Robust coding for a class of sources: Applications in control and reliable communication over limited capacity channels

Alireza Farhadi [a],[*], Charalambos D. Charalambous [b]

[a] *School of Information Technology and Engineering, University of Ottawa, Ontario, Canada*
[b] *Electrical and Computer Engineering Department, University of Cyprus, 75 Kallipoleos Avenue, Nicosia, Cyprus*

ARTICLE INFO

ABSTRACT

This paper is concerned with the control of a class of dynamical systems over finite capacity communication channels. Necessary conditions for reliable data reconstruction and stability of a class of dynamical systems are derived. The methodology is information theoretic. It introduces the notion of entropy for a class of sources, which is defined as a maximization of the Shannon entropy over a class of sources. It also introduces the Shannon information transmission theorem for a class of sources, which states that channel capacity should be greater or equal to the mini-max rate distortion (maximization is over the class of sources) for reliable communication. When the class of sources is described by a relative entropy constraint between a class of source densities, and a given nominal source density, the explicit solution to the maximum entropy, is given, and its connection to Rényi entropy is illustrated. Furthermore, this solution is applied to a class of controlled dynamical systems to address necessary conditions for reliable data reconstruction and stability of such systems.

1. Introduction

Over the last few years there has been an extensive research activity in addressing analysis and design questions associated with control of deterministic and stochastic systems over communication channels with limited channel capacity. This line of research is motivated by applications in which the communication data rates from the channel input to the controller input are limited and feedback is available from the output of the channel to the input of the channel. A typical scenario of such control/communication system is the block diagram of Fig. 1, in which the dynamical system (i.e., source of information) is controlled via a limited capacity communication channel. This system can be viewed as a general communication system with feedback in which the output of the controlled system is the information source which is transmitted over a feedback communication channel to the controller, whose output is the input to the controlled system.

The present paper is concerned with necessary conditions for reliable data reconstruction (known as observability in the literature) and stability subject to limited channel capacity and uncertainty in the source of information. Throughout, we consider the system of Fig. 1 in which the dynamical system is controlled via a limited capacity communication channel when the source statistics belong to a prescribed class.
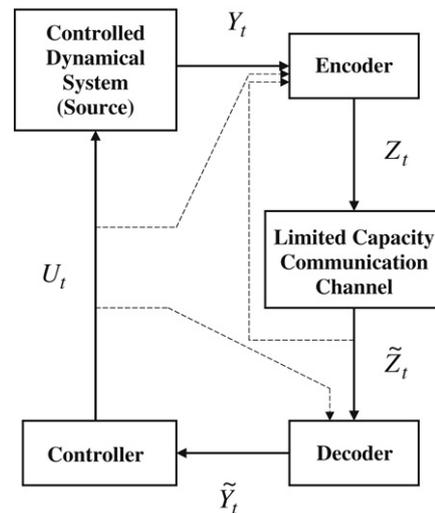
Fig. 1. Control/communication system.

The control/communication system of Fig. 1 is defined on a complete probability space $(\Omega, \mathcal{F}(\Omega), P)$ with filtration $\{\mathcal{F}_t\}_{t\geq 0}$; $t \in \mathbf{N}_+ \triangleq \{0, 1, 2, \ldots\}$, where, $Y_t, Z_t, \tilde{Z}_t, \tilde{Y}_t$, and $U_t, t \in \mathbf{N}_+$, are Random Variables (R.V.'s) denoting the source message, channel input codeword, channel output codeword, the reproduction of the source message, and the control input to the source, respectively.

* Corresponding author. Tel.: +1 613 255 0213.
E-mail address: farhadialireza@yahoo.com (A. Farhadi).

The objective is to find general necessary conditions for observability and stability of a class of sources, which is independent of the information available to the encoder, decoder and controller (i.e., the information patterns of the encoder, decoder and controller). Here, observability means reconstruction in the sense that $\tilde{Y}_t$ follows Y_t with some distortion, while the stability means that the state variable is bounded for all times, when there is a limited data rate constraint.

References [1–12] are representative although not exhaustive of the recent activity addressing necessary and sufficient conditions for stability and observability of unstable control systems subject to limited channel capacity. The necessary part is often obtained through the converse of the Information Transmission Theorem ([13], p. 72), which states that the information capacity should be at least equal to the rate at which the information is generated by the source (subject to a distortion when continuous sources are invoked).

The materials presented in this paper compliment those found in [5,14,15] in the problem formulation, methodology considered and the results obtained. Specifically, we wish to extend some of the concepts and results associated with the control/communication systems to a class of sources or dynamical systems, which are controlled over limited capacity channels. Thus, instead of considering a single source generating information which is then transmitted over the communication channel, we consider a class of sources or dynamical systems. Since the methodology is information theoretic, we introduce the notion of entropy for a class of sources which represents the amount of information generated by this class. This is defined as a maximization of the entropy of the source over the class of admissible sources considered, which corresponds to the maximum amount of information generated by the class of sources. Then, we invoke a modified version of the information transmission theorem and the Shannon lower bound to account for the class of sources, to link them to the definitions of uniform observability and robust stability. Although the approach is information theoretic, it is evident that the methodology is reminiscent to the mini-max approach associated with the investigation of robustness of the control systems [16].

This generalization is important in control applications since dynamical models are often simplified representations of the true models and hence, the robustness of the necessary conditions for reliable communication and robust stability should be investigated. This may be viewed as the first step towards introducing a robust analysis for the system of Fig. 1, in an analogous manner as it is done in robust control [17].

