}^k$ to indicate the random variable whose realizations are

computed as $V_- \begin{bmatrix} \mathbf{e}(k) \\ \vdots \\ \mathbf{e}(g(k)) \end{bmatrix}$

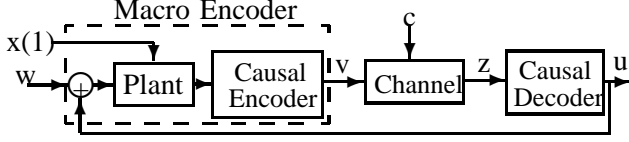


Fig. 3. Casting the feedback loop as a channel in feedback.

A. An Integral Formula

Under stationarity assumptions, the condition in theorem 4.3 can be expressed by means of an integral form.

Corollary 4.5: Let $\mathbf{x}(k)$ be the solution of the state-space equation (8). If the system is stable, i.e., $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ holds and \mathbf{e} is stationary, where $0 < m < \hat{F}_e(\omega) < M < \infty$ is Lebesgue integrable, then the following is satisfied:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [\log S(\omega)]_- d\omega + \bar{I}_{\infty}(\mathbf{v} \rightarrow \mathbf{z}) \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (33)$$

where $S(\omega) = \sqrt{\hat{F}_e(\omega)} = \sqrt{\frac{\hat{F}_{\tilde{\mathbf{e}}}(\omega)}{\hat{F}_{\tilde{\mathbf{d}}}(\omega)}} = \sqrt{\frac{\hat{F}_{\tilde{\mathbf{e}}}(\omega)}{|G(e^{j\omega})|^2}}$. The processes $\tilde{\mathbf{e}}$ and $\tilde{\mathbf{d}}$ are the ones depicted in Fig 1.

Proof: By means of a direct application of the Theorem 5.2 we find that:

$$\lim_{k \rightarrow \infty} L_-(\mathbf{e}_1^k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} [\log(\hat{F}_e(\omega))]_- d\omega \quad (34)$$

The result follows by direct substitution of (34) in (31). \square

The integral formula holds under certain asymptotic stationarity conditions. More specifically, it holds if there exists $g(k)$ such that $\Sigma(\mathbf{e}_g^k)$ converges, ‘‘sufficiently fast’’, to a stationary $\Sigma(\mathbf{e}_1^{k-g(k)})$. This extension is studied by [8] in a case study.

V. AUXILIARY RESULTS

The following theorem provides an extension of the directed data processing inequality, originally derived in [16]. Compared to the version in [16], the result presented below allows encoders and decoders that depend on past inputs indexed by $k < 1$. The quantities in the statement of the theorem refer to the scheme depicted in Fig 3.

Theorem 5.1: (Directed Data Processing Inequality) Let the following assumptions hold:

- The plant is LTI with a state-space representation where $D = 0$ (strictly proper)
- (A1) The encoder and decoder are causal operators
- (A2) The decoder satisfies:

$$\forall k > \alpha, \mathbf{u}_{\alpha+1}^k = \tilde{f}_k^d(\mathbf{u}_1^\alpha, \mathbf{z}_1^k) \quad (35)$$

for some $\alpha \in \mathbb{N}_+$ and a sequence of functions \tilde{f}_k^d .

- (A3) The fading memory condition $\limsup_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^k) = 0$ holds.

Under the above conditions, the following is true:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} I((\mathbf{x}(1), \mathbf{w}_1^k); \mathbf{u}_1^k) \leq \bar{I}_{\infty}(\mathbf{v} \rightarrow \mathbf{z}) \quad (36)$$

Proof: We separate the proof in two parts.

As a **first step** we show that $I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \leq I(\mathbf{v}_1^k \rightarrow \mathbf{z}_1^k)$.

