Shannon Entropy And Tracking Dynamic Systems Over Noisy Channels

Alireza Farhadi *

November 30, 2016

Abstract

This paper is concerned with the estimation of state trajectory of linear discrete time dynamic systems subject to parametric uncertainty over the compound erasure channel that uses feedback channel intermittently. For this combined system and channel, using the data processing inequality and a robust version of the Shannon lower bound, a necessary condition on channel capacity for estimation of state trajectory at the receiver giving almost sure asymptotically zero estimation error is presented. Then, an estimation technique over the compound erasure channel that includes an encoder, decoder and a sufficient condition under which the estimation error at the receiver is asymptotically zero almost surely is presented. This leads to the conclusion that over the compound erasure channel, a condition on Shannon capacity in terms of the rate of expansion of the Shannon entropy is a necessary and sufficient condition for estimation with uniform almost sure asymptotically zero estimation error. The satisfactory performance of the proposed technique is illustrated using simulation.

Keywords: Estimation, networked control system, Shannon entropy.

1 Introduction

1.1 Motivation and Background

One of the issues that has begun to emerge in a number of applications, such as networked control systems[1]-[22], is how to estimate the state trajectory of a dynamic system over a communication channel subject to imperfections (e.g., noise, limited capacity). In these applications, estimation 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 methodologies for designing proper estimator over, for example, data

*A. Farhadi is Assistant professor in the Department of Electrical Engineering at Sharif University of Technology, email: afarhadi@sharif.edu. This work was supported by the research office of Sharif University of Technology.

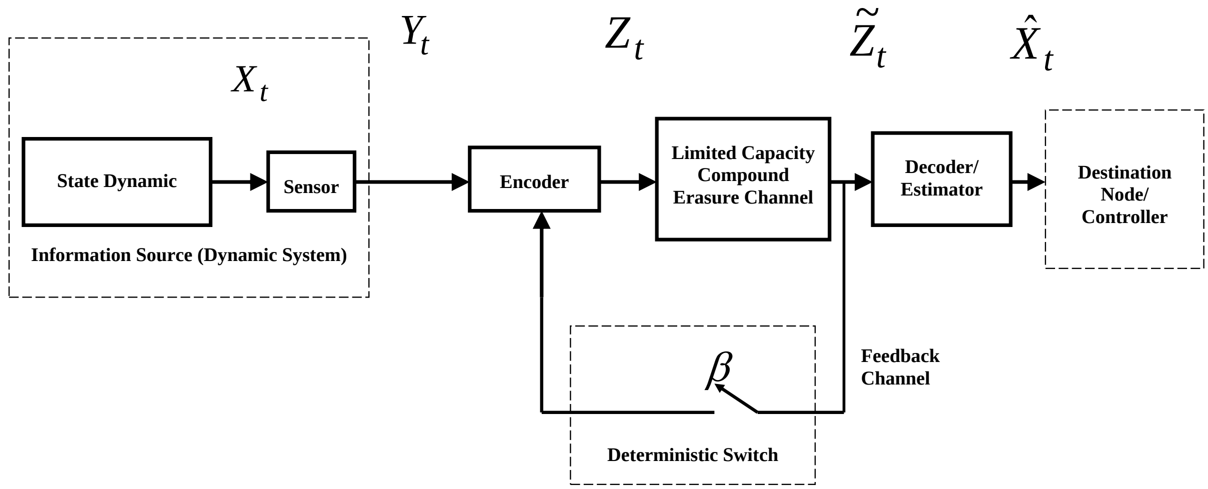

Figure 1: Communication System (CS).

links subject to data dropout and limited capacity.

In this paper we consider the problem of estimating the state trajectory of a linear time-invariant dynamic system subject to parametric uncertainty over the limited capacity compound erasure channel (i.e., the packet erasure channel with unknown erasure probability) that uses feedback channel with duty cycle $\beta \in (0, 1]$, where β is a rational number as is shown in Fig. 1. $\beta = 0$ means non-availability of feedback channel while $\beta = 1$ means full time availability of feedback channel. In some emerging applications, such as tele-operation of micro autonomous Unmanned Aerial Vehicles (UAVs) and Autonomous Underwater Vehicles (AUVs), it is necessary to transmit the observation signal of the dynamic system to remote base station where the controller is located over a limited capacity communication channel subject to imperfections. Thus, in these applications it is necessary to estimate the state trajectory of the dynamic system using the transmitted observation signal over communication channel to produce the proper control commands. The block diagram of Fig. 1 presents the block diagram of this estimation problem that occurs in tele-operation of micro UAVs and AUVs.

Although it is known in the context of information theory that feedback channel does not increase the capacity of Discrete Memoryless Channels (DMCs), it significantly simplifies the design of encoder and decoder [23] that compensate the effects of channel imperfections and reliably estimate the state of the system in control over communication problems. The simplicity in the design of encoder and decoder is a required property in control over communication as complicated design results in significant time latency between making measurements from dynamic system and applying the corresponding control commands, which obviously damages the performance. This motivated us to use feedback channel in the block diagram of Fig. 1. Nevertheless, the full time availability of feedback channel (i.e. $\beta = 1$) in the block diagram of Fig. 1 requires that the feedback channel signal is transmitted with high power full time. This results in significant power consumption at receiver in order to transmit noiseless feedback channel full time. Therefore, it is more desirable to use noiseless feedback channel that is only available intermittently (i.e., $\beta \in (0, 1)$) to avoid exhausting receiver power supply.

In the block diagram of Fig. 1, the communication channel is a limited capacity compound erasure channel. This channel is a packet erasure channel [18] with unknown erasure probability. The packet erasure channel is an abstract model for the commonly used information technologies, such as the Internet, WiFi and mobile communication; and for this reason we consider it in this paper.

As the class of linear dynamic systems is an important class of systems, in the block diagram of Fig. 1 we are concerned with the class of linear time-invariant dynamic systems. Having chosen a suitable dynamic model for the system, the defining parameters of the system are usually estimated from a sample of experimental data. But, there is always uncertainty associated with any estimation. This results in parametric uncertainty in many models. Therefore, systems subject to parametric uncertainty form a large and important class of dynamic systems and have been considered in many studies, such as [24]-[27]. Hence, in this paper we are concerned with the state estimation of linear time-invariant dynamic systems subject to parametric uncertainty over the compound erasure channel.

The problem of almost sure estimation of linear time-invariant dynamic systems over the packet erasure channel that uses feedback channel full time has been addressed in the literature (e.g., [11], [18]). In [18] it was shown that the eigenvalues rate condition described by the Shannon capacity is tight. That is, the following condition $\mathcal{C} \geq \sum_{\{i:|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$, where $\mathcal{C}$ denotes the capacity of the DMCs and $\lambda_i(A)$s denote the eigenvalues of linear time-invariant noiseless dynamic system is the necessary and sufficient condition for an estimation over DMCs with almost sure asymptotically zero estimation error. Linear time-invariant dynamic systems subject to uniformly bounded exogenous disturbances over the DMCs (with and without feedback channel) has been also considered in the literature, in which it has been shown that the following condition $\mathcal{C}_0 \geq \sum_{\{i:|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$, where $\mathcal{C}_0$ denotes the Shannon zero error capacity [23] is the necessary and sufficient condition for estimation over DMCs (with and without feedback channel) giving almost sure asymptotically bounded estimation error [22] (i.e., $\limsup_{t\to\infty} ||X_t - \hat{X}_t|| < \infty$, where X_t is the state of the system and $\hat{X}_t$ is its estimate). As the Shannon zero error capacity of noisy DMCs with and without feedback channel is zero, this result indicates that we cannot estimate the states of linear time-invariant dynamic systems subject to uniformly bounded exogenous disturbances over noisy communication channels. The problem of almost sure bounded stability of controlled nonlinear Lipschitz systems over the digital noiseless channel was addressed in [17], where a sufficient condition relating transmission rate to Lipschitz coefficient is presented for almost sure asymptotic bounded stability. Note that the problem that will be addressed in this paper is quite different as we consider different dynamic system, different communication channel and different objective. In this paper we present the necessary and sufficient condition for uniform almost sure asymptotic estimation of linear time-invariant dynamic systems over the limited capacity compound erasure channel.The problem of optimal reference tracking of linear time-invariant dynamic systems over Additive White Gaussian Noise (AWGN) channel was addressed in [19]. Also, the problem of optimal reference tracking of linear time-invariant dynamic systems over the packet erasure channel with known erasure probability was addressed in [20]. Moreover, the problem of optimal reference tracking of linear time-invariant systems over AWGN channel in feedback path or forward path was addressed in [21]. Note that in [19]-[20], the objective is that the system output follows a desired reference signal; while in this paper the objective is to estimate the state trajectory of the system at the end of communication as is shown in Fig. 1.

1.2 Paper Contributions

The main novelty of this paper is in the necessary and sufficient condition for Uniform Almost Sure Asymptotic Estimation (UASAE) over the limited capacity compound erasure channel.