The general necessary conditions presented in this note, for existence of an encoder, decoder and controller for uniform observability and robust stability state that $\mathcal{C} \geq R_r(D_v) \geq R_{S,r}(D_v)$, where $\mathcal{C}$ is the information capacity of the channel, $R_r(D_v)$ is the mini-max rate distortion in which the maximization is with respect to the class of sources and the minimization is over the class of reproduction kernels (between the source and its reconstruction), and $R_{S,r}(D_v)$ is the maximization of the Shannon lower bound over the class of sources. Throughout the paper, the class of sources is described by a relative entropy constraint between a class of source densities and a given nominal source density. An explicit solution to the maximum entropy for such class of sources is found; and its connection to Rényi entropy is shown. Furthermore, the maximum entropy is calculated for a class of partially observed Gauss Markov sources.

The paper is organized as follows. In Section 2, the robust entropy (i.e., the entropy for a class of sources) and the robust entropy rate are defined, and the explicit solution to the robust entropy and its connection to Rényi entropy are given. In Section 3, a class of partially observed Gauss Markov sources is considered, and the robust entropy rate is calculated. Finally, in Section 4, the notion of robust entropy rate is employed to derive necessary conditions for uniform observability and robust stability of the control/communication system of Fig. 1. The derivations are given in the Appendix, while the notation employed is presented in Table 1.

2. Robust entropy and entropy rate

In this section, the definitions of robust entropy and subsequently robust entropy rate for a class of sources are given. Subsequently, by considering a class of sources described by a relative entropy constraint, the explicit solution to the robust entropy and entropy rate are presented. These results are employed in Section 4 to derive necessary conditions for uniform observability and robust stability of a class of controlled systems over a limited capacity communication channel.

Robust entropy and entropy rate definition. Let $\mathcal{D}$ denote the set of all Probability Density Functions (PDF's) corresponding to a R.V. $Y : (\Omega, \mathcal{F}(\Omega)) \rightarrow (\mathfrak{R}^d, \mathcal{B}(\mathfrak{R}^d))$ and $\mathcal{D}^{0,T}$ denote the set of all joint PDF's corresponding to a sequence of such R.V.'s $Y^T \triangleq \{Y_t\}_{t=0}^T, Y_t : (\Omega, \mathcal{F}(\Omega)) \rightarrow (\mathfrak{R}^d, \mathcal{B}(\mathfrak{R}^d))$. In the real word applications, the source statistics are not entirely known; rather, they are known to belong to a specific class of sources, known as the uncertain class. This class of sources is often characterized by the source density (resp. joint source density), with respect to a nominal fixed source density $g_Y \in \mathcal{D}$ (resp. $g_{Y^T} \in \mathcal{D}^{0,T}$). Thus, the nominal source corresponds to the a priori knowledge in the absence of the true knowledge of the source. Suppose the true source density, $f_Y \in \mathcal{D}$ (resp. $f_{Y^T} \in \mathcal{D}^{0,T}$) belongs to the class $\mathcal{D}_{SU} \subset \mathcal{D}$ (resp., $\mathcal{D}_{SU}^{0,T} \subset \mathcal{D}^{0,T}$), then, the entropy associated with a class of sources is defined as follows.

Definition 2.1 (*Entropy and Entropy Rate for a Class of Sources*). Suppose a class of R.V.'s $Y : (\Omega, \mathcal{F}) \rightarrow (\mathfrak{R}^d, \mathcal{B}(\mathfrak{R}^d))$ induces a class of PDF's belonging to the class $f_Y \in \mathcal{D}_{SU}, \mathcal{D}_{SU} \subset \mathcal{D}$. The robust entropy associated with the family $\mathcal{D}_{SU}$ of the sources is defined by

$$H_r(f_Y^*) \triangleq \sup_{f_Y \in \mathcal{D}_{SU}} H_S(f_Y), \tag{1}$$

where $H_S(f_Y) \triangleq -\int_{\mathfrak{R}^d} f_Y(y) \log f_Y(y) \mathrm{d}y$ is the Shannon entropy [18] and $f_Y^* \in \arg\sup_{f_Y \in \mathcal{D}_{SU}} H_S(f_Y)$. Moreover, for a class of sequences Y^{T-1} of R.V's with length T, which induces a class of joint PDF's belonging to the class $f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}, \mathcal{D}_{SU}^{0,T-1} \subset \mathcal{D}^{0,T-1}$, the robust entropy rate associated with the family $\mathcal{D}_{SU}^{0,T-1} \subset \mathcal{D}^{0,T-1}$ is defined by

$$\mathcal{H}_r(\mathcal{Y}) \triangleq \lim_{T \to \infty} \frac{1}{T} H_r(f_{Y^{T-1}}^*) = \lim_{T \to \infty} \frac{1}{T} \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} H_S(f_{Y^{T-1}}), \tag{2}$$

provided the limit exists.

The above extension of entropy corresponds to the maximum amount of information generated by a class of sources. An application to lossless source coding for a class of finite alphabet sources is given in [19].

Class of sources described by relative entropy. Denote by $g_Y \in \mathcal{D}$ the nominal source density and let the true source density $f_Y \in \mathcal{D}$ belong to the class of sources which is described by the relative entropy constraint $\mathcal{D}_{SU}(g_Y) \triangleq \{f_Y \in \mathcal{D}; H(f_Y \parallel g_Y) \leq R_c, g_Y \in \mathcal{D}\}$, where $H(f_Y \parallel g_Y) \triangleq \int \log(\frac{f_Y(y)}{g_Y(y)}) f_Y(y) \mathrm{d}y$ (when $\log(\frac{f_Y(y)}{g_Y(y)}) f_Y(y)$ is integrable and $g_Y(y) = 0$ implies $f_Y(y) = 0$ for all such y's), $\log(.)$ is the natural logarithm and $R_c \in [0, \infty)$ is fixed.

