# Estimation of Nonlinear Dynamic Systems over Communication Channels

Vahideh Sanjaroon, Alireza Farhadi, Abolfazl Seyed Motahari and Babak. H. Khalaj

***Abstract*—Remote observation of the state trajectory of nonlinear dynamic systems over limited capacity communication channels is studied. It is shown that two extreme cases are possible: Either the system is fully observable or the error in estimation blows up. The key observation is that such behavior is determined by the relationship between the Shannon capacity and the Lyapunov exponents; the well-known characterizing parameters of a communication channel on one side, and a dynamic system from the other side. In particular, it is proved that for nonlinear systems with initial state $x_0$, the minimum capacity of an AWGN channel required for full observation of the system in the mean square sense is the sum of $\big(\kappa_i(x_0)\Delta_i(x_0)\big)$, where $\Delta_i(x_0)$s and $\kappa_i(x_0)$s denote distinct Lyapunov exponents and their multiplicity numbers, respectively. Conversely, if the capacity is less than $E[\sum_i \kappa_i(x_0)\Delta_i(x_0)]$, then observation is impossible. To show the universality of the result, we obtain the same observability conditions for the digital noiseless channel and the packet erasure channel in sure and almost sure senses, respectively.**

***Index Terms*—Nonlinear dynamic systems, Observability, Lyapunov exponents, Capacity.**

## I. INTRODUCTION

### *A. Motivation and Backgrounds*

One of the issues that has begun to emerge in a number of applications, such as networked control systems, 5G mobile communication and tactile Internet [1], [2], is how to observe the state trajectory of a dynamic system over a communication channel subject to imperfections, such as noise, dropout, or limited capacity. In these applications, observability means how to transmit information about the state trajectory of a dynamic system and reconstruct it reliably in real-time at the receiver. In these applications, it is essential to find tight (necessary and sufficient) conditions such that full observability is achievable at the receiver side.

The problem of almost sure observability of linear time-invariant dynamic systems over the limited capacity packet erasure channel that uses feedback acknowledgment is addressed in [3]. The packet erasure channel is an abstract model for the commonly used systems such as the Internet, WiFi and mobile communications. In [3], it is shown that the eigenvalues-rate condition is tight. That is, the condition $C \geq \sum_{\{i:|\lambda_i(A)\geq 1|\}} \log_2 |\lambda_i(A)|$, where $C$ denotes the channel capacity (in bits per time step) and $\lambda_i(A)$s denote the eigenvalues of the system matrix $A$, is the necessary and sufficient condition for observability with almost sure asymptotically zero estimation error. For noisy linear time-invariant dynamic systems over the digital noiseless channel, it is shown in [4] that the eigenvalues-rate condition is tight for almost sure observability. As the output of the dynamic systems is continuous alphabet and the input-output of the Additive White Gaussian Noise (AWGN) channel is also continuous alphabet, the AWGN channel is suitable for control applications over communication channels. Hence, many works in the literature (e.g., [5], [6], [7], [8]) are concerned with the observability and stability of linear noisy time-invariant systems over the capacity limited AWGN channel. In [5], it is shown that the eigenvalues-rate condition is tight for mean square observability of linear noisy time-invariant systems over the capacity limited AWGN channel.

The problem of stability of noiseless nonlinear dynamic systems over the digital noiseless channel is addressed in [9], where it is shown that the tight bound on channel capacity for stability is given by the so-called topological entropy. The extension of topological entropy is presented in [10] for stability of uncertain dynamic systems with limited information. The results in [11] shows the relation between minimum bit rate for stabilization of Lipschitz nonlinear systems and stabilization entropy. The stability of locally Lipschitz nonlinear systems is also addressed in [12]. In [13], the authors presented a sufficient condition for the observability of the Lipschitz uncontrolled noiseless nonlinear systems over the digital noiseless channel with asymptotically zero mean square estimation error. The authors in [14] also presented a sufficient condition for observability of distributed uncontrolled Lipschitz systems subject to bounded process and measurement noises over the packet erasure network with bounded mean absolute estimation error. In [15], it is shown that the desired estimation of a nonlinear systems with limited information is impossible for bit rates which are lower than estimation entropy. Furthermore, it is proved that the derived upper bound on the estimation entropy matches the average bit rate guaranteeing the desired estimation task. In [16], the authors presented a necessary condition for mean square exponential observability of noiseless nonlinear dynamic systems over the real erasure channel in terms of erasure probability and positive Lyapunov exponents. In [17], the authors also presented a necessary condition in terms of the positive Lyapunov exponents for the stability of nonlinear noiseless dynamic systems over the real erasure channel. In [16], [17], the authors only studied those systems that have unique ergodic invariant measure; therefore, they assumed that the Lyapunov exponents are independent of $x_0$.

### *B. Paper Contributions*

In this paper, we present the fundamental role of the Lyapunov exponents for observability (real-time reliable data reconstruction of the state trajectory) of noiseless nonlinear dynamic systems under limited information. Key contributions of this work compared to aforementioned earlier literature can be summarized as follows:

1) In this work, we examine the observability problem of nonlinear system over every memoryless communication channels in different senses (sure, almost sure, and mean square) and prove that $C \geq \sum_{i=1}^{s} E[\kappa_i(x_0)\Delta_i(x_0)]$ is a necessary condition over all these classes of channels (Theorem 1). However, in earlier works, the observability necessary conditions are obtained only for the digital noiseless channel [15] or packet erasure channel [16] in square exponential sense. Moreover, [15], [16] presented the necessary conditions, respectively, based on estimation entropy and constant Lyapunov exponents. Furthermore, [3], [18], focused only on linear systems.
2) The authors in earlier works presented sufficient conditions for observability of Lipschitz nonlinear system [13], [15] over digital noiseless channel, however, in Theorem 2, $\sum_{i=1}^{s} \kappa_i(x_0)\Delta_i(x_0)$ is obtained as the lower bound on the channel capacity in order to guarantee the observability of nonlinear systems (1) (not only for Lipschitz systems) over the digital noiseless channel. Moreover, this lower bound improves the previous ones. Note that [18] presented the sufficient condition only for linear dynamic systems.
3) Theorem 3 presents observability sufficient condition for nonlinear system over packet erasure channel in almost sure sense as $C > \sum_{i=1}^{s} \kappa_i(x_0)\Delta_i(x_0)$; while in [16], there is no proof of sufficiency

V. Sanjaroon, A. Farhadi and B. H. Khalaj are with the Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran.

A. S. Motahari is with the Department of Computer Engineering, Sharif University of Technology, Tehran, Iran.

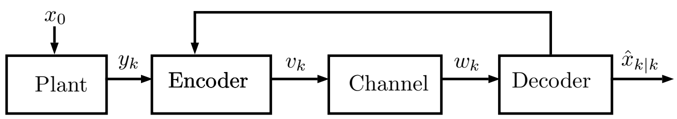


Fig. 1: Control/communication system over the Discrete Memoryless Channel (DMC) and AWGN channel

and [14] addressed this issue only for Lipschitz nonlinear systems. Also, [3] presented the sufficient condition only for linear systems.

4) To the best of our knowledge, for the first time, the observability issue of a nonlinear system over AWGN channel is addressed in Theorem 4 which presents the observability sufficient condition of such systems in the mean square sense. Note that [6], [7] presented the observability condition only for linear systems over AWGN channel.

Finally, this paper suggests that for the linear systems and also for nonlinear systems which have unique ergodic measure, the condition $C \geq \sum_{i=1}^{s} \kappa_i(x_0)\Delta_i(x_0)$, is tight for sure, almost sure, and mean square observability. Note that the conditions written in terms of the positive Lyapunov exponents result in the eigenvalues-rate condition for the special case of linear time-invariant systems as shown in [16] and [17]. Note also that the results of this paper extend the results of [3], [6], [7], [18] to the nonlinear dynamic systems.

### C. Paper Organization

The rest of the paper is organized as follows: in Section II, the problem formulation is given. In Section III, necessary conditions for sure, almost sure and mean square observability are presented; and in Section IV, the sufficient conditions and simulation results are presented for scalar nonlinear systems and then the results are extended to vector case. Finally, Section V concludes the paper.

## II. Notations and Preliminaries

### A. Notation

We use the following notation throughout this paper. $||\cdot||$ and $|\cdot|$ denote, respectively, the Frobenius norm and the absolute value. det$(\cdot)$ and $(\cdot)^{tr}$ denote, respectively, the determinant and the transpose of a matrix, $\log(\cdot)$ is the binary logarithm, $E[\cdot]$ denotes the expected value function, $E_{x|y}[\cdot]$ denotes the expected value of the random variable $x$ with respect to the random variable $y$, $\mathbb{P}(\cdot)$ is the probability mass function. The $k$th orbit of $f(\cdot)$ is denoted by $f^{(k)}(x)$ and its Jacobian matrix is denoted by $Df^{(k)}(x)$. $v_{t_1}^{t_2}$ denotes the sequence $v_{t_1}, v_{t_1+1}\ldots, v_{t_2}$ and $v^{t_2}$ is used in place of $v_0^{t_2}$. when an inequality symbol involves matrices or vectors, we mean that the inequality is satisfied elementwise.$\mathbb{R}$ and $\mathbb{Z}^+$ are, respectively, the real number and positive integer numbers. $\overset{a.s}{=}$ represents almost sure convergence.