For the block diagram of Fig. 1 we present necessary and sufficient conditions that provide UASAE when feedback channel is not necessarily available full time. Using an information theoretic approach, we derive a necessary condition for this type of estimation over the compound erasure channel. This necessary condition is given in terms of the Shannon capacity and the Shannon lower bound, which is related to the rate of expansion of the Shannon entropy of the dynamic system. This leads to the eigenvalues rate condition described by the Shannon capacity. In the absence of uncertainty in the dynamic system, we are also able to present a sufficient condition for which UASAE holds over the compound erasure channel. The sufficient condition is also given in terms of the rate of expansion of the Shannon entropy of the dynamic system. Hence, this paper extends the previous results (e.g.,

[18]) to cases where both the dynamic system and communication channel are subject to uncertainty and feedback channel is not necessarily available full time. It also complements the previous results (e.g., [22]) by considering parametric uncertainty instead of uniformly bounded disturbances and showing that in the presence of uncertainty in the communication channel, the eigenvalues rate condition described by the Shannon capacity is a tight bound for UASAE.

1.3 Paper Organization

The paper is organized as follows. Section 2 is devoted to the problem formulation. In this section we describe the compound erasure channel and the notion of uniform almost sure asymptotic estimation. In Section 3, we first describe the notions of Shannon capacity, rate distortion and Shannon entropy. Then, a necessary condition for UASAE is presented. In Section 4, a sufficient condition for this type of estimation is given. Finally, the paper is concluded in Section 5. Proofs are given in the Appendix.

2 Problem Formulation

This paper is concerned with the communication system CS shown in Fig. 1 which is defined on a complete probability space $(\Omega, \mathcal{F}(\Omega), P)$ with filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Here, X_t, Y_t, Z_t, $\tilde{Z}_t$, and $\hat{X}_t$ are random vectors denoting the state of the dynamic system, observation, channel input, channel output and the state estimate, respectively, at time $t \in \mathbf{N}_+ \equiv \{0, 1, 2, ...\}$. Note that $X_t \in \mathbf{R}^q$, $Y_t \in \mathbf{R}^d$ and $\hat{X}_t \in \mathbf{R}^q$. In the CS shown in Fig. 1 we can use feedback channel with duty cycle $\beta \in (0, 1]$, where β is a rational number. $\beta = 0$ corresponds to the case of non-availability of feedback channel while $\beta = 1$ corresponds to its full time availability. This is shown by a switch with a known policy.

Throughout the paper we adopt the following notation. Random Vectors (R.V.s) are denoted by capital letters, while a realization of a R.V. is denoted by a lower case letter. Sequences of R.V.s are denoted by $Y_0^T \equiv (Y_0, Y_1, ..., Y_T)$. We denote by A' the transpose of A, where A is either a matrix or a vector, and by A^{-1} the inverse of a square invertible matrix A. We denote by $||\cdot||$ the Euclidean norm on the vector space $\mathbf{R}^q$, by $|x|$ the absolute value of a scalar $x \in \mathbf{R}$ and by $M(q \times o)$ the space of all matrices $A \in \mathbf{R}^{q\times o}$. The space $M(q \times o)$ is endowed with the spectral norm $||A|| \equiv \sqrt{\lambda_{max}(A'A)}$ where $\lambda_{max}(A'A)$ is the largest eigenvalue of the matrix $A'A$. We denote by $\mathcal{B}(\mathcal{X})$ the Borel σ-algebra generated by the open subsets of the non-empty (arbitrary) set $\mathcal{X}$ and by $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ the Borel measurable space. We also denote by $\mathcal{M}_1(\mathcal{X})$ the space of probability measures defined on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Given a pair of measurable spaces $(\bar{A}, \mathcal{B}(\bar{A}))$ and $(\hat{A}, \mathcal{B}(\hat{A}))$, a mapping $Q : \mathcal{B}(\hat{A}) \times \bar{A} \to [0,1]$ is called a stochastic kernel if it satisfies the following two properties: i) for every $x \in \bar{A}$, the set function $Q(\cdot|x)$ is a probability measure on $\hat{A}$ and ii) for every $F \in \mathcal{B}(\hat{A})$, the function $Q(F|\cdot)$ is $\bar{A}$-measurable. Time ordered products are denoted by $\prod_{t\geq k\geq 0} a_k \equiv a_t.a_{t-1}.....a_0$.

The different blocks of Fig. 1 representing the CS are described below.

Information Source: The information source is described by a dynamic system subject to parametric uncertainty.

In this paper we are concerned with an information source described by the following uncertain linear discrete time dynamic system

$$\left\{ \begin{array}{l} X_{t+1} = (A + \Gamma_t)X_t, \quad X_0 = x_o, \;\; A \neq 0, \;\; \Gamma_t \in B_\ell(M(q \times q)), \\ \quad Y_t = CX_t \end{array} \right. \tag{1}$$

where $X_t \in \mathbf{R}^q$ is the state of the system at time $t \in \mathbf{N}_+$, $A \in M(q \times q)$ is the nominal (known) system matrix, initial condition X_0 has a known density denoted by p_0, $Y_t \in \mathbf{R}^d$ is the observation (the source output) and $C \in M(d \times q)$. The unknown matrix $\Gamma_t \in M(q \times q)$ represents the uncertainty in the system parameters. At any time $t \in \mathbf{N}_+$, it is a measurable function $\Gamma_t : (\Omega, \mathcal{F}(\Omega)) \to B_\ell(M(q \times q)) \equiv \{\Gamma \in M(q \times q) : ||\Gamma|| \leq \ell\}$ where ℓ is a known non-negative scalar representing the radius of uncertainty.

Let P_t^x, $t \in \mathbf{N}_+$ denote the probability measure associated with the state variable $X_t \in \mathbf{R}^q$, i.e., $P_t^x(B) \equiv P(X_t \in B)$, $B \in \mathcal{B}(\mathbf{R}^q)$. We assume that P_t^x admits a density function p_t, i.e., $P_t^x(B) = \int_B p_t(x)dx$, $\forall B \in \mathcal{B}(\mathbf{R}^q)$. Since the system (1) is uncertain, the information source is also uncertain, in which the uncertainty in source is characterized by a family of probability measures $\mathcal{M}_{S,t}(\subset \mathcal{M}_1(\mathbf{R}^q))$ induced by the R.V. $X_t = \Big(\prod_{t-1\geq k\geq 0}(A+\Gamma_k)\Big)X_0$ where Γ_k takes values from the set $B_\ell(M(q \times q))$. In other words, the source is represented by the family of probability measures $P_t^x \in \mathcal{M}_{S,t}$. Note that the case $\ell = 0$ corresponds to a known system, which is the nominal system.

Communication Channel: From the bit-wise perspective the channel input is '0', '1' and 'idle' and the channel output is '0', '1' and 'idle'. That is when the channel is not in use, the channel is in 'idle' mode. But, from the packet-wise, in this paper we are concerned with the discrete memoryless limited capacity compound erasure channel. It is a packet erasure channel [23] with the channel input alphabet $\mathbf{Z} = \{0,1\}^{\mathcal{R}}$ (where $\mathcal{R} \in \{1,2,3,...\}$ is the length of transmitted packet in a channel use) and the output alphabet $\tilde{\mathbf{Z}} = \mathbf{Z} \bigcup \{e\}$ (where e stands for the erasure symbol). The erasure probability θ is unknown to both transmitter and receiver. However, the unknown erasure probability θ belongs to a known set Θ, which is a compact (and proper) subset of $[0,1)$.

Encoder: We introduce three classes of encoders, Class A, Class B, and Class C. Let $E \in \mathcal{B}(\mathbf{Z})$ and $\bar{z}_0^{t-1} = (r_0\tilde{z}_0, r_1\tilde{z}_1, ..., r_{t-1}\tilde{z}_{t-1})$, where $r_j = 0$ corresponds to the case of inactive feedback channel at time instant j and $r_j = 1$ corresponds to the feedback channel being active. At any time $t \in \mathbf{N}_+$, Class A, Class B, and Class C encoders are modeled by stochastic kernels $Q_t^A(E|y_0^t, z_0^{t-1})$, $Q_t^B(E|y_0^t, z_0^{t-1}, \tilde{z}_0^{t-1})$, and $Q_t^C(E|y_0^t, z_0^{t-1}, \bar{z}_0^{t-1})$, respectively. Note that the Class A encoder does not use the channel outputs, Class B encoder can use all the channel outputs up to time $t-1$ via feedback channel, and Class C encoders can only use (via feedback) some of the channel outputs.

Decoder/Estimator: At any time $t \in \mathbf{N}_+$, the decoder is a mapping from the channel outputs $\tilde{Z}_0^t$ to the state estimate $\hat{X}_t \in \mathbf{R}^q$. It is described by a stochastic kernel $Q_t^D(E|\tilde{z}_0^t)$, $E \in \mathcal{B}(\mathbf{R}^q)$. Note that as the channel output is ternary, the decoder can identify the length of transmitted packet.

Deterministic Switch: In many practical applications providing a noiseless feedback acknowledgment from receiver to transmitter in each time step is difficult and/or expensive. Therefore, in this paper we use feedback channel with duty cycle $\beta \in (0,1]$, where $\beta = 0$ corresponds to the inactive state of the feedback channel while $\beta = 1$ corresponds to the active state (i.e., available all the time). This is shown by a switch with a known switching policy in the CS shown in Fig. 1.

In many applications a tracker of a signal process, giving almost sure zero estimation error, is desirable as it results in almost sure stability of the controlled system. Therefore, in this paper we are concerned with this type of estimation. The objective is to find necessary and sufficient conditions on the Shannon capacity for which UASAE, as defined below, holds.