Table 1
Summary of Notation

Notation	Description	Notation	Description
Y^T	$(Y_0, Y_1, \ldots, Y_T)$	$\mathcal{D}$	Set of all PDF's
$\mathcal{D}^{0,T}$	Set of all PDF's associated with Y^T	$\mathcal{D}_{SU}, \mathcal{D}_{SU}^{0,T}$	Set of admissible density functions
$\mathbf{N}_+$	$\{0, 1, 2, \ldots\}$	$\mathfrak{R}^d$	Set of d-dimensional real vectors
$\log(.)$	Natural logarithm	$L_1(\mathfrak{R}^d, \mathfrak{R}^+)$	Lebesgue measurable integrable functions
$\|.\|$	Euclidian norm	g_Y, g_{Y^T}	Nominal density functions
f_Y, f_{Y^T}	True source density function	$H_S(g_Y)$	Shannon entropy
$\mathcal{H}_S(\mathcal{Y})$	Shannon entropy rate	$H_R(g_Y)$	Rényi entropy
$R_r^{SRD}(D_v)$	Robust sequential rate distortion	$H_r(f_Y)$	Robust entropy
$\mathcal{H}_r(\mathcal{Y})$	Robust entropy rate	$H(f_Y \parallel g_Y)$	Relative entropy
$\mathcal{C}_n$	Capacity of n channel uses	$\mathcal{C}$	Capacity
$I(.;.)$	Mutual information	$R_r(D_v)$	Robust rate distortion
$R_{S,r}(D_v)$	Robust Shannon lower bound	$Var[.]$	Variance
$Cov(.)$	Covariance	$o(.)$	Residue
$E[.]$	Expected value	$Pr(A)$	Probability of the event A
$N(m, \Gamma)$	Gaussian distribution	A^{-1}	Matrix inverse
$I_{m \times m}$	$(m \times m)$ identity matrix	$'$	Matrix transpose
$\det(.)$	Determinate of a square matrix	$\sigma\{.\}$	Sigma algebra
$\mathcal{D}_{SU}(g_Y)$	$\{f_Y \in \mathcal{D}; H(f_Y \parallel g_Y) \leq R_c\}$	$\mathcal{D}_{SU}(g_{Y^{T-1}})$	$\{f_{Y^{T-1}}; H(f_{Y^{T-1}} \parallel g_{Y^{T-1}}) \leq TR_c\}$

Next, consider the robust entropy definition over the class of sources described by the relative entropy constraint defined as follows.

$$H_r(f_Y^*) \triangleq \sup_{f_Y \in \mathcal{D}_{SU}(g_Y)} H_S(f_Y). \tag{3}$$

In the following theorem, we derive the maximum information generated by this class of sources.

Theorem 2.2. *Suppose for some* $s \in [0, \infty)$, $(g_Y(y))^{\frac{s}{1+s}} \in L_1(\mathfrak{R}^d, \mathfrak{R}^+)$, *where* $L_1(\mathfrak{R}^d, \mathfrak{R}^+)$ *is the set of non-negative Lebesgue measurable integrable functions defined on* $\mathfrak{R}^d$. *Then, we have the following.*

(i) *The robust entropy associated with the class* $\mathcal{D}_{SU}(g_Y)$ *is given by*

$$H_r(f_Y^{*,s^*}) = \sup_{f_Y \in \mathcal{D}_{SU}(g_Y)} H_S(f_Y)$$

$$= \min_{s \geq 0} \left[sR_c + (1+s) \log \int (g_Y(y))^{\frac{s}{1+s}} \, dy \right]. \tag{4}$$

Moreover, if for some $s \in [0, \infty)$, $(g_Y(y))^{\frac{s}{1+s}} \log(g_Y(y))^{-\frac{1}{1+s}} \in L_1(\mathfrak{R}^d, \mathfrak{R}^+)$, *the supremum* $f_Y^{*,s}$ *is achieved by*

$$f_Y^{*,s}(y) = \frac{(g_Y(y))^{\frac{s}{1+s}}}{\int (g_Y(y))^{\frac{s}{1+s}} \, dy}. \tag{5}$$

(ii) *If for some* $s \geq 0$, $(\log g_Y(y))(g_Y(y))^{\frac{s}{1+s}} \in L_1(\mathfrak{R}^d, \mathfrak{R}^+)$ *and* $(\log g_Y(y))^2 (g_Y(y))^{\frac{s}{1+s}} \in L_1(\mathfrak{R}^d, \mathfrak{R}^+)$, *then the minimum in* (4) *with respect to* $s \geq 0$ *is the solution of* $H(f_Y^{*,s} \parallel g_Y)|_{s=s^*} = R_c$.

Moreover, $H(f_Y^{*,s} \parallel g_Y)$ *is a non-increasing function of* $s \geq 0$, *that is, for* $0 \leq s^* \leq s_1 \leq s_2$

$$0 \leq H(f_Y^{*,s} \parallel g_Y)\big|_{s=s_2} \leq H(f_Y^{*,s} \parallel g_Y)\big|_{s=s_1}$$

$$\leq H(f_Y^{*,s} \parallel g_Y)\big|_{s=s^*} = R_c. \tag{6}$$

Proof. See Appendix. $\quad\square$

Remark 2.3. (4) has a similar form as the error exponent formula for finite alphabet source coding problems. However, in the current setting this formula is obtained for a class of sources, via the relative entropy uncertainty while the alphabet is general.