### B. Preliminaries

The overall system model considered in this paper is shown in Fig. 1. In what follows, we describe the input-output relation of each building block of this system.

**Plant:** The plant is described by the following nonlinear, time invariant and fully-observed system:

$$\begin{aligned} x_{k+1} &= f(x_k), \\ y_k &= x_k, \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^d$ is the state and $y_k$ is the observed signal. The initial state $x_0 : \Omega \to \psi_0$ is a random vector, i.e., $x_0$ is a measurable function from possible outcomes $\Omega$ to the set $\psi_0 \subset \mathbb{R}^d$.

In this paper, we use the Lyapunov exponents of function $f(.)$ to present necessary and sufficient conditions for observability of the system over a communication channel. The Lyapunov exponent measures the average expansion rate of nonlinear systems which is defined as follows [19].

*Definition 1:* Consider the following symmetric-defined matrices:

$$\begin{aligned} L_k(x) &= \left(Df^{(k)}(x)^{tr} Df^{(k)}(x)\right)^{\frac{1}{2k}}, \\ L(x) &= \lim_{k\to\infty} L_k(x). \end{aligned} \tag{2}$$

The logarithm of the eigenvalues of $L(x)$ are $s$ distinct Lyapunov exponents ($s \leq d$) of the nonlinear system denoted by $\Delta_i(x)$, $1 \leq i \leq s$. The corresponding eigenvectors $p_1(x), p_2(x), \ldots, p_s(x)$ are orthogonal (due to the symmetric matrix) and the corresponding eigenspaces are denoted by $P_1(x), P_2(x), \ldots, P_s(x)$. The number $\Delta_i(x)$ is counted with the multiplicity $\kappa_i(x) = dim(P_i(x))$. For every nonzero vector $q \in P_i(x)$, $i = 1, 2, \ldots, s$, we get:

$$\Delta_i(x) = \lim_{k\to\infty} \frac{1}{k} \log\left(||Df^{(k)}(x)q||\right). \tag{3}$$

We make the following assumptions on the system dynamic.

*Assumption 1:* For an uncontrolled dynamic system $f(.)$, $L_k(x)$ is a matrix whose elements are the sequences of continuous functions which converge locally uniformly to the limit functions on domain $\psi_0$. Furthermore, it is assumed that $\log(\det\ L_k(x))$ is a sequence of bounded functions on this domain .

*Assumption 2:* We assume $f(.)$ is differentiable. Note that the differentiability of $f(\cdot)$ results in the differentiability of $f^{(k)}(.)$.

*Assumption 3:* We assume that $\Delta_i(x)$ is bounded, continuous, and positive on domain $\psi_0$ for every $i = 1, 2, \ldots, s$.

**Communication channel:** A communication channel is modeled by a probability transition $\mathbb{P}(w|v)$, where $v$ and $w$ are the channel input and output, respectively.

In this paper, we consider several types of communication channels as follows.

- The digital noiseless channel with average rate $R_{av}$: For the $k$th transmission, the input is chosen from a set with $2^{R_k}$ members, and the channel output is the same as the channel input. In fact, $R_k$ noiseless bits are transmitted at each channel use where $R_{av} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} R_k$ . Therefore, the channel is noiseless and memoryless and its capacity equals to $C = R_{av}$ bits in each channel use.
- The packet erasure channel with rate $R_{av}$ and erasure probability $\gamma$: In this case, the space of channel input is a set with $2^{R_k}$ members, and the channel output is the same as the input symbol (with probability $1-\gamma$) or the erasure symbol (with probability $\gamma$). The capacity of this channel is $C = (1-\gamma)R_{av} = (1-\gamma)\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} R_k$.
- AWGN channel: This channel is described as follows,

$$w_k = v_k + n_k, \tag{4}$$

where $v_k \in \mathbb{R}$ denotes the channel input, $w_k \in \mathbb{R}$ the channel output at time instant $k$, and $n_k$ the additive white Gaussian noise with zero mean and variance $\sigma^2$ assumed to be independent of the initial state. The capacity of this channel under the average power constraint $p_{av} \leq p$ equals to $C = \frac{1}{2}\log(1 + \frac{p}{\sigma^2})$.
Note that, for a given input signal, the average power can be computed as $p_{av} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} E[v_k^2]$ which is assumed to be less than $p$ due to the average power constraint.

**Encoder:** In general, in the availability of feedback, an encoder maps $(y^k, w^{k-1}) \to v_k$. However, as there is no process noise in the system and all the ambiguity is due to the initial state, for any encoder, there is an equivalent encoder which maps $(x_0, w^{k-1}) \to v_k$.

**Decoder:** The decoder is an operator that maps $w^k \to \hat{x}_{k|k}$, where $\hat{x}_{k|k}$ is the reconstruction of the state variable $x_k$ at time instant $k$ given the observation $w_0, w_1, \ldots, w_k$. We assume $\hat{x}_{0|-1} = 0$.

In this paper, we are interested in the observability problem at the output of communication link as defined below.

*Definition 2:* The system (1) over communication link is observable in sure, almost sure, in probability, and mean square senses if an encoder and a decoder exist such that, respectively,

– $\lim_{k\to\infty} ||x_k - \hat{x}_{k|k}|| = 0$,
– $\mathbb{P}[\lim_{k\to\infty} ||x_k - \hat{x}_{k|k}|| = 0] = 1$,
– for all $\epsilon > 0$, $\lim_{k\to\infty} \mathbb{P}\big[||x_k - \hat{x}_{k|k}|| \geq \epsilon\big] = 0$,
– $\lim_{k\to\infty} E[(x_k - \hat{x}_{k|k})^2] = 0$.

## III. Necessary conditions

In this section, we present the necessary condition for the observability of the system (1) over any memoryless communication channel in probability sense which also implies the necessary condition of observability in sure, almost sure and mean square senses. In the following theorem, we prove the necessary condition in the probability sense with the assumption that $f(\cdot)$ is one to one:

*Theorem 1:* A necessary condition for observability of the system (1) in the probability sense over memoryless channels (in particular, DMC and AWGN channel) with capacity $C$ is that $C \geq E[\sum_{i=1}^{s} \kappa_i(x_0)\Delta_i(x_0)]$ where $x_0$ is a random vector with bounded entropy.

Proof: Due to the characteristics of memoryless channels, the following relations are derived:

$$\begin{aligned} I(x_k;\hat{x}_{k|k}) &\leq I(x^k;\hat{x}_{k|k}) \overset{(a)}{\leq} I(x^k;w^k) = h(w^k) - h(w^k|x^k), \\ &= h(w^k) - \sum_{i=0}^{k} h(w_i|w_0,\ldots,w_{i-1},x^k), \\ &\overset{(b)}{=} h(w^k) - \sum_{i=0}^{k} h(w_i|w_0,\ldots,w_{i-1},v_i,x^k), \\ &\overset{(c)}{=} h(w^k) - \sum_{i=0}^{k} h(w_i|v_i) \leq \sum_{i=0}^{k} I(w_i;v_i), \\ &\overset{(d)}{\leq} (k+1)C, \end{aligned} \tag{5}$$

where $h(.)$ is the differential entropy for continuous variable. Using the data processing inequality [20] leads to inequality $(a)$. $(b)$ follows because with the knowledge of $w_0,\ldots,w_{i-1},x_i$, one can obtain $v_i$ exactly since the encoder maps $(x_i, w^{i-1})$ to $v_i$. $(c)$ is a direct result of memoryless channel [20]. $(d)$ is also a derived result for memoryless channel in [20].

For observability in the probability sense, the following function provides a measure of distortion [3]:

$$d^D(x,\hat{x}) = \begin{cases} 0 & ||x-\hat{x}|| \leq D, \\ 1 & ||x-\hat{x}|| > D. \end{cases} \tag{6}$$

Hence, $E[d^D(x,\hat{x})] = \mathbb{P}(d^D(x,\hat{x})=1) = \mathbb{P}(||x-\hat{x}|| > D)$. As the system is observable in the probability sense, for every $\epsilon > 0$, there is $k_1(\epsilon) \in \mathbb{Z}^+$ such that for every $k \geq k_1(\epsilon)$, $E[d^\epsilon(x_k;\hat{x}_{k|k})] \leq \epsilon$.

In order to derive the relation between channel capacity and distortion, an upper bound on $I(x_k;\hat{x}_{k|k})$ is deduced for $k \geq k_1(\epsilon)$ under the condition $\mathbb{P}(||x_k - \hat{x}_{k|k}|| > \epsilon) \leq \epsilon$ as follows:

$$\begin{aligned} (k+1)C \geq& I(x_k;\hat{x}_{k|k}), \\ &\overset{(a)}{\geq} (1-\epsilon)h(x_k) - \frac{1}{2} - \log(K_d\epsilon^d), \\ &\overset{(b)}{=} (1-\epsilon)\big(h(x_0) + E[\log|\det Df^{(k)}(x_0)|]\big) - \frac{1}{2} - \log(K_d\epsilon^d), \end{aligned} \tag{7}$$

where $K_d$ is the volume constant for the $d$ dimensional sphere [20]. $(a)$ is the upper bound for $I(x_k;\hat{x}_{k|k})$ which is obtained in [3] under the condition $\mathbb{P}(||x_k - \hat{x}_{k|k}|| > \epsilon) \leq \epsilon$. $(b)$ is resulted as the function $x_k = f^{(k)}(x_0)$ is assumed to be one to one [21].