Definition 2.1 *Consider the CS shown in Fig. 1, described by the uncertain system (1) and the compound erasure channel. For the system (1) UASAE holds if there exist an encoder and a decoder such that for any $\epsilon > 0$ there exists a finite time $T(\epsilon) \geq 0$, such that*

$$P(\sup_{t \geq T(\epsilon)} ||X_t - \hat{X}_t|| > \epsilon) \leq \epsilon,$$

$\forall\ \Gamma_k \in B_\ell(M(q \times q))$ *$(k \leq t-1)$ and* $\forall\ \theta \in \Theta$.

Throughout the paper, it is assumed that the erasure probability θ and the distribution of Γ_t are not known to transmitter and receiver. But, the set Θ, the non-negative scalar representing the radius of uncertainty, ℓ, and β are known a priori.

3 Necessary Condition

In this section, a necessary condition for which UASAE holds is derived for the CS shown in Fig. 1. This condition is obtained by establishing a relationship between the Shannon capacity, robust rate distortion, and a variant of the Shannon lower bound. These are information theoretic measures which we recall here.

Consider the compound erasure channel, as described earlier, and let Z_0^t and $\tilde{Z}_0^t$ be sequences of the channel input and output symbols, respectively. Let $\mathcal{Z}_{0,t} \equiv \prod_{k=0}^t \mathbf{Z}$ denote the space which contains Z_0^t (i.e., $Z_0^t \in \mathcal{Z}_{0,t}$). For any given erasure probability $\theta \in \Theta$, let $Q_{0,t}^\theta(d\tilde{z}|z)$ denote the stochastic kernel corresponding to $\tilde{Z}_0^t$ and Z_0^t. Further, let $P_{0,t}^z(B) \equiv P(Z_0^t \in B)$, $B \in \mathcal{B}(\mathcal{Z}_{0,t})$ denote the probability measure corresponding to the sequence Z_0^t. The Shannon capacity of the above channel, which is the maximum rate in bits per channel use at which information can be sent with arbitrary low probability of error, is defined as follows:

Definition 3.1 (The Shannon Capacity Of The Compound Erasure Channel) *[28] Consider the compound erasure channel, as described above. The capacity of this channel for $t+1$ channel uses is defined by $\mathcal{C}_t \equiv \sup_{P_{0,t}^z \in \mathcal{M}_1(\mathcal{Z}_{0,t})} \inf_{\theta\in\Theta} I^\theta(Z_0^t, \tilde{Z}_0^t)$ where*

$$I^\theta(Z_0^t, \tilde{Z}_0^t) \equiv \int\int \log\Big(\frac{Q_{0,t}^\theta(d\tilde{z}|z)}{\int Q_{0,t}^\theta(d\tilde{z}|z)P_{0,t}^z(dz)}\Big)Q_{0,t}^\theta(d\tilde{z}|z)P_{0,t}^z(dz)$$

denotes the mutual information between sequences Z_0^t and $\tilde{Z}_0^t$ (and the superscript θ emphasizes the dependency of the mutual information on parameter θ). Then, the Shannon capacity in bits per channel use is given by $\mathcal{C} \equiv \lim_{t\to\infty} \frac{1}{t+1}\mathcal{C}_t$.

For the compound erasure channel, the capacity achieving input probability mass function is the same for all the channels in the family $\theta \in \Theta$. Therefore, for these channels we have [29]

$$\mathcal{C}_t \equiv \sup_{P_{0,t}^z \in \mathcal{M}_1(\mathcal{Z}_{0,t})} \inf_{\theta\in\Theta} I^\theta(Z_0^t, \tilde{Z}_0^t) = \inf_{\theta\in\Theta} \sup_{P_{0,t}^z \in \mathcal{M}_1(\mathcal{Z}_{0,t})} I^\theta(Z_0^t, \tilde{Z}_0^t). \tag{2}$$

It is shown in [30] that the feedback capacity of the memoryless compound channel is given by

$$\mathcal{C} = \lim_{t\to\infty} \frac{1}{t+1}\mathcal{C}_t, \text{ where } \mathcal{C}_t = \inf_{\theta\in\Theta} \sup_{P_{0,t}^z \in \mathcal{M}_1(\mathcal{Z}_{0,t})} I^\theta(Z_0^t, \tilde{Z}_0^t). \tag{3}$$

Hence, it follows from equality (2) that feedback does not increase the capacity of the compound erasure channel. However, as shown in [23], it can help significantly in simplifying coding scheme. Note that for a compound erasure channel with the channel input alphabet $\mathbf{Z} = \{0,1\}^{\mathcal{R}}$ (where $\mathcal{R}$ is the length of transmitted packet in each channel use), output alphabet $\tilde{\mathbf{Z}} = \mathbf{Z} \cup \{e\}$, and unknown erasure probability $\theta \in \Theta$, it is verified that $\mathcal{C}_t = \inf_{\theta\in\Theta}(1-\theta)(t+1)\mathcal{R}$ and therefore, $\mathcal{C} = \inf_{\theta\in\Theta}(1-\theta)\mathcal{R}$.

Next, we recall the definition of robust rate distortion and then we establish a relationship between the Shannon capacity and the robust rate distortion for reliable data reconstruction. Let $X_t \in \mathbf{R}^q$ denote the source message with distribution P_t^x and $\hat{X}_t \in \mathbf{R}^q$ denote the corresponding reconstruction. Suppose that the source is uncertain in the sense that its probability measure is unknown; but the set $\mathcal{M}_{S,t} \subset \mathcal{M}_1(R^q)$ to which it belongs is known. That is, $P_t^x \in \mathcal{M}_{S,t} \subset \mathcal{M}_1(\mathbf{R}^q)$. Let $D \geq 0$ denote the distortion level and let

$$\mathcal{M}_D(P_t^x) \equiv \{Q_t : \int_{\mathbf{R}^q}\int_{\mathbf{R}^q} \rho(x-\hat{x})Q_t(d\hat{x}|x) \times P_t^x(dx) \leq D\}$$

represent the set of stochastic kernels satisfying the distortion constraint where $\rho(x-\hat{x})$ is the difference distortion measure. For example, the distortion measure ρ can be chosen either as r-moment measure (i.e., $\rho_r(x-\hat{x}) \equiv ||x-\hat{x}||^r$, $r > 0$) or indicator measure, i.e.,

$$\rho_\epsilon(x-\hat{x}) \equiv \begin{cases} 0 & \text{if} \quad ||x-\hat{x}|| \leq \epsilon \\ 1 & \text{if} \quad ||x-\hat{x}|| > \epsilon \end{cases} \quad (\epsilon > 0).$$

Then, the robust rate distortion for the uncertain source is defined as follows.

Definition 3.2 (The Robust Rate Distortion) *[31] Consider the information source as described above. The robust rate distortion corresponding to the family $\mathcal{M}_{S,t}$ is given by*

$$R_t(D) \equiv \inf_{Q_t\in\mathcal{M}_D(P_t^x)} \sup_{P_t^x\in\mathcal{M}_{S,t}} I(X_t,\hat{X}_t). \tag{4}$$

As shown in [31] when $\mathcal{M}_{S,t}$ is a compact set, $\inf\sup$ in (4) can be exchanged with $\sup\inf$. Furthermore, when the messages are produced according to an i.i.d. distribution, the rate distortion $R_t(D)$, as defined above, has an operational meaning and it represents the minimum rate for which uniform reliable data reconstruction up to the distortion level D holds.

For most of distortion measures and source distributions, finding an explicit analytical expression for the rate distortion as a function of t and D is difficult. Therefore, approximating the rate distortion function by a lower bound, which can be easily computed, is useful. In the following lemma we present a lower bound for the (robust) rate distortion in terms of the (robust) entropy of the source message. We use this lemma to present a necessary condition for UASAE.

Toward this goal, consider the uncertain information source, as described above, and suppose that P_t^x admits a density function p_t. Denote the Shannon (differential) entropy associated with the density function p_t by $H(p_t)$, which is given by [23]

$$H(p_t) \equiv -\int_{\mathbf{R}^q} p_t(x)\log(p_t(x))dx.$$

Also, let $R_{S,t}(D)$ be the corresponding Shannon lower bound given by [32]

$$R_{S,t}(D) \equiv H(p_t) - \max_{h \in G_D} H(h)$$

where

$$G_D \equiv \{h : \mathbf{R}^q \to [0,\infty) : \int_{\mathbf{R}^q} h(\xi)d\xi = 1, \int_{\mathbf{R}^q} \rho(\xi)h(\xi)d\xi \le D\}$$

and $H(h)$ is the Shannon entropy associated with the density h. Note that when

$$\int_{\mathbf{R}^q} e^{s\rho(\xi)}d\xi < \infty \text{ for all } s < 0,$$

the density $h^* \in G_D$ that maximizes $H(h)$ is given by

$$h^*(\xi) = \frac{e^{s^*\rho(\xi)}}{\int_{\mathbf{R}^q} e^{s^*\rho(\xi)}d\xi}$$

for some $s^* < 0$ satisfying $\int_{\mathbf{R}^q} \rho(\xi)h^*(\xi)d\xi = D$.
Then, a relationship between the robust rate distortion and the Shannon lower bound for the uncertain information source is given by the following result:

Lemma 3.3 (The Shannon Lower Bound) *Consider the information source, as described above. Suppose that P_t^x admits a density function p_t. Then*
(i) *for the case of single source (i.e., $\mathcal{M}_{S,t} = \{P_t^x\}$) a lower bound for the rate distortion function $R_t(D)$ is given by $R_t(D) \ge R_{S,t}(D) \equiv H(p_t) - \max_{h\in G_D} H(h)$.*
(ii) *for the case of uncertain source, a lower bound for the robust rate distortion is given by $R_t(D) \ge \sup_{P_t^x \in \mathcal{M}_{S,t}} H(p_t) - \max_{h\in G_D} H(h)$.*

Proof: See Appendix.