Using (P2) we can write the following equality, for any given $i \in \{1, \dots, k\}$:

$$\begin{aligned} I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1}) | \mathbf{z}_1^{i-1}) &= I(\mathbf{z}(i); \mathbf{v}_1^i | \mathbf{z}_1^{i-1}) + \\ &I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1}) | \mathbf{z}_1^{i-1}, \mathbf{v}_1^i) - \\ &I(\mathbf{z}(i); \mathbf{v}_1^i | \mathbf{z}_1^{i-1}, \mathbf{x}(1), \mathbf{w}_1^{i-1}) \end{aligned} \quad (37)$$

Now notice that (P2) allows us to rewrite:

$$\begin{aligned} I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1}) | \mathbf{z}_1^{i-1}, \mathbf{v}_1^i) &= \\ I((\mathbf{z}_1^i, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) - I((\mathbf{z}_1^{i-1}, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) \end{aligned} \quad (38)$$

But, from (P3), we know that

$$\begin{aligned} I((\mathbf{z}_1^i, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) &= \\ I((\mathbf{c}(i), \mathbf{z}_1^{i-1}, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) \end{aligned} \quad (39)$$

where we used the fact that, from the definition 2.1 (channel), the following map is invertible:

$$(z(i), v(i)) \mapsto (c(i), v(i))$$

Causality makes $\mathbf{c}(i)$ independent of $(\mathbf{z}_1^{i-1}, \mathbf{v}_1^i, \mathbf{x}(1), \mathbf{w}_1^{i-1})$, so that (39) implies the following:

$$I((\mathbf{z}_1^i, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) = I((\mathbf{z}_1^{i-1}, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) \quad (40)$$

By making use of (40) and (38) we infer that $I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1}) | \mathbf{z}_1^{i-1}, \mathbf{v}_1^i) = 0$. Such fact, together with (P1) and (37), leads to:

$$I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1}) | \mathbf{z}_1^{i-1}) \leq I(\mathbf{z}(i); \mathbf{v}_1^i | \mathbf{z}_1^{i-1}) \quad (41)$$

The first part of the proof is concluded once we notice that, from causality, \mathbf{w}_1^k is independent of $(\mathbf{x}(1), \mathbf{w}_1^{i-1}, \mathbf{z}_1^i)$, which implies:

$$I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^{i-1}) = I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1}) | \mathbf{z}_1^{i-1}) \quad (42)$$

so that (41) implies:

$$\begin{aligned} I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) &= \sum_{i=1}^k I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1}) | \mathbf{z}_1^{i-1}) \leq \\ &\sum_{i=1}^k I(\mathbf{z}(i); \mathbf{v}_1^i | \mathbf{z}_1^{i-1}) = I(\mathbf{v}_1^k \rightarrow \mathbf{z}_1^k) \end{aligned} \quad (43)$$

In the **second step** we prove that:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \leq \limsup_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k))$$

Once again, we use (P2) to write:

$$I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) = I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) + I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^k) - I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{u}_1^k) \quad (44)$$

It follows from (P2), (P4) and assumption (A7) that:

$$I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^k) = I(\mathbf{u}_{\alpha+1}^k; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^k, \mathbf{u}_1^\alpha) + I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^k) = I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^k) \quad (45)$$

Substitution of (45) in (44), together with property (P1), leads to:

$$I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \leq I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) + I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^k) \quad (46)$$

Accordingly, (46) and the assumption (A8), which requires $\limsup_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^k) = 0$, imply that:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \leq \limsup_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \quad (47)$$

which, together with (43), concludes the proof. \square

The following is the statement of the *main theorem* of Chapter 5 of [6], repeated here for convenience:

Theorem 5.2: (Reproduced from [6], pp.64-65) Let $\hat{F}_e(\omega)$ be a real-valued function of the class \mathcal{L}_1 ($|\hat{F}_e(\omega)|$ is integrable in the sense of Lebesgue). We denote by m and M the essential lower bound and upper bound of $\hat{F}_e(\omega)$, respectively, and assume that m and M are finite. If $G(\lambda)$ is any continuous function defined in the finite interval $m \leq \lambda \leq M$, we have:

$$\lim_{k \rightarrow \infty} \frac{\sum_{i=1}^n G(\lambda_i(\Sigma(\mathbf{e}_1^k)))}{k+1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\hat{F}_e(\omega)) d\omega \quad (48)$$

VI. ACKNOWLEDGMENTS

The authors would like to thank Prof. Nicola Elia (Iowa State University) for interesting suggestions. The first author is grateful to Prof. John Doyle (Caltech) for his enthusiastic support. We also would like to thank Prof. Sanjoy Mitter for interesting discussions on related problems. This work was sponsored by the University of California - Los Angeles, MURI project title: “*Cooperative Control of Distributed Autonomous Vehicles in Adversarial Environments*”, award: 0205-G-CB222. Nuno C. Martins was partially supported by the Portuguese Foundation for Science and Technology and the European Social Fund, PRAXIS BD19630/99.