Therefore, based on (7), the following inequality is true for every $\epsilon > 0$ and $k \geq k_1(\epsilon)$:

$$\begin{aligned} C \geq& \frac{1-\epsilon}{k+1}E[\log|\det Df^{(k)}(x_0)|] - \frac{1}{k+1}\log(K_d\epsilon^d) \\ &+ \frac{1}{k+1}\big((1-\epsilon)h(x_0) - \frac{1}{2}\big). \end{aligned} \tag{8}$$

We take the limit with $k$ going to infinity:

$$\begin{aligned} C \geq& \lim_{k\to\infty} \frac{1-\epsilon}{k+1}E[\log|\det Df^{(k)}(x_0)|] - \frac{1}{k+1}\log(K_d\epsilon^d) \\ &+ \frac{1}{k+1}\big(h(x_0)(1-\epsilon) - \frac{1}{2}\big), \\ &= (1-\epsilon)\lim_{k\to\infty} E[\frac{1}{k+1}\log|\det Df^{(k)}(x_0)|], \\ &\overset{(a)}{=} (1-\epsilon)E[\lim_{k\to\infty}\frac{1}{k+1}\log|\det Df^{(k)}(x_0)|], \\ &\overset{(b)}{=} (1-\epsilon)E\big[\sum_{i=1}^{s}\kappa_i(x_0)\Delta_i(x_0)\big]. \end{aligned} \tag{9}$$

Note that in the above equalities/inequalities, $(a)$ follows since for bounded sequence $\frac{1}{k+1}\log\Big|\det Df^{(k)}(x_0)\Big|$, we can interchange limits and expected value functions [22]. Equation $(b)$ is resulted since $\lim_{k\to\infty}\frac{1}{k+1}\log|\det Df^{(k)}(x_0)| = \sum_{i=1}^{s}\kappa_i(x_0)\Delta_i(x_0)$ [19]. Note that the inequalty (9) is true for arbitrary small value of $\epsilon > 0$ which implies that $C > E\big[\sum_{i=1}^{s}\kappa_i(x_0)\Delta_i(x_0)\big]$ is the necessary condition. ■

## IV. Sufficient conditions

In this section, we address the sufficient conditions for observability of the control/communication system of Fig. 1 consisting of the digital noiseless, packet erasure, and AWGN channels. For this purpose, we propose encoders and decoders that map $(x_0, w^{k-1}) \to v_k$ and $w^k \to \hat{x}_{k|k}$, respectively. This means that as the ambiguity in the states of the system is only due to the ambiguity in the initial state, it is possible to guarantee reliable observability by focusing on reconstruction of the initial state at the receiver and sending it back to the transmitter throughout communication steps. We explain the proposed communication scheme for scalar case and then extend it to vector case.

### A. Scalar case

In order to observe a scalar nonlinear system, the innovation process $(x_0 - \hat{x}_{0|k-1})$ is sent at every time slot $k$. Upon receiving $w_k$, $\hat{x}_{0|k}$ is updated and sent back to the transmitter. Moreover, $\hat{x}_{k|k} = f^{(k)}(\hat{x}_{0|k})$ is also obtained to track the system. In this section, we prove that using such coding scheme, we can obtain $C > \Delta(x_0)$ as the sufficient condition for observability.

*Theorem 2:* The sufficient condition for sure observability of the control/communication system consisting of the system (1) and the digital noiseless channel is $C = R_{av} > \Delta(x_0)$, where $x_0$ takes any value in $\psi_0 = [-L_0, L_0]$.

Proof: At every time instant $k$, we quantize the innovation process to $R_k$ bits with a zero centroid quantizer whose dynamical range is $\Big(2L_0/2^{\sum_{i=0}^{k-1}R_i}\Big)$ [18]. We can adaptively change the value of $R_k$ based on $\hat{x}_{0|k-1}$, since $\hat{x}_{0|k-1}$ is available at both transmitter and receiver. In fact, if this adaptive change of rates is possible, the proposed coding scheme requires different average bit rates for different initial states. Therefore, for designing a casual encoding/decoding scheme, we can set the value of bit rates such that $R_k = \mathcal{H}(\hat{x}_{0|k-1})$ where $\mathcal{H}(x)$ is positive, bounded, and continuous on domain $\psi_0$.[1] Due to the positive bit rates, the dynamical rang of quantizer goes to zero which implies the convergence of $\hat{x}_{0|k}$ to $x_0$. Note that due to the convergence of $\hat{x}_{0|k-1}$

[1]Note that for fractional bit rates, time sharing technique should be used as explained in Appendix A.

to $x_0$ and continuity of $\mathcal{H}(.)$, $R_k$ converges to $\mathcal{H}(x_0)$ and as a result, $R_{av}$ equals to $\mathcal{H}(x_0)$. By reconstruction of $\hat{x}_{k|k} = f^{(k)}(\hat{x}_{0|k})$, we have:

$$|x_k - \hat{x}_{k|k}| \leq \max_{-L_0 \leq a \leq L_0} |f^k(\hat{x}_{0|k} + \frac{a}{2^{\sum_{i=0}^{k-1} R_i}}) - f^k(\hat{x}_{0|k})|. \tag{10}$$

Now, the mean value theorem guaranties the existence of $c \in [0,1]$ for every $a \in [-L_0, L_0]$, such that:

$$|x_k - \hat{x}_{k|k}| \leq \max_{-L_0 \leq a \leq L_0} \varepsilon \big|(Df^k(\hat{x}_{0|k} + \varepsilon c))\big|, \tag{11}$$

where $\varepsilon \triangleq a/2^{\sum_{i=0}^{k-1} R_i}$. If the right hand of the above inequality goes to zero everywhere, then the system is observable. Towards this purpose, it is sufficient that:

$$\lim_{k\to\infty} \left( \max_{-L_0 \leq a \leq L_0} \varepsilon \big|(Df^k(\hat{x}_{0|k} + \varepsilon c))\big| \right)^{\frac{1}{k}}$$
$$= \max_{-L_0 \leq a \leq L_0} \lim_{k\to\infty} (\varepsilon)^{\frac{1}{k}} \big|(Df^k(\hat{x}_{0|k} + \varepsilon c))\big|^{\frac{1}{k}} < 1. \tag{12}$$

Note that if $R_{av} = \Delta(x_0) + \epsilon$ for an arbitrary positive value of $\epsilon$, the above condition is satisfied, since:

$$\max_{-L_0 \leq a \leq L_0} \lim_{k\to\infty} (\varepsilon)^{\frac{1}{k}} \big|(Df^k(\hat{x}_{0|k} + \varepsilon c))\big|^{\frac{1}{k}} = \frac{L_0}{2^{R_{av}}} 2^{\Delta(x_0)}$$
$$= \frac{L_0}{2^{\Delta(x_0)+\epsilon}} 2^{\Delta(x_0)} < 1. \tag{13}$$

Therefore, with choosing function $\mathcal{H}(.)$ as $\Delta(.) + \epsilon$, (i.e., $R_k = \Delta(\hat{x}_{0|k-1}) + \epsilon$), the system is observable. Obviously, the minimum required channel capacity for this transmission is $\Delta(x_0) + \epsilon$.

Hence, under the condition $C > \Delta(x_0)$, there is an encoder/decoder which makes the system sure observable. This proves that $C > \Delta(x_0)$ is a sufficient condition for sure observability of the nonlinear system over the digital noiseless channel. ■

In the next theorem, the sufficient condition for the observability of the dynamic system over the packet erasure channel is derived.

*Theorem 3:* The sufficient condition for almost sure observability of the control/communication system consisting of the system (1) and the packet erasure channel with erasure probability $\gamma$, average bit rate $R_{av}$ and feedback acknowledgment is $C = R_{av}(1-\gamma) > \Delta(x_0)$, where $R_{av}(1-\gamma)$ is the channel capacity and $x_0$ takes the values in the set $\psi_0 = [-L_0, L_0]$.

Proof: At each time instant $k$, we apply a zero centroid quantizer with dynamical range $\left[2L_0/\prod_{i=0}^{k-1} 2^{\beta_i R_i}\right]$ in order to quantize the innovation process into $R_k$ bits, where $\beta_i$s are independent indicator random variables with the following distribution [3]:

$$p(\beta_i = 0) = \gamma; \qquad p(\beta_i = 1) = 1 - \gamma. \tag{14}$$

In fact, if the acknowledgment is not received at the transmitter, transmitter resends the previous signal and therefore, the ambiguity of $x_0$ will not be changed.

As the transmitter and receiver have access to the value of $\hat{x}_{0|k-1}$ at time instant $k$, we can adaptively set the value of $R_k$ as a bounded, positive, and continuous function of $\hat{x}_{0|k-1}$, (i.e., $R_k = \mathcal{H}(\hat{x}_{0|k-1})$). Using this encoder-decoder pair, $\hat{x}_{0|k}$ and $\mathcal{H}(\hat{x}_{0|k})$, respectively, converges to $x_0$ and $\mathcal{H}(x_0)$ almost surely and as the results, $R_{av} \overset{a.s}{=} \mathcal{H}(x_0)$.