Consider the CS shown in Fig. 1, with the corresponding mathematical model as described earlier. At time t, the state of the system X_t is observed and the observation is encoded and transmitted via the channel to the receiver. Then, the receiver produces the state estimate $\hat{X}_t$. The encoder uses (via feedback) either the past channel outputs: $\tilde{Z}_0^{t-1}$ or $\bar{Z}_0^{t-1}$ and/or the past channel inputs Z_0^{t-1} to produce the current channel input Z_t. The decoder also uses all channel outputs $\tilde{Z}_0^t$ to produce $\hat{X}_t$. This system is subject to the conditional independence property. That is, $X_t \to Z_0^t \to \tilde{Z}_0^t \to \hat{X}_t$ forms a Markov chain. From now on we denote this communication system by the pair (B_ℓ, Θ) to emphasize on the uncertainty in the source and channel.

A necessary condition for uniform reliable data reconstruction capability of the communication system (B_ℓ, Θ), in the sense that $E[\rho(X_t - \hat{X}_t)] \leq D$, $\forall \theta \in \Theta$ and $\forall \Gamma_k \in B_\ell(M(q \times q))$ $(k \leq t-1)$, is given in the following theorem.

Theorem 3.4 *Consider the communication system (B_ℓ, Θ), as described above. For a given distortion level D, suppose that the limit, $\lim_{t\to\infty} \frac{1}{t+1} R_t(D)$, exists. Then, a necessary condition for the existence of an encoder and a decoder (estimator) for the uniform reliable data reconstruction is that the following inequality holds*

$$\mathcal{C} \geq \lim_{t\to\infty} \frac{1}{t+1} R_t(D). \tag{5}$$

Proof: See Appendix.

We have the following remarks regarding above result.

Remark 3.5 **(i)** *The necessary condition (5) is independent of the class of encoders and therefore it holds for all the encoders: Class A, Class B and Class C.*
(ii) *The difference distortion measure $\rho(\cdot)$, used for the expression of the rate distortion $R_t(D)$, is chosen according to the desired reliable data reconstruction capability. For example, for moment reconstructability (i.e., $E||X_t - \hat{X}_t||^r \leq D$) we choose $\rho(x - \hat{x}) = \rho_r(x - \hat{x})$.*

By combining Lemma 3.3 and Theorem 3.4 we have the following necessary condition for UASAE.

Theorem 3.6 *Consider the communication system (B_ℓ, Θ), as described above. Suppose that $\lim_{t\to\infty} \frac{1}{t+1} \sup_{P_t^x \in \mathcal{M}_{S,t}} H(p_t)$ exists. Then, a necessary condition for which UASAE holds is that the Shannon capacity $\mathcal{C}$ must satisfy the following inequality*

$$\mathcal{C} \geq \lim_{t\to\infty} \frac{1}{t+1} \sup_{P_t^x \in \mathcal{M}_{S,t}} H(p_t). \tag{6}$$

Proof: See Appendix.

Corollary 3.7 *Consider the communication system (B_ℓ, Θ). Suppose that the initial state X_0 of the information source (1) has finite entropy. Then, a necessary condition for which UASAE holds is that*

$$\mathcal{C} \geq \sum_{i=1}^{q} \max\{0, \log |\lambda_i(A+\Gamma^o)|\}, \tag{7}$$

where $\Gamma^o \equiv argmax\{|\det(A+\Gamma)| = \prod_{i=1}^{q} |\lambda_i(A+\Gamma)|, \Gamma \in B_\ell(M(q\times q))\}$.

Proof: See Appendix.

We have the following observation regarding the result of Corollary 3.7.

Remark 3.8 **(i)** *Since the set $B_\ell(M(q\times q))$ is compact and $\det(\cdot)$ is a continuous function, there always exists a $\Gamma^o \in B_\ell(M(q\times q))$ satisfying (7).*
(ii) *It is clear that the larger the ℓ is, the larger is the required capacity for UASAE to hold. In other words, the larger the system uncertainty is, the larger is the required capacity.*

For the purpose of illustration of the necessary condition (7), we present the following example.

Example 3.9 *Consider the CS shown in Fig. 1. Suppose that the dynamic system is the uncertain system (1) with $A = \begin{pmatrix} -3 & 1 \\ 0 & 2 \end{pmatrix}$ and $\Gamma_t = \begin{pmatrix} \delta_t & 0 \\ 0 & \gamma_t \end{pmatrix}$ where $||\Gamma_t|| \leq \ell = 1$ (recall that, for each $t \geq 0$, $||\Gamma_t|| \equiv \sqrt{\lambda_{max}(\Gamma_t'\Gamma_t)}$, that is, $\max\{|\delta_t|, |\gamma_t|\} \leq 1$, $\forall t \in \mathbf{N}_+$). The channel is the compound erasure channel in which at each time step it transmits $\mathcal{R}$ bits in each channel use. The erasure probability θ is unknown and belongs to the set $\Theta = [0.1, 0.5]$. Therefore, the capacity of this channel is $\mathcal{C} = 0.5\mathcal{R}$ (bits/time step). From Corollary 3.7 we have $\Gamma^0 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ and therefore the necessary condition (7) is given by $0.5\mathcal{R} \geq \log 12$. This means that for the rates $\mathcal{R} = 1, 2, 3, 4, 5, 6, 7$ we can not find any encoder and decoder for which UASAE holds.*

4 Sufficient Condition

In this section it is shown that the lower bound given by (7) can be also a sufficient condition for UASAE of the nominal version of the system (1) (i.e., $\ell = 0$) over the compound erasure channels. This is shown by developing a differential coding strategy, which uses feedback channel with duty cycle $\beta \in (0, 1]$. Recall that $\beta = 0$ implies non-availability of feedback channel and $\beta = 1$ implies its full time availability.

In this section, it is assumed that $\beta = \frac{M}{N}$ ($M, N \in \{1, 2, 3, ...\}$, $M \leq N$), where for each N updates of the state estimate, in the first M updates feedback channel is used in the encoder of Class C. Again, for the simplicity of presentation, it is assumed that the measurement matrix C in (1) is an identity matrix (if C is not an identity matrix but $C'C$ is invertible, then $\bar{Y}_t \equiv (C'C)^{-1}C'Y_t = X_t$ is treated as the observation signal). Moreover, it is assumed that the encoder and decoder are aware of each others policies and X_0 has a bounded support in $\mathbf{R}^q$.

The differential coding strategy is described in the proof of the following proposition which shows that UASAE holds for the system (1) over the compound erasure channel.

Proposition 4.1 *Consider the communication system (B_ℓ, Θ) whose capacity is given by $\mathcal{C} = \inf_{\theta\in\Theta}(1-\theta)\mathcal{R}$, where $\mathcal{R}$ is the average transmission bit rate. Suppose that the information source is given by the nominal version of the system (1) (i.e., $\ell = 0$) and the encoder is of the Class C. Let X_0 have a bounded support and $C^{'}C$ be invertible. Then, a sufficient condition for which UASAE holds is that the capacity $\mathcal{C}$ (measured in bits/time step) satisfies the following lower bound*

$$\mathcal{C} > \sum_{i=1}^{q} \max\{0, \log|\lambda_i(A)|\} = \lim_{t\to\infty} \frac{1}{t+1} H(p_t). \tag{8}$$

Proof: See Appendix.

We have the following observation regarding the above result.

Remark 4.2 **(i)** *The sufficient condition (8) implies that there exists a coding strategy for which UASAE holds for the nominal version of the system (1) over the compound erasure channel that uses feedback channel with duty cycle $\beta \in (0, 1]$. This strategy results in UASAE if the transmission rate $\mathcal{R}$ is greater than the rate $\frac{1}{\inf_{\theta\in\Theta}(1-\theta)} \sum_{i=1}^{q} \max\{0, \log|\lambda_i(A)|\}$.*
(ii) *From Corollary 3.7 it follows that the eigenvalues rate condition described by the Shannon capacity is tight (i.e., the necessary and sufficient condition) for which UASAE holds for the nominal version of the system (1) over the compound erasure channel. In other words, the eigenvalues rate: $\sum_{i=1}^{q} \max\{0, \log|\lambda_i(A)|\}$ is the minimum capacity required for UASAE.*
(iii) *In this paper we have been concerned with uncertainty in communication channel; and as we have shown above, in the presence of uncertainty in the channel, the eigenvalues rate condition described by the Shannon capacity is tight for almost sure estimation.*

When a communication channel has limited capacity, it is desirable to have estimation using the minimum possible capacity. As shown above, the minimum required capacity for almost sure estimation equals the eigenvalues rate: $\sum_{i=1}^{q} \max\{0, \log|\lambda_i(A)|\}$. In the problem considered in this paper, the information source (dynamic system) is given; and therefore, the eigenvalues rate is fixed and in many cases non-integer. But non-integer rate $\frac{M}{N}\frac{1}{\inf_{\theta\in\Theta}(1-\theta)}\sum_{i=1}^{q}\max\{0,\log|\lambda_i(A)|\}$ cannot be put on communication channel. Hence, all we can do is a proper modification of the proposed coding strategy to achieve almost sure estimation by the use of the minimum required capacity. This modification is described below.