REFERENCES

- [1] Bode, H. W., “*Network Analysis and Feedback Amplifier Design*”, D. Van Nostrand, Princeton, 1945
- [2] Cover, T.M.; Thomas, J. A., “*Elements of Information Theory*”, Wiley-Interscience Publication, 1991
- [3] Doyle, J.C.; Francis, B.A.; Tannenbaum, A.R., “*Feedback Control Theory*”, Macmillan, New York, 1992
- [4] Elia, N., “*Control-Oriented feedback communication schemes*”, Allerton Conference in Communication and Control, 2003
- [5] Freudenberg, J.S.; Looze, D.P. “*Frequency Domain Properties of Scalar and Multivariable Systems*”, Springer-Berlin, 1988
- [6] Grenander, U.; Szego, G., “*Toeplitz Forms and Their Applications*”, University of California Press, 1958
- [7] Liberzon, D. “*On Stabilization of Non-Linear Systems With Limited Information Feedback*”, Proc. of the IEEE Conf. on Dec. and Control, pp. 182-6, 2003
- [8] Martins, N. C. “*Information Theoretic Aspects of the Control and Mode Estimation of Stochastic Systems*”, Ph.D. Thesis, MIT, 2004 (The work pertaining to this paper corresponds to Chapter 3 of the aforementioned Thesis. It can be downloaded from web.mit.edu/~nmartins/www/thesischap3.pdf)
- [9] Martins, N.C.; Dahleh, M. A. and Elia, N., “*Feedback Stabilization of Uncertain Stochastic Systems Using a Stochastic Digital Link*”, to appear in the IEEE Conf. on Decision and Control, 2004
- [10] Massey, J., “*Causality, Feedback and Directed Information*”, Proc. of the 1990 Int. Symp. on Information Theory and its Applications (ISITA-90), pp. 303-305
- [11] Nair, G. N. and Evans, R. J., “*Stabilization with Data-Rate-Limited Feedback: Tightest Attainable Bounds.*” Systems and Control Letters, Vol 41, pp. 49-76, 2000
- [12] Nair, G. N. and Evans, R. J., Mareels, I. M. Y. and Moran W. “*Topological Entropy and Nonlinear Stabilization*” to appear in the Special Issue on Networked Control Systems, IEEE Transactions on Automatic Control, 2004
- [13] Pinsker, M. S., “*Information and Information Stability of Random Variables and Processes*”, Holden Day, 1964
- [14] Sahai, A., “*Evaluating Channels for Control: Capacity Reconsidered*”, Proc. ACC., pp. 2358 - 2362, 2000
- [15] Tatikonda, S., “*Control under Communication Constraints: Part I and II*”, submitted to the IEEE Transactions on Automatic Control
- [16] Tatikonda, S., “*Control under Communication Constraints*”, Ph.D. Thesis, M.I.T. 2000
- [17] Witsenhausen, H., “*Separation of Estimation and Control for Discrete-Time Systems*”, Proceeding of the IEEE, Volume 59, No 11, November 1971
- [18] Wong, W.S.; Brockett, R.W.; “*Systems with finite communication bandwidth constraints -II: Stabilization with Limited Information Feedback*” IEEE Trans. Automat. Control, Vol 44, No. 5 pp. 1049-1053, 1999
- [19] Wong, W.S.; Brockett, R.W.; “*Systems with finite communication bandwidth constraints -I: State estimation problems*” IEEE Trans. Automat. Control, Vol 42, pp. 1294-1298, Sept 1997
- [20] Yuksel, S.; Basar, T. “*Quantization and Coding for Decentralized LTI Systems*”, Proc. IEEE CDC, Hawaii, December 2003
- [21] Zang, G.; Iglesias, P. A., “*Nonlinear extension of Bode’s integral based on an information theoretic interpretation*”, Systems and Control Letters, 50 (2003) pp. 11-19