Note that reconstruction $\hat{x}_{k|k} = f^{(k)}(\hat{x}_{0|k})$ leads to:

$$|x_k - \hat{x}_{k|k}| \leq \max_{a\in[-L_0, L_0]} |f^k(\hat{x}_{0|k} + \frac{a}{2^{\sum_{i=0}^{k-1} \beta_i R_i}}) - f^k(\hat{x}_{0|k})|. \tag{15}$$

For almost sure observability of the system, it is sufficient that the right hand side of the above inequality converges to zero almost surely. Following similar steps as taken in the proof of Theorem 2, one can obtain the sufficient condition for almost sure observability as follows:

$$\Delta(x_0) \overset{a.s}{<} \lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \beta_i R_i = \lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \beta_i \mathcal{H}(\hat{x}_{0|i-1}). \tag{16}$$

In Appendix B, we prove that $\frac{1}{k}\sum_{i=0}^{k-1} \beta_i \mathcal{H}(\hat{x}_{0|i-1})$ converges almost surely to $(1-\gamma)\mathcal{H}(x_0)$ for bounded, positive, and continuous function $\mathcal{H}(\cdot)$. Therefore, if we choose $\mathcal{H}(\cdot) = \frac{1}{1-\gamma}\Delta(\cdot) + \epsilon$ (i.e., $R_k = \frac{1}{1-\gamma}\Delta(\hat{x}_{0|k-1}) + \epsilon$), for an arbitrary value of $\epsilon > 0$, the system is observable in almost sure sense, since:

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \beta_i \mathcal{H}(\hat{x}_{0|i-1}) \overset{a.s}{=} (1-\gamma)\mathcal{H}(x_0),$$
$$= \Delta(x_0) + (1-\gamma)\epsilon > \Delta(x_0), \tag{17}$$

Now, with dependency assumption of channel capacity on the initial state, we obtain the minimum required channel capacity for this procedure as follows:

$$C = (1-\gamma)E[R_{av}|x_0],$$
$$= \lim_{k\to\infty} \frac{(1-\gamma)}{k}(R_0 + \frac{1}{1-\gamma}\sum_{i=1}^{k-1} E[\Delta(\hat{x}_{0|i-1})|x_0]) + \epsilon,$$
$$= \Delta(x_0) + (1-\gamma)\epsilon. \tag{18}$$

Hence, under the condition $C > \Delta(x_0)$, there is an encoder/decoder which makes the system almost surely observable. ■

The final theorem in this section, is about the sufficient condition for observability of the control system over AWGN channel in the mean square sense. To achieve this condition, the following additional assumptions will be also placed on the plant (1):

*Assumption 4:* We assume that the function $f(.)$ is analytic. Note that the analyticity of $f(\cdot)$ results in the analyticity of $f^{(k)}(.)$.

*Assumption 5:* We assume that on domain $\psi_0$, $\Delta(x)$ is bounded, positive, and differentiable function whose derivative is also bounded.

*Assumption 6:* We assume that $f(.)$ is a function such that for every $x \in \psi_0$ and $n \geq 2$, we have,

$$\lim_{k\to\infty} \left(D^n f^{(k)}(x)\right)\left(Df^{(k)}(x)\right)^{-n} \leq \bar{f},$$

where $\bar{f}$ is constant and $D^n f^{(k)}(\cdot)$ is the $n$th derivative for function $f^{(k)}(\cdot)$. Note that conventional functions have this property.

*Theorem 4:* The sufficient condition for mean square observability of the control/communication system consisting of the system (1) and AWGN channel (4) with noiseless feedback link and capacity $C$ is $C > \Delta(x_0)$, where the initial state has Gaussian distribution.

Proof: We use a coding scheme that is based on [23] to address this question. At every time slot $k$, $v_k = T(k)(x_0 - \hat{x}_{0|k-1})$ is transmitted. Upon receiving $w_k$, $\hat{x}_{0|k} = \hat{x}_{0|k-1} + R(k)w_k$ is computed and sent back to the transmitter and then, $\hat{x}_{k|k}$ is also obtained in order to track the system.

The functions $T(.)$ and $R(.)$ should be chosen as follows:

$$T(k) = \begin{cases} \alpha_0 & k = 0, \\ (\prod_{i=0}^{k-1}\alpha_i)(\alpha_k^2 - 1)^{\frac{1}{2}} & k \geq 1, \end{cases}$$
$$R(k) = \begin{cases} \alpha_0^{-1} & k = 0, \\ (\prod_{i=0}^{k-1}\alpha_i)^{-1}\frac{(\alpha_k^2-1)^{\frac{1}{2}}}{\alpha_k^2} & k \geq 1. \end{cases}$$

Consequently, the estimation of the initial state after the first step is $\hat{x}_{0|0} = x_0 + \frac{n_0}{\alpha_0}$, and after the second step, it is

$$\hat{x}_{0|1} = \hat{x}_{0|0} + \frac{1}{\alpha_0}\frac{(\alpha_1^2-1)^{1/2}}{\alpha_1^2}w_1 = x_0 + \frac{n_0}{\alpha_0\alpha_1^2} + \frac{1}{\alpha_0}\frac{(\alpha_1^2-1)^{1/2}}{\alpha_1^2}n_1,$$

The recursive solution is then given by:

$$\hat{x}_{0|k} = x_0 + \rho_{0|k}n_0 + \rho_{1|k}n_1 + \ldots + \rho_{k|k}n_k, \tag{19}$$

where, $\rho_{0|k} = \frac{1}{\alpha_0\prod_{i=1}^{k}\alpha_i^2}$ and $\rho_{i|k}$ for $i \geq 1$ is obtained through the following equation:

$$\rho_{i|k} = \frac{(\alpha_i^2-1)^{\frac{1}{2}}}{\prod_{j=0}^{i-1}\alpha_j\prod_{j=i}^{k}\alpha_j^2}. \tag{20}$$

Due to the knowledge of transmitter and receiver about the value of $\hat{x}_{0|k-1}$ at time instant $k$, they can set $\log(\alpha_k), k \neq 0$ as a bounded, positive, and continuous function of $\hat{x}_{0|k-1}$. This choice leads to mean squarely convergence of $\hat{x}_{0|k-1}$ to $x_0$. With definition ($e_{0|k} \triangleq x_0 - \hat{x}_{0|k}$) and reconstruction $\hat{x}_{k|k} = f^{(k)}(\hat{x}_{0|k})$, and using the Taylor expansion for analytic function $f^{(k)}(x)$, we have:

$$\hat{x}_{k|k} = x_k + Df^{(k)}(x_0)e_{0|k} + \sum_{n=2}^{\infty} \frac{1}{n!} D^n f^{(k)}(x_0) e_{0|k}^n. \tag{21}$$

Consequently, we obtain error variance from (21) as follows:

$$\begin{aligned} E[(x_k - \hat{x}_{k|k})^2|x_0] = \\ & E\big[\big(Df^{(k)}(x_0)e_{0|k} + \sum_{n=2}^{\infty} \frac{D^n f^{(k)}(x_0)}{n!} e_{0|k}^n\big)^2|x_0\big], \\ & = Df^{(k)}(x_0)^2 E[e_{0|k}^2|x_0] \\ & + \sum_{\substack{t=1 \\ (t,n)\neq(1,1)}}^{\infty} \sum_{n=1}^{\infty} \frac{1}{t!n!} D^t f^{(k)}(x_0) D^n f^{(k)}(x_0) E[e_{0|k}^{n+t}|x_0]. \end{aligned} \tag{22}$$

Note that based on Assumption 6, we have:

$$\begin{aligned} \lim_{k\to\infty} E[(x_k - \hat{x}_{k|k}))^2|x_0] \leq \lim_{k\to\infty} \Big\{ Df^{(k)}(x_0)^2 E[e_{0|k}^2|x_0] \\ + \bar{f}^2 \sum_{\substack{t=1 \\ (t,n)\neq(1,1)}}^{\infty} \sum_{n=1}^{\infty} \frac{1}{t!n!} Df^{(k)}(x_0)^{(t+n)} E[e_{0|k}^{n+t}|x_0] \Big\}. \end{aligned} \tag{23}$$

Therefore, if we set $\alpha_k = 2^{\Delta(\hat{x}_{0|k-1})+\epsilon}$ for an arbitrary positive value of $\epsilon$, $\alpha_k$ converges to $\alpha(x_0) \triangleq 2^{\Delta(x_0)+\epsilon}$ in probability and according to Appendix C, the error variance converges to zero and system is observable mean squarely.