Consider a small positive real number η and define $\mathcal{R}_{min}$ by the following expression

$$\mathcal{R}_{min} \equiv \frac{M}{N}\frac{1}{\inf_{\theta\in\Theta}(1-\theta)}\sum_{i=1}^{q}\max\{0,\log|\lambda_i(A)|\} + \frac{\eta}{\inf_{\theta\in\Theta}(1-\theta)}.$$

If $\mathcal{R}_{min}$ is an integer number this rate can be put into the channel by implementing the proposed differential coding strategy; and therefore, we have UASAE by transmission with the minimum required capacity. But, in general, $\mathcal{R}_{min}$ may not be an integer number. For this case, we use the following time-sharing strategy to achieve UASAE by the use of the minimum required capacity.

Time-Sharing Strategy: For simplicity consider the scalar case first and without loss of generality suppose $M = N = 1$. Let $i \in \mathbf{N}_+$ be the smallest integer such that $i \leq \mathcal{R}_{min} < 1 + i$. Also, let $\xi \equiv \mathcal{R}_{min} - i$. Suppose that both rates i and $1 + i$ can be put into the channel. Then, unlike the proposed differential coding strategy, the encoder now partitions the box $[-L_t, L_t]$, where the estimation error lives in, into $2^{\mathcal{R}_t}$ equal size non-overlapping intervals, where $\mathcal{R}_t$ takes values from the set $\mathcal{R}_t \in \{i, 1+i\}$ according to the following time-sharing strategy: $\begin{cases} P(\mathcal{R}_t = i) = 1-\xi, \\ P(\mathcal{R}_t = 1+i) = \xi. \end{cases}$

Subsequently, by the strong law of large numbers, the average transmitted rate denoted here by $\mathcal{R}^{av} \equiv \lim_{t\to\infty}\frac{1}{t+1}\sum_{j=0}^{t}\mathcal{R}_j$ equals $E[\mathcal{R}_0]$ almost surely, where

$$E[\mathcal{R}_0] = (1-\xi)i + \xi(1+i) = i - \xi i + \xi + \xi i = i + \mathcal{R}_{min} - i = \mathcal{R}_{min}.$$

In other words, here we are transmitting with the minimum required capacity

$$\mathcal{C} = \inf_{\theta\in\Theta}(1-\theta)\mathcal{R}^{av} = \inf_{\theta\in\Theta}(1-\theta)\mathcal{R}_{min} = \sum_{i=1}^{q}\max\{0,\log|\lambda_i(A)|\} + \eta.$$

Now we must also show that for this time-sharing strategy UASAE holds. By implementing the proposed coding strategy and noting that here the transmission rate is specified by the above time-sharing law, we have $|X_t - \hat{X}_t| \leq V_t$, where here the random variables F_t are i.i.d. with common distribution given by: $P(F_t = 1) = \theta$, $P(F_t = 2^{-i}) = (1-\xi)(1-\theta)$,

$P(F_t = 2^{-(1+i)}) = \xi(1-\theta)$. Subsequently, along the same lines of the proof of Proposition 4.1, it is shown that using the above time-sharing strategy $V_t \to 0$; and therefore, UASAE holds. Note that when the erasure channel transmits a packet of data, $\mathcal{R}_t$, successfully, from the number of received bits, the decoder can identify whether $\mathcal{R}_t = i$ or $\mathcal{R}_t = 1+i$ has been selected by the encoder as the communication channel is assumed to be ternary with input and output 0, 1, idle. This type of channel can be obtained, for example, by implementing an Amplitude Shift Keying (ASK) type modulation scheme, in which when the amplitude of the sampled received signal is above V ($V > 0$), it is decoded as 1, when it is less than $-V$, it is decoded as 0, and otherwise the channel is assumed to be in the idle mode. Consequently, at each time instant by counting the number of the received 0 and 1 bits the length of the received packet can be identified. Hence, the decoder can identify which of them (i or $1+i$) was chosen for transmission. Another way for the decoder to know the length of transmitted bits is to choose i or $1+i$ deterministically instead of randomly.

Extension of the above strategy to the vector case is straightforward. It is realized by implementing a similarity transformation that turns the system matrix A to the real Jordan form, and noting that now the transmission rate is defined following a time-sharing strategy, as specified below:
Let $\lambda_k(A)$ be the eigenvalue of the system matrix A corresponding to the kth ($k = 1, 2, ..., m$, $m \le q$) Jordan block of the matrix $A \in M(q \times q)$. For each $\lambda_k(A)$, let i_k be the smallest integer such that

$$i_k \le \mathcal{R}^{(k)}_{min} \equiv \max\{0, \frac{1}{\inf_{\theta\in\Theta}(1-\theta)} \log|\lambda_k(A)|\} + \frac{\eta}{\inf_{\theta\in\Theta}(1-\theta)} < 1 + i_k.$$

Also, let $\xi_k = \mathcal{R}^{(k)}_{min} - i_k$. Then, for each $\lambda_k(A)$ the encoder partitions the box associated with $\lambda_k(A)$ into $2^{\mathcal{R}^{(k)}_t}$ equal size non-overlapping intervals, where $\mathcal{R}^{(k)}_t$ is chosen according to the following i.i.d. distribution: $\begin{cases} P(\mathcal{R}^{(k)}_t = i_k) = 1 - \xi_k, \\ P(\mathcal{R}^{(k)}_t = 1 + i_k) = \xi_k. \end{cases}$

Simulation Result: The results shown in Fig. 2 illustrate the performance of the proposed time-sharing strategy when the communication is via the compound erasure channel, which uses feedback channel all the time (i.e., $\beta = 1$). Here, it is assumed that the unknown erasure probability θ belongs to the set $\Theta = [0.1, 0.5]$. The dynamic system (1) is assumed to be scalar with $A = -3$ and $\ell = 0$. The initial condition X_0 is uniformly distributed in the interval $[-1, 1]$, i.e., $L_0 = 1$. From Corollary 3.7 and Proposition 4.1 it follows that $\frac{1}{\inf_{\theta\in\Theta}(1-\theta)} \log|A| = \frac{1}{0.5}\log 3 = 3.1699$ (bits/time step) is the minimum transmission rate for which UASAE holds. But this rate is not an integer number. Therefore, to have almost sure estimation requiring the minimum capacity $\mathcal{C} = \log|A| = 1.58$ (bits/time step) for the

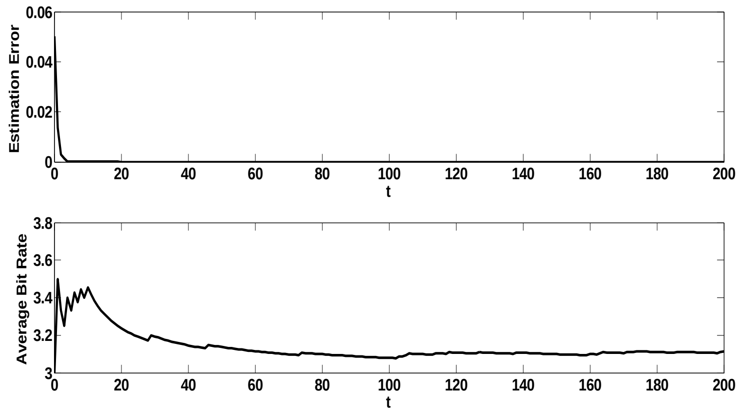

Figure 2: Simulation result for $\beta = 1$, $A = -3$, $\ell = 0$, $L_0 = 1$, $\Theta = [0.1, 0.5]$ and $\eta = 0.0005$. Top figure: estimation error $|X_t - \hat{X}_t|$, bottom figure: the average bit rate.

case of $\beta = 1$ we need to implement the proposed time-sharing strategy. Fig. 2 illustrates the performance of the proposed strategy for $A = -3$, $\ell = 0$, $L_0 = 1$, $\Theta = [0.1, 0.5]$ and $\eta = 0.0005$. It is clear from Fig. 2 that, although the erasure probability θ is unknown, after a few iterations the absolute value of the estimation error (i.e., $|X_t - \hat{X}_t|$) converges to zero; while the average transmission rate $\mathcal{R}^{av} = \lim_{t\to\infty} \frac{1}{t+1} \sum_{j=0}^{t} \mathcal{R}_j$ equals the minimum required transmission rate $\frac{1}{\inf_{\theta\in\Theta}(1-\theta)} \log|A|$. That is, here the capacity $\mathcal{C} \equiv \inf_{\theta\in[0.1,0.5]}(1-\theta)\mathcal{R}^{av}$ is going to equal the minimum required capacity 1.58 (bits/time step).

In addition, it is observed that if the above technique is applied to the case where $\theta \in \bar{\Theta} \subset \Theta$, the estimation error is asymptotically zero. If $\bar{\Theta}_1$, $\bar{\Theta}_2 \subset \Theta$, where each point in $\bar{\Theta}_2$ is greater than all the points in $\bar{\Theta}_1$, the performance of the case of $\theta \in \bar{\Theta}_1 \subset \Theta$ is better than the performance of the case of $\theta \in \bar{\Theta}_2 \subset \Theta$, as expected. And if $\theta \in \widetilde{\Theta}$, where each point in $\widetilde{\Theta}$ is greater than all points in Θ, the performance is poor and in some cases the estimation error may become unbounded.