(ii) An alternative expression for (4), using Taylor's expansion, is given by $H_r(f_Y^{*,s^*}) = \min_{s \geq 0}[sR_c + H_S(g_Y) + \frac{1}{2(1+s)} Var[\log \frac{1}{g_Y(y)}] +$

$o(\frac{1}{1+s})]$. This shows that robust entropy also includes the variance of self information $-\log g_Y(y)$ around the entropy $H_S(g_Y)$.

(iii) The above solution to the robust entropy is related to the Rényi entropy defined by $H_R(g_Y) \triangleq \frac{1}{1-\alpha} \log \int g_Y^\alpha(y) dy$, $\alpha > 0$, $\alpha \neq 1$, $g_Y^\alpha(y) \in L_1(\mathfrak{R}^d, \mathfrak{R}^+)$ [20] which is obtained by relaxing the mean value property of the Shannon entropy from an arithmetic to an exponential mean. Assume $s^* > 0$ and let $\alpha = \frac{s}{1+s}$, $s > 0$; then

$$\min_{\alpha \in (0,1)} H_R(g_Y) \leq H_r(f_Y^{*,s^*}) = \min_{\alpha \in (0,1)} \left\{ \frac{\alpha}{1-\alpha} R_c + H_R(g_Y) \right\}$$

$$\leq \frac{\alpha}{1-\alpha} R_c + H_R(g_Y). \tag{7}$$

Although, it is well known that Rényi entropy gives as a special case the Shannon entropy [21], as discussed in [22] one special applications of Rényi entropy is to measure the complexity of a signal through its so called Time-Frequency Representation (TFR). The negative values taken on by most TFR's prohibit the application of the Shannon entropy. As it is shown in [22], for certain values of $\alpha > 0$, the Rényi entropy measures the signal complexity. Note that the observation that maximization of entropy over a relative entropy constraint yields Rényi entropy, was first investigated in [19] in the context of lossless source coding. Note also that in (4) the term $(1 + s) \log \int (g_Y(y))^{\frac{s}{1+s}} dy$ can be view as the Rényi entropy with $\alpha = \frac{s}{1+s}$, in which this entropy can be computed for a large class of probability densities (see the anthology of densities given in [21]).

The extension of Theorem 2.2 to a class of sequences is presented in the following corollary which is a direct consequence of Theorem 2.2.

Corollary 2.4. *Consider a class of sequences* Y^{T-1} *of R.V.'s with corresponding joint density function* $f_{Y^{T-1}} \in \mathcal{D}_{SU}(g_{Y^{T-1}}) \subset \mathcal{D}^{0,T-1}$, *described by*

$$\mathcal{D}_{SU}(g_{Y^{T-1}}) = \left\{ f_{Y^{T-1}} \in \mathcal{D}^{0,T-1}; H(f_{Y^{T-1}} \parallel g_{Y^{T-1}}) \leq TR_c \right\}. \tag{8}$$

Then, the statements of Theorem 2.2 hold for the class of sequences Y^{T-1} *by replacing* R_c *with* TR_c. *Consequently, the robust entropy* $H_r(f_{Y^{T-1}}^{*,s^*})$ *corresponding to the class* (8) *is given by* (4), *in which* R_c *is replaced by* TR_c *and* g_Y *by* $g_{Y^{T-1}}$. *Subsequently, the robust entropy rate corresponding to the class* (8) *is given by* $\mathcal{H}_r(\mathcal{Y}) = \lim_{T \to \infty} \frac{1}{T} H_r(f_{Y^{T-1}}^{*,s^*})$, *provided the limit exists.*

Note that if the sequence Y^{T-1} is generated via a stochastic partially observed dynamical system, the class of uncertainty described by (8) can model unknown dynamics in both the state and the observation. The observation is affected by state dynamics. Therefore, uncertainty in the state dynamics also affects the observation, in which this uncertainty in the dynamics can be described by (8).

Next, in the following Lemma, the robust entropy rate for a class of sources described via the relative entropy constraint (8) and a Gaussian distributed nominal source, is calculated. This Lemma follows from the result of Corollary 2.4.

Lemma 2.5. *Consider a class of sequences $Y^{T-1} = \{Y_t\}_{t=0}^{T-1}$, $Y_t \in \Re^d$, for which the nominal source joint density $g_{Y^{T-1}} \in \mathcal{D}^{0,T-1}$ is a Td-dimensional Gaussian distributed vector (i.e., $Y^{T-1} \sim N(m_T, \Gamma_T)$, $\Gamma_T \triangleq Cov[(Y'_0, \ldots, Y'_{T-1})']$ with Shannon entropy rate $\mathcal{H}_S(\mathcal{Y}) \triangleq \lim_{T \to \infty} \frac{1}{T} H_S(g_{Y^{T-1}})$. Then,*

$$\mathcal{H}_r(\mathcal{Y}) = \frac{d}{2} \log \left(\frac{1+s^*}{s^*} \right) + \mathcal{H}_S(\mathcal{Y}), \tag{9}$$

where for a given $R_c \in [0, \infty)$, $s^* > 0$ is the unique solution of the following non-linear equation

$$R_c = -\frac{d}{2} \log \left(\frac{1+s^*}{s^*} \right) + \frac{d}{2s^*}, \quad R_c \in [0, \infty). \tag{10}$$

Proof. See Appendix. □

3. Class of sources

In this section, the entropy rate is calculated for a class of partially observed Gauss Markov sources.