Now, we should calculate the minimum required power for such procedure:

$$\begin{aligned} P_{av} &= \lim_{k\to\infty} P_{av}(k) = \lim_{k\to\infty} \frac{1}{k} E[\alpha_0^2 (x_0 - \hat{x}_{0|-1})^2 \\ &+ \frac{1}{k} \sum_{i=1}^{k-1} (\prod_{l=0}^{i-1} \alpha_l (\alpha_i^2 - 1)^{\frac{1}{2}})^2 (x_0 - \hat{x}_{0|i-1})^2 |x_0], \\ = \lim_{k\to\infty} & \frac{1}{k} \alpha_0^2 x_0^2 + \frac{1}{k} \sum_{i=1}^{k-1} E[(\prod_{l=0}^{i-1} \alpha_l (\alpha_i^2 - 1)^{\frac{1}{2}})^2 (x_0 - \hat{x}_{0|i-1})^2 |x_0]. \end{aligned} \tag{24}$$

Using (41) in Appendix C leads to the following equation:

$$\begin{aligned} & \lim_{i\to\infty} E[(\prod_{l=0}^{i-1} \alpha_l (\alpha_i^2 - 1)^{\frac{1}{2}})^2 (x_0 - \hat{x}_{0|i-1})^2 |x_0] \\ & = \lim_{i\to\infty} E[(\alpha_i^2 - 1)\sigma^2 |x_0]. \end{aligned} \tag{25}$$

Note that as $\hat{x}_{0|k}$ converges to $x_0$ mean squarely, $\Delta(\hat{x}_{0|i-1})$ converges to $\Delta(x_0)$ in probability. Additionally, $\Delta(x_0)$ is bounded and continuous, Therefore:

$$\lim_{i\to\infty} E[(\prod_{l=0}^{i-1} \alpha_l (\alpha_i^2 - 1)^{\frac{1}{2}})^2 (x_0 - \hat{x}_{0|i-1})^2 |x_0] \quad = (2^{2\Delta(x_0)+2\epsilon} - 1)\sigma^2.$$

Now, as $E[(\prod_{l=0}^{i-1} \alpha_l (\alpha_i^2 - 1)^{\frac{1}{2}})^2 (x_0 - \hat{x}_{0|i-1})^2 |x_0]$ is a convergent sequence, its arithmetic mean is also convergent with the same limit point. Therefore,

$$P_{av} = \lim_{k\to\infty} P_{av}(k) = \lim_{k\to\infty} \frac{\alpha^2(x_0^2)}{k} + (2^{2\Delta(x_0)+2\epsilon} - 1)\sigma^2, \tag{26}$$

for any $x_0$, it is easy to derive the minimum required channel capacity for this transmission, as follows:

$$C = \frac{1}{2} \log(1 + \frac{P}{\sigma^2}) = \Delta(x_0) + \epsilon. \tag{27}$$

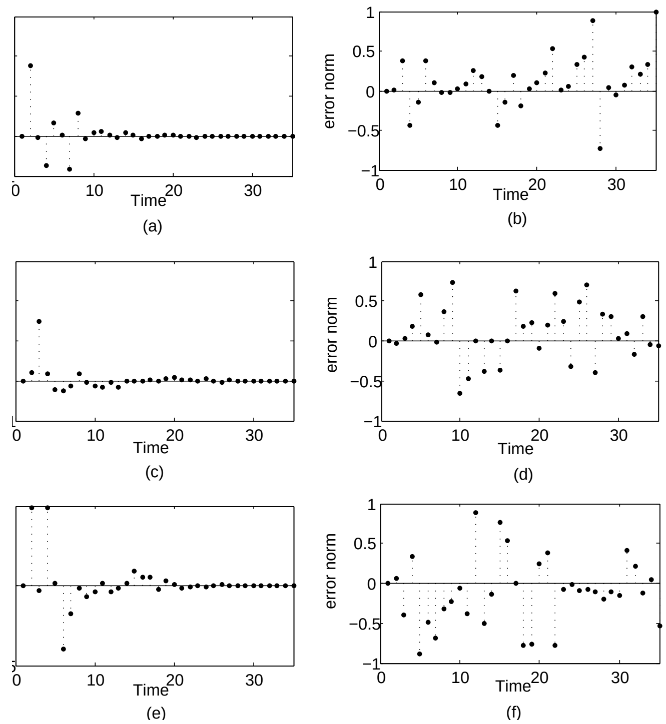


Fig. 2: (a), (c), (e): Estimation error of the dynamic system over, respectively, the digital noiseless, packet erasure and AWGN channels with capacity $C = 1.3$ bit per channel use. (b), (d), (f): Estimation error of the dynamic system over, respectively, the digital noiseless, packet erasure and AWGN channels with $C = 0.9$ bit per channel use.

Consequently, if $C > \Delta(x_0)$ there is a coding scheme in order to achieve observability of the dynamic system over AWGN channel in the mean square sense. ∎

### B. Simulation results

We consider the following scalar nonlinear system satisfying Assumption 1 and others:

$$x_{k+1} = 4x_k(1 - x_k),$$

where $x_0 \in [0.1, 0.4]$ is a random variable. It is assumed that $\hat{x}_{0|-1} = 0.2$, and therefore, $(x_0 - \hat{x}_{0|-1}) \in [-0.2, 0.2]$. The Lyapunov exponent of the above dynamic system on domain $[0.1, 0.4]$ is constant and equals to $log(2) = 1$ bit [24]. We simulate the estimate of the state $x_k$ over the digital noiseless, packet erasure, and AWGN channels using the proposed encoding/decoding schemes. The results of the estimation error are presented in Fig. 2 for two different values of channel capacity. The results illustrate that using the proposed communication schemes, the above nonlinear system over digital noiseless, packet erasure and AWGN channels is observable if $C > \Delta(x_0)$.

### C. Vector case

In this part, we prove that the sufficient condition for observability in Theorem 2, Theorem 3, and Theorem 4 is expressed as $C > \sum_{i=1}^{s} \kappa_i(x_0)\Delta(x_0)$ for multi-dimensionl system. Without loss of generality, we assume that the system has $d$ distinct and positive Lyapunov exponents. Due to the space limitation, we prove the sufficient condition for a two dimension system over the digital noiseless channel. For other channels and higher order, a similar methodology can be used.

We assume that $x_k = [a_k, b_k]^{tr}$ is the state of the system where $x_0 \in \psi_0$ is a bounded random vector and $\hat{x}_{0|-1} = [0, 0]^{tr}$ (i.e., there are two constant numbers $L_0^1$ and $L_0^2$ such that $[-L_0^1, -L_0^2]^{tr} \leq x_0 - \hat{x}_{0|-1} \leq [L_0^1, L_0^2]^{tr}$). Based on the Definition 1, for every $x_0 \in \psi_0$, there are two eigenvectors $p_1(x_0)$ and $p_2(x_0)$ corresponding to the Lyapunov exponents, hence, we can define $\mathcal{P}(x_0) = [p_1(x_0), p_2(x_0)]$ as a unitary matrix.

In our proposed scheme, at every time instant $k$, we transmit $\mathcal{P}^{tr}(\hat{x}_{0|k-1})(x_0 - \hat{x}_{0|k-1})$ over the digital noiseless channel by assigning $R_k^1$ and $R_k^2$ bits to its first and second elements, respectively. Subsequently, uponing the channel output $w_k$, the initial state estimate is updated at the reciev$\mathcal{P}^{tr}(\hat{x}_{0|k-1})$ is a unitary matrix, therefore, for $k \geq 1$, we can define $L_k^1$ and $L_k^2$ through the following ittiartive equation such that $[-L_k^1, -L_k^2]^{tr} \leq x_0 - \hat{x}_{0|k-1} \leq [L_k^1, L_k^2]^{tr}$.

$$\begin{bmatrix} L_k^1 \\ L_k^2 \end{bmatrix} = \mathcal{P}(\hat{x}_{0|k-1}) \begin{bmatrix} 1/2^R & 0 \\ 0 & 1/2^R \end{bmatrix} \mathcal{P}^{tr}(\hat{x}_{0|k-1}) \begin{bmatrix} L_k^1 \\ L_k^2 \end{bmatrix} \quad (28)$$

dyamivc (29) is stable, $[L_k^1, L_k^2]$ converges to zer(i.e., $x_0$ converges to $\hat{x}_{0|k}$)

based on the multi variable mean value theorem, there exist $c, c' \in [0 \;\; 1]$ such that:

$$a_k - \hat{a}_{k|k} = Df_1^{(k)}\big((1-c)x_0 + c\hat{x}_{0|k}\big(x_0 - \hat{x}_{0|k}\big),$$
$$b_k - \hat{b}_{k|k} = Df_2^{(k)}\big((1-c')x_0 + c'\hat{x}_{0|k}\big)(x_0 - \hat{x}_{0|k}),$$

where $Df_i^{(k)}$ is the $i$th row of the Jacobian matrix $Df^{(k)}$. The above terms converge to zero if:

$$\lim_{k\to\infty} norm \begin{bmatrix} Df_1^{(k)}\big((1-c)x_0 + c\hat{x}_{0|k}\big) \\ Df_2^{(k)}\big((1-c')x_0 + c'\hat{x}_{0|k}\big) \end{bmatrix} \begin{bmatrix} L_{k+1}^1 & \pm L_{k+2}^2 \end{bmatrix} = 0.$$

or equivalently subsituting (29),

$$\lim_{k\to\infty} norm \begin{bmatrix} Df_1^{(k)}\big((1-c)x_0 + c\hat{x}_{0|k}\big) \\ Df_2^{(k)}\big((1-c')x_0 + c'\hat{x}_{0|k}\big) \end{bmatrix} p(\hat{x}_{0|k})$$
$$\prod_{i=0}^{k} \left( \begin{bmatrix} 1/2^{\sum R_i^1} & 0 \\ 0 & 1/2^{\sum R_i^2} \end{bmatrix} p^{tr}(\hat{x}_{0|k}) p(\hat{x}_{0|k-1}) \right) = 0.$$

As $\hat{x}_{0|k}$ converges to $x_0$, $p(\hat{x}_{0|k})$ and $p^{tr}(\hat{x}_{0|k})p(\hat{x}_{0|k-1})$ respectively converges to $p(x_0)$ and unit matrix. therefore the suffiicent condition is reduced to :

$$\lim_{k\to\infty} norm \begin{bmatrix} Df_1^{(k)}\big((1-c)x_0 + c\hat{x}_{0|k}\big) \\ Df_2^{(k)}\big((1-c')x_0 + c'\hat{x}_{0|k}\big) \end{bmatrix} p(x) \begin{bmatrix} 1/2^{\sum R_i^1} & 0 \\ 0 & 1/2^{\sum R_i^2} \end{bmatrix} = 0.$$