Fig. 3 illustrates the performance of the proposed coding and time-sharing strategy when the erasure probability θ is known and is equivalent to $\theta = 0.1$ and $\beta = 1$. As is clear from Fig. 3 the proposed strategy is able to estimate the state of the system with asymptotically zero estimation error by transmission with the minimum rate of 1.7611 bits/time step. However, the performance for this case is not as good as the performance of the other case (Fig. 2). This is due to the fact that in Fig. 2 by taking a conservative approach more bits are transmitted which results in a better performance.

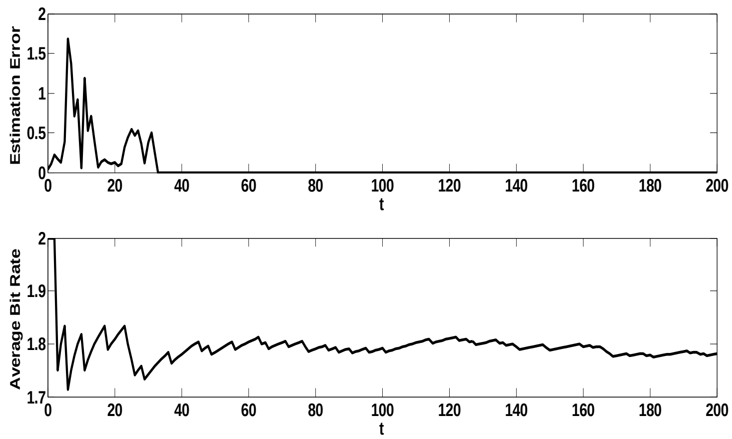

Figure 3: Simulation result for $\beta = 1$, $A = -3$, $\ell = 0$, $L_0 = 1$, $\theta = 0.1$ and $\eta = 0.0005$. Top figure: estimation error $|X_t - \hat{X}_t|$, bottom figure: the average bit rate.

5 Conclusion

This paper extended the previous results to cover the cases where both the dynamic system and communication channel are uncertain, and the feedback channel has a duty cycle $\beta \in (0, 1]$. A necessary condition for which UASAE holds was derived under this general situation. Moreover, when the dynamic system is not uncertain; but the channel is, a sufficient condition for which UASAE holds over the compound erasure channels, was presented. Consequently, it was concluded that over the compound erasure channel, a condition on the Shannon capacity in terms of the rate of expansion of the Shannon entropy is a necessary and sufficient condition for uniform almost sure asymptotic zero estimation error. Possible extension for future includes addressing the estimation problem of stochastic nonlinear uncertain dynamic systems over noisy communication channels, which use feedback links with a Markov chain model.

6 Appendix

Proof of Lemma 3.3. (i) The first part is well known and for the detailed proof see [32]. **(ii)** From the classical Shannon lower bound, as given in part i, it follows that for each

$P_t^x \in \mathcal{M}_{S,t}$, we have $\inf_{Q_t \in \mathcal{M}_D(P_t^x)} I(X_t, \hat{X}_t) \geq H(p_t) - \max_{h \in G_D} H(h)$. Consequently, the robust rate distortion $R_t(D)$ satisfies the following inequalities:

$$\begin{aligned} R_t(D) \equiv \inf_{Q_t \in \mathcal{M}_D(P_t^x)} \sup_{P_t^x \in \mathcal{M}_{S,t}} I(X_t, \hat{X}_t) \quad &\geq \sup_{P_t^x \in \mathcal{M}_{S,t}} \inf_{Q_t \in \mathcal{M}_D(P_t^x)} I(X_t, \hat{X}_t) \\ &\geq \sup_{P_t^x \in \mathcal{M}_{S,t}} \Big(H(p_t) - \max_{h \in G_D} H(h)\Big). \end{aligned}$$

Now, as G_D is independent of P_t^x, from above expression, we have

$$R_t(D) \geq \sup_{P_t^x \in \mathcal{M}_{S,t}} H(p_t) - \max_{h \in G_D} H(h).$$

This completes the proof.

Proof of Theorem 3.4. Let $\{\bar{P}_{0,t}^z\}$ denote the set of all distributions corresponding to the channel input sequence Z_0^t when the source message X_t with distribution $P_t^x \in \mathcal{M}_{S,t}$ is transmitted. Suppose that there exists an encoder-decoder pair such that we have the uniform reliable data reconstruction. Then, it follows from the data processing inequality [23], as described by $I^\theta(Z_0^t, \tilde{Z}_0^t) \geq I^\theta(X_t, \hat{X}_t)$ ($\forall P_t^x \in \mathcal{M}_{S,t}$ and $\forall \theta \in \Theta$) that, for a given $\theta \in \Theta$, we have

$$\sup_{\{\bar{P}_{0,t}^z\}} I^\theta(Z_0^t, \tilde{Z}_0^t) \;\geq\; \sup_{P_t^x \in \mathcal{M}_{S,t}} I^\theta(X_t, \hat{X}_t). \tag{9}$$

Let $\{P_{0,t}^z\}$ denote the set of all channel input distributions. Evidently $\{\bar{P}_{0,t}^z\} \subseteq \{P_{0,t}^z\}$ and hence $\sup_{P_{0,t}^z \in \mathcal{M}_1(\mathcal{Z}_{0,t})} I^\theta(Z_0^t, \tilde{Z}_0^t) \geq \sup_{\{\bar{P}_{0,t}^z\}} I^\theta(Z_0^t, \tilde{Z}_0^t)$. Therefore, for a given $\theta \in \Theta$, it follows from (9) that

$$\sup_{P_{0,t}^z \in \mathcal{M}_1(\mathcal{Z}_{0,t})} I^\theta(Z_0^t, \tilde{Z}_0^t) \geq \inf_{Q_t^\theta \in \mathcal{M}_D(P_t^x)} \sup_{P_t^x \in \mathcal{M}_{S,t}} I^\theta(X_t, \hat{X}_t), \tag{10}$$

where Q_t^θ is the stochastic kernel corresponding to X_t and $\hat{X}_t$ given $\theta \in \Theta$. By definition

$$I^\theta(X_t, \hat{X}_t) \equiv \int\int \log\Big(\frac{Q_t^\theta(d\hat{x}|x)}{\int Q_t^\theta(d\hat{x}|x) P_t^x(dx)}\Big) Q_t^\theta(d\hat{x}|x) P_t^x(dx).$$

Evidently infimum of the term $\sup_{P_t^x \in \mathcal{M}_{S,t}} I^\theta(X_t, \hat{X}_t)$ with respect to $Q_t^\theta \in \mathcal{M}_D(P_t^x)$ is independent of Q_t^θ. Hence,

$$\inf_{Q_t^\theta \in \mathcal{M}_D(P_t^x)} \sup_{P_t^x \in \mathcal{M}_{S,t}} I^\theta(X_t, \hat{X}_t)$$

is independent of θ. That is,

$$\inf_{Q_t^\theta \in \mathcal{M}_D(P_t^x)} \sup_{P_t^x \in \mathcal{M}_{S,t}} I^\theta(X_t, \hat{X}_t) = R_t(D). \tag{11}$$

Therefore, by taking infimum with respect to θ over the set Θ, it follows from (10) and (11) that the following inequality holds:

$$\inf_{\theta\in\Theta} \sup_{P_{0,t}^z \in \mathcal{M}_1(\mathcal{Z}_{0,t})} I^\theta(Z_0^t, \tilde{Z}_0^t) \geq R_t(D).$$

Hence, it follows from the Definition 3.1 and equality (2) that

$$\mathcal{C}_t \equiv \sup_{P_{0,t}^z \in \mathcal{M}_1(\mathcal{Z}_{0,t})} \inf_{\theta\in\Theta} I^\theta(Z_0^t, \tilde{Z}_0^t) = \inf_{\theta\in\Theta} \sup_{P_{0,t}^z \in \mathcal{M}_1(\mathcal{Z}_{0,t})} I^\theta(Z_0^t, \tilde{Z}_0^t) \geq R_t(D).$$

Therefore, under the assumption of the existence of an encoder-decoder pair that yields an average distortion $E[\rho(X_t - \hat{X}_t)] \leq D$, for all $\Gamma_k \in B_\ell(M(q\times q))$ $(k \leq t-1)$ and $\forall\theta\in\Theta$, we have $\mathcal{C}_t \geq R_t(D)$, and therefore

$$\mathcal{C} \equiv \lim_{t\to\infty} \frac{1}{t+1}\mathcal{C}_t \geq \lim_{t\to\infty} \frac{1}{t+1} R_t(D).$$

That is, $\mathcal{C} \geq \lim_{t\to\infty} \frac{1}{t+1} R_t(D)$ is a necessary condition for the existence of an encoder-decoder pair. This completes the proof.