Consider the following partially observed Gauss Markov nominal source model

$$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t \geq 0}) : \begin{cases} X_{t+1} = AX_t + BW_t, & X_0 = X, \\ Y_t = CX_t + DG_t, & t \in \mathbf{N}_+, \end{cases} \tag{11}$$

where $X_t \in \Re^q$ denotes the unobserved (state) process, $Y_t \in \Re^d$ is the observed process, $W_t \in \Re^m, G_t \in \Re^l$, W_t is Independent Identically Distributed (i.i.d.) $\sim N(0, I_{m \times m})$, G_t is i.i.d. $\sim N(0, I_{l \times l})$, $X_0 \sim N(\bar{x}_0, \bar{V}_0)$, and $\{X_0, G_t, W_t\}$ are mutually independent, $\forall t \in \mathbf{N}_+$. Here it is assumed that (C, A) is detectable, $(A, (BB')^{\frac{1}{2}})$ is stabilizable, and $D \neq 0$.

Denote by $g_{Y^{T-1}} \in \mathcal{D}^{0,T-1}$ the joint density function of the sequence Y^{T-1} produced by the nominal source model (11). Then, the class of source densities $f_{Y^{T-1}} \in \mathcal{D}^{0,T-1}$ corresponding to the nominal source density $g_{Y^{T-1}}$ is described by (8). One possible uncertain class of densities is the one generated by letting $A \to A + \Delta A$ and $C \to C + \Delta C$ in (11), while satisfying the relative entropy constraint. This will lead to a specific constraint described by ΔA and ΔC. Next, in the following theorem, we recall the results of [15]. We shall use these results to find the robust entropy rate.

Theorem 3.1 (*[15]*).

(i) *Let $\{Y_t; t \in \mathbf{N}_+\}$, $Y_t \in \Re^d$, be a Gaussian vector and define $\Gamma_T \triangleq Cov[(Y'_0, \ldots, Y'_{T-1})']$, $Z_t \triangleq Y_t - E[Y_t|\sigma\{Y_0, \ldots, Y_{t-1}\}]$, $\Lambda_t \triangleq Cov(Z_t)$, and $\Lambda_\infty = \lim_{T \to \infty} \Lambda_T \neq 0$, where σ- denotes the sigma algebra generated by the sequence Y^{T-1}.*

Then, the Shannon entropy rate of $\{Y_t; t \in \mathbf{N}_+\}$ in nats per time step is given by

$$\mathcal{H}_S(\mathcal{Y}) = \frac{d}{2} \log(2\pi e) + \lim_{T \to \infty} \frac{1}{2T} \log \det \Gamma_T$$

$$= \frac{d}{2} \log(2\pi e) + \frac{1}{2} \log \det \Lambda_\infty. \tag{12}$$

(ii) *Consider the nominal source model (11). Then, $\Lambda_\infty = CV_\infty C' + DD'$, where $V_t = E[\tilde{X}_t \tilde{X}'_t]$, $\tilde{X}_t \triangleq X_t - E[X_t|\sigma\{Y_0, \ldots, Y_{t-1}\}]$, $t \in \mathbf{N}_+$, and $V_\infty = \lim_{T \to \infty} V_T$ is the unique positive semi-definite solution of the following Algebraic Riccati-equation*

$$V_\infty = AV_\infty A' - AV_\infty C'(CV_\infty C' + DD')^{-1}CV_\infty A' + BB'. \tag{13}$$

Subsequently, the Shannon entropy rate of the observed process $\{Y_t; t \in \mathbf{N}_+\}$ of the nominal source model (11) is given by

$$\mathcal{H}_S(\mathcal{Y}) = \frac{d}{2} \log(2\pi e) + \frac{1}{2} \log \det \Lambda_\infty. \tag{14}$$

Proof. The proof of the first part follows from Cholesky decomposition ([23], pp. 44) and the proof of the second part follows from ([23], p. 156–158, and [23] Theorem 4.2). □

Next, in the following corollary, using the results of Lemma 2.5 and Theorem 3.1, we compute the robust entropy rate.

Corollary 3.2. *The robust entropy rate of the class of sources described via the relative entropy constraint (8) with the nominal source model (11), is given by $\mathcal{H}_r(\mathcal{Y}) = \frac{d}{2} \log(\frac{1+s^*}{s^*}) + \mathcal{H}_S(\mathcal{Y})$, $\mathcal{H}_S(\mathcal{Y}) = \frac{d}{2} \log(2\pi e) + \frac{1}{2} \log \det \Lambda_\infty$, where $s^* > 0$ is the unique solution of (10) and Λ_∞ is given in Theorem 3.1.*

Remark 3.3. From (10), it follows that the case $R_c = 0$ corresponds to $s^* \to \infty$. Letting $s^* \to \infty$ in above result, we obtain $\mathcal{H}_r(\mathcal{Y}) = \mathcal{H}_S(\mathcal{Y})$. That is, the robust entropy rate is equal to the Shannon entropy rate of the nominal source. This is expected since $R_c = 0$ corresponds to a single source.

4. Application in control/communication systems

In this section, the robust entropy rate is used to establish necessary conditions for uniform observability and robust stability of the control/communication system of Fig. 1, described by a class of controlled sources.

The precise definitions of uniform observability and robust stability of the system of Fig. 1 considered in this paper are defined as follows.