Therefore, we can write an upper bound on the above norm as the summation of the following two terms:

$$\mathcal{A} = \frac{2}{2^{2\sum_{i=0}^{k} R_i^1}} \left( norm( \begin{bmatrix} Df_1^{(k)}\big((1-c)x_0 + c\hat{x}_{0|k}\big) \\ Df_2^{(k)}\big((1-c')x_0 + c'\hat{x}_{0|k}\big) \end{bmatrix} p_1) \right)^2, \quad (29)$$

$$\mathcal{B} = \frac{21}{2^{2\sum_{i=0}^{k} R_i^2}} \left( norm( \begin{bmatrix} Df_1^{(k)}\big((1-c)x_0 + c\hat{x}_{0|k}\big) \\ Df_2^{(k)}\big((1-c')x_0 + c'\hat{x}_{0|k}\big) \end{bmatrix} p_2) \right)^2. \quad (30)$$

Similar to the scalar case, the system is observable in sure sense provided $\lim_{k\to\infty} \frac{1}{k}\log(\mathcal{A}) < 0$ and $\lim_{k\to\infty} \frac{1}{k}\log(\mathcal{B}) < 0$. Note that $\hat{x}_{0|k}$ converges to $x_0$, therefore, (3) results in:

$$\lim_{k\to\infty} \frac{2}{k} \log(norm( \begin{bmatrix} Df_1^{(k)}\big((1-c)x_0 + c\hat{x}_{0|k}\big) \\ Df_2^{(k)}\big((1-c')x_0 + c'\hat{x}_{0|k}\big) \end{bmatrix} .p_i)) = 2\Delta_i(x_0), \quad (31)$$

where $i = 1, 2$. Now, substituting (31) in the aforementioned conditions, results in the following sufficient conditions for observability.

$$\lim_{k\to\infty} \frac{1}{k}\log(\mathcal{A}) = -2R_{av}^1 + 2\Delta_i(x_0) < 0, \quad (32)$$

$$\lim_{k\to\infty} \frac{1}{k}\log(\mathcal{A}) = -2R_{av}^2 + 2\Delta_2(x_0) < 0, \quad (33)$$

where $R_{av}^i = \lim_{k\to\infty} \frac{1}{k+1} \sum_{j=0}^{k} R_j^i$ for $i = 1, 2$. Similar to the scalar case, if we set $R_k^i = \Delta_i(\hat{x}_{0|k-1}) + \epsilon$ for $i = 1, 2$ and for every $\epsilon > 0$, the above two conditions are satisfied and the required capacity is $C = \Delta_1(x_0) + \Delta_1(x_0) + 2\epsilon$. Consequently, $C > \sum_{i=1}^{s} \kappa_i(x_0)\Delta_i(x_0)$ is the observability sufficient condition, i.e., under this condition, there is an encoding/decoding scheme to guarantee system observability in sure sense.

Finally, the following points should be taken into account:

*Remark 1:* For some communication/control systems, we are able to allocate average power/bit rate immediately after observation of the initial state $x_0$. Consequently, the channel capacity is assumed to be dependent of $x_0$ which is the assumption made in deriving sufficient conditions. However, if the channel capacity is determined independent of $x_0$, the sufficient condition is modified as $C > \sup_{x_0 \in \psi_0} \sum_{i=1}^{s} \kappa_i(x_0)\Delta_i(x_0)$.

*Remark 2:* Note that for the channels subject to fixed bit rate, the proposed encoding/decoding scheme is not practical. In this situation, we order Lyapunov exponent as $\Delta_1(x_0) \geq \Delta_2(x_0) \geq \ldots \geq \Delta_d(x_0)$ for every $x_0$ $\big($Note that if $k_i(x_0) \geq 2$, then we consider $\Delta_{i+j}(x_0)$ equals to $\Delta_i(x_0)$ for $j = 1, \ldots, k_i(x_0) - 1\big)$. Hence, it is easy to verify that applying aforementioned scheme with fixed rate $R = \sum_{i=1}^{d} \sup_{x_0} \Delta_i(x_0) + \epsilon$ for every $\epsilon > 0$ results in the observability of systems.

*Remark 3:* For those systems that have unique ergodic invariant measure, $L(x_0)$ in (2) is independent of $x_0$, then fixed rate encoding/decoding can be employed which leads to the tight sufficient and necessary conditions of observability.

*Remark 4:* For discrete linear system $x_{k+1} = f_k(x_k) = Ax_k$, as a special case, the Lyapunov exponents are the logarithm of the magnitudes of the eigenvalues of the matrix $A$ [19]. Hence, the necessary and sufficient condition is reduced to the well-known eigenvalues-rate condition $C > \sum_i \max\{0, \log(|\lambda_i(A)|)\}$ (also appearing in e.g., [25]–[29]).

## V. Conclusion

This paper addresses real-time reliable data reconstruction of the state trajectory of nonlinear dynamic systems over capacity limited communication channels. It is shown that over the memoryless channels (in particular two DMC and AWGN channels), the condition $C \geq E[\sum_{i=1}^{s} \kappa_i(x_0)\Delta_i(x_0)]$ is a necessary condition for sure; almost sure, and mean square observability of noiseless nonlinear dynamic systems, where $C$ is the Shannon channel capacity and $\Delta_i(x_0)$s are the positive Lyapunov exponents. It is also shown that the condition $C > \sum_{i=1}^{s} \kappa_i(x_0)\Delta_i(x_0)$ is a sufficient condition for sure, almost sure and mean square observability over, respectively, the digital noiseless channel, packet erasure channel and AWGN channel with noiseless feedback channel. Finally, it is shown that for the special case of linear noiseless time-invariant systems, the necessary and sufficient condition for observability is reduced to the well-known eigenvalues-rate condition.

For future work, it is interesting to find the necessary and sufficient conditions for sure, almost sure and mean square stability of nonlinear controlled dynamic systems over limited capacity communication channels. This research direction is currently under investigation.

## Appendix A

Time sharing technique for fractional bit:

As explained before, we set $R_i = \mathcal{H}(\hat{x}_{0|i-1})$ in order to obtain $R_{av} = \mathcal{H}(x_0)$. However, these bit rates may be fractional numbers. In the appearance of such problem, we can set new bit rates as follows:

$$\bar{R}_1 = \lfloor R_1 \rfloor, e_1 \triangleq R_1 - \bar{R}_1,$$
$$\bar{R}_2 = \lfloor R_2 + e_1 \rfloor, e_2 \triangleq R_2 + e_1 - \bar{R}_2,$$
$$\vdots$$
$$\bar{R}_{i+1} = \lfloor R_{i+1} + e_i \rfloor, e_{i+1} \triangleq R_{i+1} + e_i - \bar{R}_{i+1}.$$

Sum of new bit rates equals to:

$$\sum_{i=1}^{k}\bar{R}_i = R_1 - e_1 + \sum_{i=2}^{k}(R_i + e_{i-1} - e_i) = \sum_{i=1}^{k} R_i - e_k, \tag{34}$$

where $0 \leq e_k < 1$. Therefore, this technique keeps the average power as $\mathcal{H}(x_0)$ with transmission just integer bit rates.

## Appendix B

Note that $\mathcal{H}(\hat{x}_{0|k-1})$ converges to $\mathcal{H}(x_0)$ almost surely if and only if for every $\epsilon_1 > 0$, $\lim_{k_1\to\infty}\mathbb{P}(A_{k_1}(\epsilon_1)) = 1$, where $A_{k_1}(\epsilon_1) \triangleq \{|\mathcal{H}(\hat{x}_{0|k-1}) - \mathcal{H}(x_0)| < \epsilon_1, k \geq k_1\}$ [30].

Similarly, and based on the strong law of large numbers, we have: for every $\epsilon_2 > 0$, $\lim_{k_2\to\infty}\mathbb{P}(B_{k_2}(\epsilon_2)) = 1$ where $B_{k_2}(\epsilon_2) \triangleq \{|\frac{1}{k}\sum_{i=0}^{k-1}\beta_i - (1-\gamma)| < \epsilon_2, k \geq k_2\}$.

Proving $\lim_{k_3\to\infty}\mathbb{P}(C_{k_3}(\epsilon_3)) = 1$, for every $\epsilon_3 > 0$, leads to the almost sure convergence of the sequence $\frac{1}{k}\sum_{i=0}^{k-1}\beta_i\mathcal{H}(\hat{x}_{0|i-1})$ to $(1-\gamma)\mathcal{H}(x_0)$ where $C_{k_3}(\epsilon_3)$ is defined as $C_{k_3}(\epsilon_3) \triangleq \{|\frac{1}{k}\sum_{i=0}^{k-1}\beta_i\mathcal{H}(\hat{x}_{0|i-1}) - (1-\gamma)\mathcal{H}(x_0)| < \epsilon_3, k \geq k_3\}$.