Proof of Theorem 3.6. Suppose that UASAE holds. This implies that for any $\epsilon > 0$ there exists a $T(\epsilon) < \infty$ such that the following inequality holds:

$$P(\sup_{t\geq T(\epsilon)} ||X_t - \hat{X}_t|| > \epsilon) \leq \epsilon,\ \forall\Gamma_k \in B_\ell(M(q\times q))(k \leq t-1),\ \forall\theta\in\Theta.$$

Now, if we choose $\rho(\cdot)$ as the indicator measure, i.e., $\rho(\xi) = \rho_\epsilon(\xi) = \begin{cases} 0 & \text{if} \quad ||\xi|| \leq \epsilon \\ 1 & \text{if} \quad ||\xi|| > \epsilon \end{cases}$, for $t \geq T(\epsilon)$ we have $E[\rho(X_t - \hat{X}_t)] = P(||X_t - \hat{X}_t|| > \epsilon) \leq \epsilon$, uniformly with respect to $B_\ell(M(q\times q))$ and Θ. Therefore, from Theorem 3.4 and Lemma 3.3, the capacity and robust rate distortion must, for all $t \geq T(\epsilon)$, satisfy the following inequalities:

$$\frac{1}{t+1}\mathcal{C}_t \geq \frac{1}{t+1}R_t(\epsilon) \geq \frac{1}{t+1} \sup_{P_t^x \in \mathcal{M}_{S,t}} H(p_t) - \frac{1}{t+1} \max_{h\in G_\epsilon} H(h).$$

It is known that for the indicator distortion measure, $\max_{h\in G_\epsilon} H(h)$ is finite [34]. Hence, it follows from above expression that $\mathcal{C} \equiv \lim_{t\to\infty} \frac{1}{t+1}\mathcal{C}_t \geq \lim_{t\to\infty} \frac{1}{t+1} \sup_{P_t^x \in \mathcal{M}_{S,t}} H(p_t)$. This proves that the inequality (6) is a necessary condition for UASAE.

Proof of Corollary 3.7. From Theorem 3.6 we have the following inequality as a necessary condition for UASAE: $\mathcal{C} \geq \lim_{t\to\infty} \frac{1}{t+1} \sup_{P_t^x \in \mathcal{M}_{S,t}} H(p_t)$, where the density function p_t is induced by the R.V. $X_t = \Big(\prod_{t-1\geq k\geq 0}(A+\Gamma_k)\Big)X_0$. Hence, it follows from ([23], p. 234) that $H(p_t) = \log\Big|\det\Big(\prod_{t-1\geq k\geq 0}(A+\Gamma_k)\Big)\Big| + H(p_0)$. Therefore,

$$\sup_{P_t^x \in \mathcal{M}_{S,t}} H(p_t) = H(p_0) + \sup_{\{||\Gamma_k||\leq l, 0\leq k\leq t-1\}} \Big(\log\Big|\det\Big(\prod_{t-1\geq k\geq 0}(A+\Gamma_k)\Big)\Big|\Big).$$

Consequently

$$\lim_{t\to\infty}\frac{1}{t+1}\sup_{P_t^x\in\mathcal{M}_{S,t}} H(p_t)=\lim_{t\to\infty}\frac{1}{t+1}\sup_{\{||\Gamma_k||\leq l, 0\leq k\leq t-1\}}\Big(\log\Big|\det\Big(\prod_{t-1\geq k\geq 0}(A+\Gamma_k)\Big)\Big|\Big). \tag{12}$$

We can always find a similarity transformation T such that the matrix $A+\Gamma^o$ can be written in the following form

$$T^{-1}(A+\Gamma^o)T=\begin{pmatrix}(A+\Gamma^o)_s & \\ & (A+\Gamma^o)_{us}\end{pmatrix}, \tag{13}$$

where $(A+\Gamma^o)_s$ has eigenvalues inside the unit circle and $(A+\Gamma^o)_{us}$ has eigenvalues on or outside the unit circle. Accordingly, we split the state space into the following two disjoint subspaces: i) stable subspace which corresponds to $(A+\Gamma^o)_s$; and ii) unstable subspace which corresponds to $(A+\Gamma^o)_{us}$. Let $\mathcal{P}_S$ be the projection onto the stable subspace. Then, $\lim_{t\to\infty}\mathcal{P}_S(X_t)=0$. Hence, for sufficiently large t, the projection of the state onto the stable subspace is reconstructed as zero. That is, for large t the stable subspace does not contribute to the entropy of the R.V. X_t. Therefore, without loss of generality, in computing the entropy of the R.V. X_t we can restrict our attention to the matrix $A+\Gamma^o$ that contains only unstable eigenvalues.

In view of this fact, it follows from (12) that

$$\lim_{t\to\infty}\frac{1}{t+1}\sup_{P_t^x\in\mathcal{M}_{S,t}} H(p_t) = \log|\det(A+\Gamma^o)|_{us}=\log\prod_{i=1}^{q}|\lambda_i(A+\Gamma^o)_{us}|=\sum_{i=1}^{q}\log|\lambda_i(A+\Gamma^o)_{us}|.$$

Therefore, for the matrix $A+\Gamma^o$ (with some stable eigenvalues) we have the following necessary condition for UASAE

$$\mathcal{C}\geq\sum_{i=1}^{q}\max\{0,\log|\lambda_i(A+\Gamma^o)|\}.$$

This completes the proof.

Proof of Proposition 4.1. In what follows we consider the scalar system. Extension of the results to the general vector case is straightforward and it follows by implementing a similarity transformation that turns the system matrix A to the real Jordan form [18].

As the initial state is bounded, we have $|X_0|\leq L_0$, where L_0 is known a priori. At time instant $t=0$ the encoder partitions the interval $[-L_0,L_0]$ into $2^{\frac{N}{M}\mathcal{R}}$ equal size bins, and upon observing X_0 it identifies the bin, where X_0 is located and represents the corresponding index by $\frac{N}{M}\mathcal{R}$ bits and transmits the corresponding packet. Then, the output of the decoder is updated by (14)

$$\hat{X}_t=\begin{cases}\gamma_j+\hat{X}_t^e & \text{if erasure does not occur}\\ \hat{X}_t^e & \text{if erasure occurs,}\end{cases} \tag{14}$$

where γ_j is the center of the $j+1$ bin, which contains $X_t - \hat{X}_t^e$. Note that $\hat{X}_0^e = 0$ and the encoder and decoder are aware of each others policies when the feedback channel is available; and hence, the decoder can determine $\hat{X}_t^e$ when feedback channel is available. Consequently, for the time instant $t=0$, the decoding error is bounded above by

$$|X_0 - \hat{X}_0| \leq V_0,$$

where

$$V_0 = \begin{cases} \frac{L_0}{2^{\frac{N}{M}\mathcal{R}}} & \text{if erasure does not occur} \\ L_0 & \text{if erasure occurs.} \end{cases}$$

At time instant $t=1$, using feedback channel, the encoder can determine V_0 and $\hat{X}_0$. Subsequently, it computes $\hat{X}_1^e = A\hat{X}_0$ and $L_1 = |A|V_0$ (note that during the time period between two time instants $t=0$ and $t=1$, feedback channel is used). Then, it partitions the interval $[-L_1, L_1]$ into $2^{\frac{N}{M}\mathcal{R}}$ bins. Upon observing X_1, the encoder computes $X_1 - \hat{X}_1^e$ and determines the bin, where $X_1 - \hat{X}_1^e$ is located. Then, it represents the index of this bin by $\frac{N}{M}\mathcal{R}$ bits and transmits the corresponding packet. Subsequently, the decoder output is updated by (14). For this case the decoding error (if the feedback channel is available) is bounded above by

$$|X_1 - \hat{X}_1| \leq V_1,$$

where

$$V_1 = \begin{cases} \frac{L_1}{2^{\frac{N}{M}\mathcal{R}}} & \text{if erasure does not occur} \\ L_1 & \text{if erasure occurs.} \end{cases}$$

Then, by following this procedure, we have:
at time instant $t \in \{1, 2, ..., M-1\}$, where the feedback channel is available, $L_t = |A|V_{t-1}$ and

$$|X_t - \hat{X}_t| \leq V_t,$$

where

$$V_t = \begin{cases} \frac{L_t}{2^{\frac{N}{M}\mathcal{R}}} & \text{if erasure does not occur} \\ L_t & \text{if erasure occurs.} \end{cases}$$

at time instant $t=M$, as the feedback channel is not available noting is sent to the decoder; and hence, $\hat{X}_M = \hat{X}_M^e = A\hat{X}_{M-1}$, $V_M = L_M$, $L_M = |A|V_{M-1}$. Similarly, at time instant $t \in \{M+1, ..., N-1\}$, $\hat{X}_t = \hat{X}_t^e = A^{(t-M+1)}\hat{X}_{M-1}$, $V_t = L_t$, $L_t = |A^{(t-M+1)}|V_{M-1}$.
Consequently, in general, at time instant t:

$$|X_t - \hat{X}_t| \leq V_t,$$

where

- for $t = 0$

$$V_0 = F_0 L_0;$$

- for $t = N, 2N, 3N, ...$, where feedback channel is available up to the next $M-1$ time instants, we have

$$V_t = F_t |A^{N-M+1}| V_{t-N+M-1},\ t = Nj,\ j \in \{1,2,3,...\}$$

where the sequence F_t ($t \in \mathbf{N}_+$) is i.i.d. with the following common distribution

$$F_t = \begin{cases} \frac{1}{2^{\frac{N}{M}\mathcal{R}}} & P(F_t = \frac{1}{2^{\frac{N}{M}\mathcal{R}}}) = 1-\theta \\ 1 & P(F_t = 1) = \theta \end{cases}$$