Definition 4.1 (*Uniform Observability in Probability and r-Mean*). Consider the block diagram of Fig. 1 described by a class of sources. Let $f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1} \subset \mathcal{D}^{0,T-1}$ be the joint density function of the sequence Y^{T-1} produced by the class of sources.

(i) The controlled source is called uniform observable in probability if for a given $\delta \geq 0$ and $D_v \in [0, 1)$, there exists (a control sequence), an encoder and a decoder such that $\lim_{T \to \infty} \frac{1}{T} \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} \sum_{k=0}^{T-1} E \rho(Y_k, \tilde{Y}_k) \leq D_v$, where $\rho(Y, \tilde{Y})$ is defined by $\rho(Y, \tilde{Y}) = 1$ if $\|Y - \tilde{Y}\| > \delta$ and $\rho(Y, \tilde{Y}) = 0$ if $\|Y - \tilde{Y}\| \leq \delta$.

(ii) For a given $r > 0$ and a finite $D_v \geq 0$, the controlled source is called uniform observable in r-mean if $\lim_{T \to \infty} \frac{1}{T} \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} \sum_{k=0}^{T-1} E \|Y_k - \tilde{Y}_k\|^r \leq D_v$.

Definition 4.2 (*Robust Stability in Probability and r-Mean*). Consider the block diagram of Fig. 1 described by a class of controlled sources, in which $Y_t = H_t + \Upsilon_t$, where H_t is the signal to be controlled and Υ_t is a function of perturbation and the measurement noise. Let $f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1} \subset \mathcal{D}^{0,T-1}$ denote the joint density function of the sequence Y^{T-1} produced by the class of sources.

(i) The controlled source is called robust stabilizable in probability if for a given $\delta \geq 0$ and $D_v \in [0, 1)$, there exists a controller, an encoder, and a decoder such that $\lim_{T\to\infty} \frac{1}{T} \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} \sum_{k=0}^{T-1} E\rho(H_k, 0) \leq D_v$, where $\rho(.,.)$ is the one defined in Definition 4.1.

(ii) For a given $r > 0$ and a finite $D_v \geq 0$, the controlled source is called robust stabilizable in r-mean if $\lim_{T\to\infty} \frac{1}{T} \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} \sum_{k=0}^{T-1} E\|H_k - 0\|^r \leq D_v$.

In this section, by finding a connection between capacity, robust rate distortion (i.e., mini-max rate distortion) and a variant of the Shannon lower bound, necessary conditions for uniform observability and robust stability in the form of a lower bound on the capacity are derived. Since the derivation of these conditions involves capacity and rate distortion, we shall recall the definition of these measures.

Definition 4.3 (*Information Capacity [18]*). Consider a memoryless channel without feedback and let Z^{n-1} and $\tilde{Z}^{n-1}$ be the channel input and output sequences, respectively. Let $\mathcal{D}_{CI}$ denotes the set of joint density functions $f_{Z^{n-1}}$ associated with the sequence Z^{n-1} which satisfy certain channel input power constraint. The Shannon information capacity for n channel uses, is defined by

$$\mathcal{C}_n \triangleq \sup_{f_{Z^{n-1}} \in \mathcal{D}_{CI}} I(Z^{n-1}; \tilde{Z}^{n-1})$$

$$I(Z^{n-1}; \tilde{Z}^{n-1}) = \int \log\left(\frac{f_{\tilde{Z}^{n-1}|Z^{n-1}}}{f_{\tilde{Z}^{n-1}}}\right) f_{\tilde{Z}^{n-1}|Z^{n-1}} f_{Z^{n-1}} \mathrm{d}z^{n-1} \mathrm{d}\tilde{z}^{n-1}.$$
(15)

Subsequently, the information capacity, in nats per channel use, is $\mathcal{C} \triangleq \lim_{n\to\infty} \frac{1}{n} \mathcal{C}_n$ provided the limit exists.

For Discrete Memoryless Channels (DMC's) without feedback and memoryless Additive White Gaussian Noise (AWGN) channels without feedback, the information channel capacity of Definition 4.3 represents the operational capacity [18]; or simply the channel capacity. Furthermore, when feedback is employed the mutual information in (15) is replaced by the directed information $I(Z^{n-1} \to \tilde{Z}^{n-1}) \triangleq \sum_{i=0}^{n-1} I(Z^i; \tilde{Z}_i|\tilde{Z}^{i-1}) = \sum_{i=0}^{n-1} \int \log(\frac{f_{\tilde{Z}_i|\tilde{Z}^{i-1},Z^i}}{f_{\tilde{Z}_i|\tilde{Z}^{i-1}}}) f_{\tilde{Z}_i|Z^i} f_{Z^i} \mathrm{d}z^i \mathrm{d}\tilde{z}^i$ [24]. Under this modification the capacity of DMC's and AWGN channels with and without feedback is the same. Please note that if Z and $\tilde{Z}$ represents the channel input-output codewords, respectively, associated with the source message Y, the capacity is given in nats per source message; and subsequently in nats per time step.

Definition 4.4 (*Robust Information Rate Distortion [25]*). Let Y^{T-1} and $\tilde{Y}^{T-1}$ be sequences with length T of the source and reproduction of the source messages, respectively. Let $f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1} \subset \mathcal{D}^{0,T-1}$ denote the probability density function of Y^{T-1} which belongs to the class $\mathcal{D}_{SU}^{0,T-1} \subset \mathcal{D}^{0,T-1}$. Denote by $f_{\tilde{Y}^{T-1}|Y^{T-1}} \in \mathcal{D}^{0,T-1}$ the conditional density function of $\tilde{Y}^{T-1}$ given Y^{T-1} and let $\mathcal{D}_{DC} \triangleq \{f_{\tilde{Y}^{T-1}|Y^{T-1}} \in \mathcal{D}^{0,T-1}; E\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v\}$ denote the set of distortion constraint, in which $D_v \geq 0$ is the distortion value and $\rho_T \in [0, \infty)$ is the distortion measure.