Using the total law of probability:

$$\begin{aligned}\mathbb{P}(C_{k_3}(\epsilon_3)) &= \mathbb{P}(C_{k_3}(\epsilon_3) \mid A_{k_1}(\epsilon_1))\mathbb{P}(A_{k_1}(\epsilon_1)) \\ &+ \mathbb{P}(C_{k_3}(\epsilon_3) \mid A^c_{k_1}(\epsilon_1))\mathbb{P}(A^c_{k_1}(\epsilon_1)),\end{aligned} \tag{35}$$

where $k_1 = \frac{k_3}{T}$, $T \in \mathbb{Z}^+$ and $A^c_{k_1}(\epsilon_1)$ is the complement of event $A_{k_1}(\epsilon_1)$, we obtain a lower bound for $\mathbb{P}(C_{k_3}(\epsilon_3) \mid A_{k_1}(\epsilon_1))$ in the following. It is clear that for $k \geq k_3$, we have:

$$\begin{aligned}&\left|\frac{1}{k}\sum_{i=0}^{k-1}\beta_i\mathcal{H}(\hat{x}_{0|i-1}) - (1-\gamma)\mathcal{H}(x_0)\right| \leq \\ &\frac{1}{k}\left|\sum_{i=0}^{k-1}\beta_i\mathcal{H}(\hat{x}_{0|i-1}) - \sum_{i=0}^{k-1}\beta_i\mathcal{H}(x_0)\right| + \left|\frac{1}{k}\sum_{i=0}^{k-1}\beta_i - (1-\gamma)\right|\mathcal{H}(x_0).\end{aligned} \tag{36}$$

The following inequalities are obtained under the assumption that $A_{k_1}(\epsilon_1)$ has occurred:

$$\begin{aligned}&\frac{1}{k}\Big|\sum_{i=0}^{k-1}\beta_i\mathcal{H}(\hat{x}_{0|i-1}) - \sum_{i=0}^{k-1}\beta_i\mathcal{H}(x_0)\Big| \leq \\ &\Big|\frac{1}{k}\sum_{i=0}^{k_1-1}\beta_i\big(\mathcal{H}(\hat{x}_{0|i-1}) - \mathcal{H}(x_0)\big)\Big| + \Big|\frac{1}{k}\sum_{i=k_1}^{k-1}\beta_i\big(\mathcal{H}(\hat{x}_{0|i-1}) - \mathcal{H}(x_0)\big)\Big|, \\ &\leq \frac{k_1}{k}M + \frac{k-k_1}{k}\epsilon_1 \leq \frac{M}{T} + \epsilon_1,\end{aligned} \tag{37}$$

where $M \triangleq \max\big(\mathcal{H}(\hat{x}_{0|i-1}) - \mathcal{H}(x_0)\big)$ which is a finite number. Therefore,

$$\begin{aligned}&\mathbb{P}(C_{k_3}(\epsilon_3) \mid A_{k_1}(\epsilon_1)) \\ &\geq \mathbb{P}\big(\{\frac{M}{T} + \epsilon_1 + \big|\frac{1}{k}\sum_{i=0}^{k-1}\beta_i - (1-\gamma)\big|\mathcal{H}(x_0) < \epsilon_3, k \geq k_3\}\big), \\ &= \mathbb{P}\big(\{\big|\frac{1}{k}\sum_{i=0}^{k-1}\beta_i - (1-\gamma)\big| < \frac{\epsilon_3 - \frac{M}{T} - \epsilon_1}{\mathcal{H}(x_0)}, k \geq k_3\}\big).\end{aligned} \tag{38}$$

For every $\epsilon_3 > 0$, we choose $T$ and $\epsilon_1$ such that $(\epsilon_3 - \frac{M}{T} - \epsilon_1) > 0$. Therefore, when $k_3$ goes to infinity:

$$\begin{aligned}&\lim_{k_3\to\infty}\mathbb{P}(C_{k_3}(\epsilon_3) \mid A_{k_1}(\epsilon_1)) \\ &\geq \lim_{k_3\to\infty}\mathbb{P}\left(B_{k_3}(\frac{1}{\mathcal{H}(x_0)}(\epsilon_3 - \frac{M}{T} - \epsilon_1))\right) = 1.\end{aligned} \tag{39}$$

On the other hand, $k_1 = \frac{k_3}{T}$ also goes to infinity and therefore, $\mathbb{P}(A_{k_1}(\epsilon_1))$ converges to one. Consequently, for every $\epsilon_3 > 0$, $\lim_{k_3\to\infty}\mathbb{P}(C_{k_3}(\epsilon_3)) = 1$.

## Appendix C

We start the proof from the following relation:

$$e_{0|k} = \rho_{0|k}n_0 + \rho_{1|k}n_1 + \ldots + \rho_{k|k}n_k. \tag{40}$$

Substituting (20) in error variance leads to the following equation:

$$\begin{aligned}E(e^2_{0|k}|x_0) &= E\Bigg[\frac{1}{\prod_{i=0}^{k}\alpha_i^2}\Bigg(n_k^2\sum_{i=1}^{k}\frac{n_{i-1}^2 - n_i^2}{\prod_{j=i}^{k}\alpha_j^2} + 2\sum_{i=1}^{k}\frac{n_0 n_i(\alpha_i^2-1)^{\frac{1}{2}}}{\prod_{m=1}^{i-1}\alpha_m\prod_{n=i}^{k}\alpha_n^2} \\ &+ 2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{n_i n_j(\alpha_i^2-1)^{\frac{1}{2}}(\alpha_j^2-1)^{\frac{1}{2}}}{\prod_{m=i}^{j-1}\alpha_m\prod_{n=j}^{k}\alpha_n^2}\Bigg)|x_0\Bigg], \\ &= E_{\alpha_0^k|x_0}\Bigg[\frac{1}{\prod_{i=0}^{k}\alpha_i^2}\Bigg(\sigma^2 + \sum_{i=1}^{k}\frac{1}{\prod_{j=i}^{k}\alpha_j^2}E[n_{i-1}^2 - n_i^2|x_0,\alpha_i^k] \\ &+ 2\sum_{i=1}^{k}\frac{(\alpha_i^2-1)^{\frac{1}{2}}}{\prod_{m=1}^{i-1}\alpha_m\prod_{n=i}^{k}\alpha_n^2}E[n_0 n_i|x_0,\alpha_1^k] \\ &+ 2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{(\alpha_i^2-1)^{\frac{1}{2}}(\alpha_j^2-1)^{\frac{1}{2}}}{\prod_{m=i}^{j-1}\alpha_m\prod_{n=j}^{k}\alpha_n^2}E[n_i n_j|x_0,\alpha_{i+1}^k]\Bigg)\Bigg].\end{aligned} \tag{41}$$

We show that the term inside the parenthesis in (41) converges to $\sigma^2$ as $k$ goes to infinity. Toward this purpose, we consider the following steps:

1) A new sequence $u_k$ is defined as

$$u_{k+1} \triangleq \sum_{i=1}^{k}\frac{1}{\prod_{j=i}^{k}\alpha_j^2}E[n_{i-1}^2 - n_i^2|x_0,\alpha_i^\infty].$$

Note that $u_k$ can be also determined thorough the following iteration formula:

$$u_{k+1} = \frac{1}{\alpha_{k+1}^2}u_k + \frac{1}{\alpha_{k+1}^2}E[n_k^2 - n_{k+1}^2|x_0,\alpha_{k+1}^\infty]. \tag{42}$$

Due to the convergence of $\alpha_i$ to $\alpha(x_0)$ in the probability sense, the zero convergence of $E[n_k^2 - n_{k+1}^2|x_0,\alpha_{k+1}^\infty]$ is resulted, i. e.,

$$\lim_{k\to\infty}E[n_k^2 - n_{k+1}^2|x_0,\alpha_{k+1}^\infty] = \lim_{k\to\infty}E[n_k^2 - n_{k+1}^2|x_0,\alpha(x_0)] = 0.$$

Furthermore, $\alpha_{k+1}$ is greater than one and hence, the limit point of sequence $u_k$ which is also the limit point of the sequence $\sum_{i=1}^{k}\frac{1}{\prod_{j=i}^{k}\alpha_j^2}E[n_{i-1}^2 - n_i^2|x_0,\alpha_i^k]$ converges to zero [3].

2) As $\alpha_i$s are greater than one, it is obvious that

$$\lim_{k\to\infty}\sum_{i=1}^{k}\frac{(\alpha_i^2-1)^{\frac{1}{2}}}{\prod_{m=1}^{i-1}\alpha_m\prod_{n=i}^{k}\alpha_n^2}E[n_0 n_i|x_0,\alpha_1^k] = 0.$$

3) We define new sequence

$$r_k = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{(\alpha_i^2-1)^{\frac{1}{2}}(\alpha_j^2-1)^{\frac{1}{2}}}{\prod_{m=i}^{j-1}\alpha_m\prod_{n=j}^{k}\alpha_n^2}E[n_i n_j|x_0,\alpha_{i+1}^\infty].$$

By deriving the iteration formula of $r_k$, we can prove that $r_k$ converges to zero. Then, note that the sequence $\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{(\alpha_i^2-1)^{\frac{1}{2}}(\alpha_j^2-1)^{\frac{1}{2}}}{\prod_{m=i}^{j-1}\alpha_m\prod_{n=j}^{k}\alpha_n^2}E[n_i n_j|x_0,\alpha_{i+1}^k]$ converges to the limit point of $r_k$ which equals to zero.