- for $t \in \{Nj+1, ..., Nj+M-1\}$, $j \in \mathbf{N}_+$, where the feedback channel is available, we have:

$$V_t = F_t |A| V_{t-1},$$

and

- for $t \in \{Nj+M, ..., Nj+N-1\}$, where the feedback channel is not available, we have

$$V_t = |A^{t-Nj-M+1}| V_{Nj+M-1}.$$

Consequently,

$$\begin{aligned} V_{Nj} &= F_{Nj} F_{N(j-1)+M-1} ... F_{N(j-1)+1} |A^N| ... F_{2N} F_{N+M-1} ... F_{N+1} |A^N| F_N F_{M-1} ... F_1 |A^N| V_0 \\ &= F_{Nj}|A|F_{N(j-1)+M-1}|A| ... F_{N(j-1)+1}|A||A^{N-M}| ... F_{2N}|A|F_{N+M-1}|A| ... F_{N+1}|A||A^{N-M}| \\ &\quad \times F_N|A|F_{M-1}|A| ... F_1|A||A^{N-M}|V_0 \\ &= 2^{Mj(\frac{1}{Mj}(\log|A|F_{Nj} + \log|A|F_{N(j-1)+M-1} + ... + \log|A|F_{N(j-1)+1} + ... + \log|A|F_N + \log|A|F_{M-1} + ... + \log|A|F_1))} \\ &\quad \times |A|^{j(N-M)} V_0. \end{aligned}$$

Now, from the strong low of large numbers [33] we have the following equality, almost surely:

$$\begin{aligned} &\lim_{j\to\infty} \frac{1}{Mj}(\log|A|F_{Nj} + \log|A|F_{N(j-1)+M-1} + ... + \log|A|F_{N(j-1)+1} + ... \\ &+ \log|A|F_N + \log|A|F_{M-1} + ... + \log|A|F_1) = E[\log(|A|F_1)] = (1-\theta)\log\frac{|A|}{2^{\frac{N}{M}\mathcal{R}}} + \theta\log|A|. \end{aligned}$$

Consequently, as $j \to \infty$, we have

$$V_{Nj} \to 2^{Mj((1-\theta)\log \frac{|A|}{2^{\frac{N}{M}\mathcal{R}}}+\theta \log |A|)} |A|^{j(N-M)} V_0 = (2^{M((1-\theta)\log \frac{|A|}{2^{\frac{N}{M}\mathcal{R}}}+\theta \log |A|)} |A|^{N-M})^j V_0.$$

But, as we have assumed $\inf_{\theta \in \Theta}(1-\theta)\mathcal{R} > \max\{0, \log |A|\}$, for each $\theta \in \Theta$, we have $2^{M((1-\theta)\log \frac{|A|}{2^{\frac{N}{M}\mathcal{R}}}+\theta \log |A|)} |A|^{N-M} < 1$; and hence, V_{Nj} along with the sequence V_{Nj+1}, ..., V_{Nj+N-1} converge to zero, almost surely, as $j \to \infty$. This completes the proof as $|X_t - \hat{X}_t| \le V_t$.

For the vector case, the encoder encodes each element of vector $X_t - \hat{X}_t$ into $\mathcal{R}_i$, $i = \{1, 2, ..., q\}$ bits and transmits a packet with length $\frac{N}{M}\sum_{i=1}^{q} \mathcal{R}_i$ over the compound erasure channel when the feedback channel is available.

References

[1] L. Brinon Arranz, A. Seuret, and C. Canudas de Wit, "Translation control of a fleet circular formation of AUVs under finite communication range," in *Proc. 48th IEEE Conference on Decision and Control*, 2009, pp. 8345-8350.

[2] L. Brinon Arranz, A. Seuret, and C. Canudas de Wit, "Collaborative estimation of gradient direction by a formation of AUVs under communication constraints," in *Proceedings of the 50th IEEE Conference on Decision and Control*, 2011, pp. 5583-5588.

[3] N. Elia,"When bode meets Shannon: control-oriented feedback communication schemes,", *IEEE Trans. Automat. Contr.*, 49(9), pp. 1477-1488, 2004.

[4] N. Elia and J. N. Eisenbeis,"Limitations of linear control over packet drop networks," *IEEE Trans. Automat. Contr.*, 56(4), pp. 826-841, 2011.

[5] N. C. Martins, A. Dahleh, and N. Elia,"Feedback stabilization of uncertain systems in the presence of a direct link," *IEEE Trans. Automat. Contr.*, 51(3), pp. 438-447, 2006.

[6] P. Minero, M. Franceschetti, S. Dey, and N. Nair,"Data rate theorem for stabilization over time-varying feedback channels," *IEEE Trans. Automat. Contr.*, 54(2), pp. 243-255, Feb. 2009.

[7] P. Minero, L. Coviello, and M. Franceschetti,"Stabilization over Markov feedback channels: the general case," *IEEE Trans. Automat. Contr.*, 58(2), pp. 349-362, 2013.

[8] G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran,"Topological feedback entropy and nonlinear stabilization," *IEEE Trans. Automat. Contr.*, 49(9), pp. 1585-1597, 2004.

[9] G. N. Nair and R. J. Evans,“Stabilizability of stochastic linear systems with finite feedback data rates,” *SIAM J. Control Optimization*, 43(3), pp. 413-436, 2004.

[10] C. Canudas de Wit, F. Gomez-Estern, and F. Rodrigues Rubio,“Delta-modulation coding redesign for feedback-controlled systems,” *IEEE Transactions on Industrial Electronics*, 56(7), pp. 2684-2696, 2009.

[11] A. Farhadi and N. U. Ahmed,“Tracking nonlinear noisy dynamic systems over noisy communication channels,” *IEEE Transactions on Communications*, 59(4), pp. 955-961, 2011.

[12] C. D. Charalambous and A. Farhadi, “A mathematical framework for robust control over uncertain communication channels,” in Proceedings of the 44th IEEE Conference on Decision and Control, Seville, Spain, pp. 2530-2535, December 2005.

[13] C. D. Charalambous and A. Farhadi, A, “Control of feedback systems subject to the finite rate constraints via Shannon lower bound,” in Proceedings of the 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Cyprus, pp. 1-7, April 2007.

[14] A. Farhadi, J. Domun, and C. Canudas de Wit, “A supervisory control policy over an acoustic communication network,” International Journal of Control, 88(5), pp. 946-958, May 2015.

[15] A. Farhadi, “Feedback channel in linear noiseless dynamic systems controlled over the packet erasure network,” International Journal of Control, 88(8), pp. 1490-1503, August 2015.

[16] A. Farhadi, “Stability of linear dynamic systems over the packet erasure channel: a co-design approach,” International Journal of Control, 88(12), pp. 2488-2498, December 2015.

[17] A. Farhadi, “Stability of nonlinear uncertain Lipschitz systems over the digital noiseless channel,” Scientia Iranica, Transactions on Computer Science and Engineering and Electrical Engineering, (accepted in October 2016).

[18] S. Tatikonda and S. Mitter, “Control over noisy channels,” *IEEE Transactions on Automatic Control*, 49(7), pp. 1196-1201, 2004.

[19] X. W. Jiang, Z. H. Guan, G. Feng, Y. Wu and F. S. Yuan,"Optimal tracking performance of networked control systems with channel input power constraint," IET Control Theory and Applications, 6(11), pp. 1690-1698, 2012.

[20] X. S. Zhan, J. Wu, T. Jiang and X. W. Jiang,"Optimal performance of networked control systems under the packet dropouts and channel noise," ISA Transactions, 58, pp. 214-221, 2015.

[21] X. W. Jiang, B. Hu, Z. H. Guan, X. H. Zhang and L. Yu,"Best achievable tracking performance for networked control systems with encoderdecoder," Information Sciences, 305, pp. 184-195, 2015.

[22] 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, pp. 241-255, February 2007.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

[24] J. C. Geromel, "Optimal linear filtering under parameter uncertainty," *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 168-175, January 1999.

[25] S. Xu, P. V. Dooren, R. Stefan, and J. Lam, "Robust stability and stabilization for singular systems with state delay and parameter uncertainty," *IEEE Transactions on Automatic Control*, vol. 47, no. 7, pp. 1122-1128, July 2002.

[26] H. Gao, X. Meng, and T. Chen, "A parameter-dependent approach to robust H_∞ filtering for time -delay systems," *IEEE Transactions on Automatic Control*, vol. 53, no. 10, pp. 2420-2425, November 2008.

[27] L. El Ghaoui and G. Calafiore, "Robust filtering for discrete - time systems with bounded noise and parametric uncertainty," *IEEE Transactions on Automatic Control*, vol. 46, no. 7, pp. 1084-1089, July 2001.

[28] J. Wolfowitz, "Simultaneous channels," *Archive for Rational Mechanics and Analysis*, vol. 4, pp. 371-386, 1959.

[29] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2148-2177, October 1998.

[30] B. Shrader and H. Permuter, “Feedback capacity of the compound channel,” *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3629-3644, August 2009.

[31] D. J. Sakrison, “The rate distortion function for a class of sources,” *Information and Control*, vol. 15, pp. 165-195, 1969.

[32] T. Linder and R. Zamir, “ On the asymptotic tightness of the Shannon lower bound,” *IEEE Transactions on Information Theory*, vol. 40, no. 6, November 1994.

[33] P. Billingsley, *Probability and Measure*, John Wiley and Sons, 1995.

[34] 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*, 44(12), pp. 3181-3188, December 2008.