Then, the robust information rate distortion for the class $\mathcal{D}_{SU}^{0,T-1} \in \mathcal{D}^{0,T-1}$ is defined by

$$R_r(D_v) \triangleq \lim_{T\to\infty} \frac{1}{T} R_{T,r}(D_v),$$

$$R_{T,r}(D_v) \triangleq \inf_{f_{\tilde{Y}^{T-1}|Y^{T-1}} \in \mathcal{D}_{DC}} \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} I(Y^{T-1}; \tilde{Y}^{T-1})$$
(16)

provided the limit exists.

In [25] it is shown that for single letter distortion measure, i.e., $\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) = \frac{1}{T} \sum_{t=0}^{T-1} \rho(Y_t, \tilde{Y}_t)$ and a class of memoryless sources in which $\mathcal{D}_{SU}^{0,T-1}$ is compact, (16) represents the minimum rate for uniform reliable communication up to the distortion value D_v. For sources which are part of the control loop (see Fig. 1), it is desirable that the time ordering for encoding and decoding to be causal (as described in [6]). For such sources the minimum in (16) must be taken over the set $\mathcal{D}_{DC}^{SRD} \triangleq \{\{f_{\tilde{Y}_t|\tilde{Y}^{t-1},Y^t}\}_{t=0}^{T-1}; \frac{1}{T} \sum_{t=0}^{T-1} E[\rho(Y_t, \tilde{Y}_t)] \leq D_v\}$. That is (16) must be replaced by the following

$$R_r^{SRD}(D_v) \triangleq \lim_{T\to\infty} \frac{1}{T} R_{T,r}^{SRD}(D_v),$$

$$R_{T,r}^{SRD}(D_v) \triangleq \inf_{\{\{f_{\tilde{Y}_t|\tilde{Y}^{t-1},Y^t}\}_{t=0}^{T-1} \in \mathcal{D}_{DC}^{SRD}\}} \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} I(Y^{T-1} \to \tilde{Y}^{T-1}) \quad (17)$$

More elaboration on this issue is found in [15].

Next, in the following theorem, we find a connection between capacity and robust rate distortion. This connection is valid under the assumption that the outputs of the source, encoder, channel, and decoder of the control/communication system of Fig. 1 are subject to the conditional independence assumption. That is, for any $T, n \in \{1, 2, \ldots\}$, the conditional independence assumption of the source output sequence Y^{T-1}, the channel input sequence Z^{n-1} and the channel output $\tilde{Z}^{n-1}$ is equivalent to $f_{\tilde{Z}_t|Z^t, \tilde{Z}^{t-1}, Y^{T-1}} = f_{\tilde{Z}_t|Z^t, \tilde{Z}^{t-1}}$ ($t \in \{0, 1, \ldots, n-1\}$), and similarly for the rest of the blocks.

Theorem 4.5 (*Robust Information Transmission Theorem*). *Consider the block diagram of Fig. 1 subject to the conditional independence assumption and uncertainty in the source. Then*

(i) $I(Z^{n-1}; \tilde{Z}^{n-1}) \geq I(Z^{n-1} \to \tilde{Z}^{n-1}) \geq I(Y^{T-1}; \tilde{Z}^{n-1}) \geq I(Y^{T-1}; \tilde{Y}^{T-1}) \geq I(Y^{T-1} \to \tilde{Y}^{T-1}), \forall T, n \in \{1, 2, 3, \ldots\}$.

(ii) *When the control/communication system of Fig. 1 is described by DMC's or memoryless AWGN channels, a necessary condition for reproducing a sequence Y^{T-1} of the source messages, up to the distortion value D_v by $\tilde{Y}^{T-1}$, at the decoder end (i.e., $E\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v, \forall f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}$) using a sequence of the channel inputs-outputs with length n ($T \leq n$), is $\mathcal{C}_n \geq R_{T,r}(D_v)$.*

Proof. (i) The second inequality follows from ([24], Theorem 3). The first and the last inequalities also follow from ([24], Theorem 1) and the third inequality is a direct result of data processing inequality. (ii) See Appendix. □

Please note that under causal time-ordering, a tight necessary condition is given by $\mathcal{C}_n \geq R_{T,r}^{SRD}(D_v)$ which also follows from the data processing inequality of Theorem 4.5, (i). Next, we present a lower bound for the robust rate distortion in terms of the robust entropy rate of the source. We use this Lemma and Theorem 4.5 to relate the capacity to the robust entropy rate for uniform observability and robust stability.