Consequently, we can write,

$$Df^{(k)}(x_0)^2 E(e_{0|k}^2|x_0) = Df^{(k)}(x_0)^2 E\left[\frac{\sigma^2+\zeta_k}{\prod_{i=0}^{k}\alpha_i^2} \mid x_0\right],$$

where the sequence $\zeta_k$ goes to zero. In the following, we show that the sequence $Df^{(k)}(x_0)^2 E\left[\frac{1}{\prod_{i=0}^{k}\alpha_i^2} \mid x_0\right]$ converges to zero (Note that this convergence also implies the zero convergence of $Df^{k}(x_0)^2 E(e_{0|k}^2|x_0)$). By setting $\alpha_i = 2^{(\Delta(\hat{x}_{0|i-1})+\epsilon)}$, we have:

$$\begin{aligned}
&Df^{(k)}(x_0)^2 E\left[\frac{1}{\prod_{i=0}^{k}\alpha_i^2} \mid x_0\right] \\
&= \frac{Df^{(k)}(x_0)^2}{\alpha_0 2^{2k\epsilon}} E\left[\frac{1}{2^{2\sum_{i=0}^{k-1}\Delta(\hat{x}_{0|i})}} \mid x_0\right], \\
&\overset{(a)}{\leq} \frac{Df^{(k)}(x_0)^2}{\alpha_0 2^{2k(\Delta(x_0)+\epsilon)}} E\left[2^{2\bar{c}(\sum_{i=0}^{k-1}|e_{0|i}|)}|x_0\right], \\
&\leq \frac{Df^{(k)}(x_0)^2}{\alpha_0 2^{2k(\Delta(x_0)+\epsilon)}} E\left[2^{2\bar{c}(\sum_{i=0}^{k-1}\sum_{j=0}^{i}\rho_{j|i}|n_j|)}|x_0\right], \\
&\overset{(b)}{\leq} \frac{Df^{(k)}(x_0)^2}{\alpha_0 2^{2k(\Delta(x_0)+\epsilon)}} E\left[2^{2\bar{c}(\sum_{j=0}^{k-1}\sum_{i=j}^{k-1}\frac{\bar{\rho}}{2^{\epsilon(2i-j)}}|n_j|)}|x_0\right], \\
&\overset{(c)}{=} \frac{Df^{(k)}(x_0)^2}{\alpha_0 2^{2k(\Delta(x_0)+\epsilon)}} \prod_{j=0}^{k-1} E\left[2^{2\bar{c}(\sum_{i=j}^{k-1}\frac{\bar{\rho}}{2^{\epsilon(2i-j)}}|n_j|)}\right].
\end{aligned} \tag{43}$$

Note that $(a)$ follows because based on Assumption 5, there is a constant number $\bar{c}$ such that $|d\Delta(x_0)/dx| \leq \bar{c}$ (which results in $|\Delta(\hat{x}_{0|i-1}) - \Delta(x_0)| \leq |e_{0|i-1}|\bar{c}$). $(b)$ follows since we have $\alpha_i \geq 2^{\epsilon}$ for every $i = 1, \ldots, k-1$ (Due to the positive value of $\Delta(x)$) and therefore, $\rho_{j|i} \leq \frac{\bar{\rho}}{2^{\epsilon(2i-j)}}$, where $\bar{\rho}$ is a constant number. Because of the indecency of noise samples, $(c)$ is deduced. It can be proved that $\prod_{j=0}^{k-1} E\left[2^{2\bar{c}(\sum_{i=j}^{k-1}\frac{\bar{\rho}}{2^{\epsilon(2i-j)}}|n_j|)}\right]$ is less than a constant number. Furthermore, we have for every $x_0 \in \psi_0$:

$$\lim_{k\to\infty}\left(\frac{Df^{(k)}(x_0)^2}{2^{2k(\Delta(x_0)+\epsilon)}}\right)^{\frac{1}{k}} = \frac{2^{2\Delta(x_0)}}{2^{2\Delta(x_0)+2\epsilon}} = \frac{1}{2^{2\epsilon}} < 1, \tag{44}$$

Therefore, (43) converges to zero and

$$\lim_{k\to\infty} Df^{(k)}(x_0)^2 E[e_{0|k}^2|x_0] = 0. \tag{45}$$

Similarly, it can be proved that for every $x_0$:

$$\lim_{k\to\infty}\sum_{t=1}^{\infty}\sum_{n=2}^{\infty} Df^{(k)}(x_0)^{(t+n)} E[e_{0|k}^{n+t}|x_0] = 0. \tag{46}$$

Substituting (45) and (46) in (22) results in the convergence of $E[(x_k - \hat{x}_{k|k})^2|x_0]$ to zero for every $x_0$, therefore, $E[(x_k - \hat{x}_{k|k})^2]$ also converges to zero and the system is observable in the mean square sense.

## References


[1] G. Fettweis and S. Alamouti, "5G: personal mobile internet beyond what cellular did to telephony," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 140–145, 2014.

[2] G. Fettweis, "The tactile internet: applications and challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, p. 6470, 2014.

[3] S. Tatikonda and S. Mitter, "Control over noisy channels," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1196–1201, 2004.

[4] A. S. Matveev and A. V. Savkin, "Shannon zero error capacity in the problems of state estimation and stabilization via noisy communication channels," *International Journal of Control*, vol. 80, no. 2, pp. 241–255, 2007.

[5] C. D. Charalambous, A. Farhadi, and S. Z. Denic, "Control of continuous-time linear Gaussian systems over additive Gaussian wireless fading channels: A separation principle," *IEEE Trans. Autom. Control*, vol. 53, no. 4, pp. 1013–1019, 2008.

[6] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 31 549 – 1561, 2004.

[7] C. D. Charalambous and A. Farhadi, "LQG optimality and separation principle for general discrete time partially observed stochastic systems over finite capacity communication channels," *Automatica*, vol. 44, no. 12, pp. 3181–3188, 2008.

[8] F. Lian, A. Chakrabortty, and A. Duel-Hallen, "Game-theoretic multi-agent control and network cost allocation under communication constraints," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 2, pp. 330–340, 2017.

[9] G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran, "Topological feedback entropy and nonlinear stabilization," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1585–1597, 2004.

[10] A. V. Savkin, "Analysis and synthesis of networked control systems: Topological entropy, observability, robustness and optimal control," *Automatica*, vol. 42, no. 1, pp. 51–62, 2006.

[11] F. Colonius, "Minimal bit rates and entropy for exponential stabilization," *SIAM J. Control Optim.*, vol. 50, no. 5, pp. 2988–3010, 2012.

[12] D. Liberzon and J. P. Hespanha, "Stabilization of nonlinear systems with limited information feedback," *IEEE Trans. Autom. Control*, vol. 50, no. 6, pp. 910–915, 2005.

[13] G. N. Nair and R. J. Evans, "State estimation via a capacity-limited communication channel," *Proceedings of the 36th Conference on Decision and Control*, vol. 49, no. 9, pp. 2488–2498, 1997.

[14] A. Farhadi and N. U. Ahmed, "Tracking nonlinear noisy dynamic systems over noisy communication channels," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 955–961, 2011.

[15] D. Liberzon and S. Mitra, "Entropy and minimal data rates for state estimation and model detection," *In Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control. ACM*, pp. 247–256, 2016.

[16] A. Diwadkar and U. Vaidya, "Limitation for nonlinear observation over erasure channel," *IEEE Trans. Autom. Control*, vol. 58, no. 2, pp. 454–459, 2013.

[17] U. Vadia. and N. Elia, "Stabilization of nonlinear systems over packet-drop links: scalar case," *Systems and Control Letters*, vol. 61, no. 9, pp. 959–966, 2012.

[18] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1056–1068, 2004.

[19] C. Skokos, "The lyapunov characteristic exponents and their computation," *Dynamics of Small Solar System Bodies and Exoplanets. Springer Berlin Heidelberg*, pp. 63–135, 2010.

[20] T. M. cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2006.

[21] J. V. Michalowicz, J. M. Nichols, and F. Bucholtz, *Handbook of differential entropy*. CRC Press., 2013.

[22] R. Durrett, *Probability: Theory and Examples*. Cambridge university press., 2010.

[23] J. Schalkwijk, "Coding scheme for additive noise channels with feedback - part II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 183–189, 1966.

[24] L. S. Young, "Entropy, lyapunov exponent, and hausdorff dimension in differentiable dynamic systems," *IEEE Trans. Circuits Syst.*, vol. 30, no. 8, pp. 599–607, 1983.

[25] N. Elia, "When bode meets Shannon: control-oriented feedback communication schemes," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1477–1488, 2004.

[26] N. Elia and J. N. Eisenbeis, "Limitations of linear control over packet drop networks," *IEEE Trans. Autom. Control*, vol. 56, no. 4, pp. 826–841, 2011.

[27] N. C. Martins, A. Dahleh, and N. Elia, "Feedback stabilization of uncertain systems in the presence of a direct link feedback stabilization of uncertain systems in the presence of a direct link," *IEEE Trans. Autom. Control*, vol. 51, no. 3, pp. 438–447, 2006.

[28] P. Minero, M. Franceschetti, S. Dey, and N. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 243–255, 2009.

[29] P. Minero, L. Coviello, and M. Franceschetti, "Stabilization over Markov feedback channels: the general case," *IEEE Trans. Autom. Control*, vol. 58, no. 2, pp. 349–362, 2013.

[30] H. Pishro-Nik, *Introduction to probability, statistics and random processes*. Kappa Research, LLC., 2014.