Lemma 4.6 (*Robust Shannon Lower Bound*). *Let $Y^{T-1} = \{Y_t\}_{t=0}^{T-1}$, $Y_t \in \mathfrak{R}^d$ be a sequence with length T of the messages produced by a class of sources with corresponding joint density function $f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1} \subset \mathcal{D}^{0,T-1}$. Consider the following single letter distortion measure $\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) = \frac{1}{T} \sum_{t=0}^{T-1} \rho(Y_t, \tilde{Y}_t)$, where $\rho(Y_t, \tilde{Y}_t) = \rho(Y_t - \tilde{Y}_t) : \mathfrak{R}^d \to [0, \infty)$ is continuous.
Then, a lower bound for $\frac{1}{T} R_{T,r}(D_v)$ is given by*

$$\frac{1}{T} R_{T,r}(D_v) \geq \frac{1}{T} H_r(f_{Y^{T-1}}^*) - \max_{h \in G_D} H_S(h), \quad (18)$$

where G_D is defined by $G_D \triangleq \{h : \mathfrak{R}^d \to [0, \infty); \int_{\mathfrak{R}^d} h(\xi)\mathrm{d}\xi = 1, \int_{\mathfrak{R}^d} \rho(\xi)h(\xi)\mathrm{d}\xi \leq D_v, \xi \in \mathfrak{R}^d\}$. Moreover, when $\int_{\mathfrak{R}^d} e^{s\rho(\xi)}\mathrm{d}\xi < \infty$ for all $s < 0$, then $h^*(\xi) \in G_D$ that maximizes $H_S(h)$ is

given by $h^*(\xi) = \frac{e^{s\rho(\xi)}}{\int_{\mathfrak{R}^d} e^{s\rho(\xi)}\mathrm{d}\xi}$, where $s < 0$ satisfies the following $\int_{\mathfrak{R}^d} \rho(\xi)h^*(\xi)\mathrm{d}\xi = D_v$. Subsequently, when $R_r(D_v)$ and $\mathcal{H}_r(\mathcal{Y})$ exist, the robust Shannon lower bound $R_{S,r}(D_v)$ is given by the following

$$R_r(D_v) \geq \mathcal{H}_r(\mathcal{Y}) - \max_{h \in G_D} H_S(h) \triangleq R_{S,r}(D_v). \qquad (19)$$

Proof. See Appendix. □

Note that for distortion measure $\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) = \frac{1}{T}\sum_{t=0}^{T-1} \|Y_t - \tilde{Y}_t\|^2$, when $R_{T,r}(D_v) = \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} R_T(D_v)$ ($R_T(D_v)$ is the rate distortion function for a single source), the robust Shannon lower bound is exact for sufficiently small D_v.

Next, combining Theorem 4.5 and Lemma 4.6, necessary conditions for uniform observability and robust stability of the control/communication system of Fig. 1 are derived in the following theorem.

Theorem 4.7. *Consider the control/communication system of Fig. 1, under conditional independence assumption, described by a class of sources over DMC's or memoryless AWGN channels. Assume the robust entropy rate of the class of sources exists and it is finite.*

Then, (i) A necessary condition for uniform observability in probability, in nats per time step, is

$$\mathcal{C} \geq \mathcal{H}_r(\mathcal{Y}) - \frac{1}{2}\log[(2\pi e)^d \det \Gamma_g] = R_{S,r}(D_v), \qquad (20)$$

where $\mathcal{H}_r(\mathcal{Y})$ is the robust entropy rate of the class of sources and Γ_g is the covariance matrix of the Gaussian distribution $h^*(\xi) \sim N(0, \Gamma_g)$, $(\xi \in \mathfrak{R}^d)$ which satisfies $\int_{\|\xi\|>\delta} h^*(\xi)\mathrm{d}\xi = D_v$.

(ii) A necessary condition for r-mean uniform observability, in nats per time step, is

$$\mathcal{C} \geq \mathcal{H}_r(\mathcal{Y}) - \frac{d}{r} + \log\left(\frac{r}{dV_d\Gamma(\frac{d}{r})}\left(\frac{d}{rD_v}\right)^{\frac{d}{r}}\right) = R_{S,r}(D_v), \qquad (21)$$

where $\Gamma(.)$ is the gamma function and V_d is the volume of the unit sphere (e.g., $V_d = Vol(S_d)$; $S_d \triangleq \{\xi \in \mathfrak{R}^d; \|\xi\| < 1\}$).

Furthermore, when $Y_t = H_t + \Upsilon_t$, (20) and (21) are also necessary conditions for robust stability in probability and r-mean, respectively.

Proof. See Appendix. □

In Theorem 4.7, $\mathcal{H}_r(\mathcal{Y})$ can be a function of the control signal. Nevertheless, in the following proposition, it is shown that the robust entropy rate of a class of controlled sources can be bounded below by the robust entropy rate of the uncontrolled analogous sources no matter what the information patterns for the encoder and decoder are. Subsequently, for such uncertain controlled sources, the robust entropy rate of the uncontrolled analogous sources can be used in Theorem 4.7.

Proposition 4.8. *Consider a class of controlled sources described via (8), in which the nominal source model is a controlled version of the nominal source model (11) described by the following state space model*

$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t \geq 0}):$

$$\begin{cases} X_{t+1} = AX_t + BW_t + NU_t, & X_0 = X, \\ Y_t = H_t + DG_t, \ H_t = CX_t, & t \in \mathbf{N}_+, \end{cases} \qquad (22)$$

where $U_t \in \mathfrak{R}^o$, $N \in \mathfrak{R}^{q \times o}$, (C, A) is detectable, $(A, (BB')^{\frac{1}{2}})$ is stabilizable and $D \neq 0$.

Then, the robust entropy rate of this class of controlled sources is bounded below by $\mathcal{H}_r(\mathcal{Y}) \geq \frac{d}{2}\log(\frac{1+s^*}{s^*}) + \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log\det \Lambda_\infty$, where $s^* > 0$ is the unique solution of (10) and Λ_∞ is given in Theorem 3.1.

Proof. Follows from standard chain rule inequality of the Shannon entropy ([18], p. 239) and from the property that the conditioning reduces entropy ([18], p. 232) as well as by Gaussianity. □

We have the following remarks regarding the results of Theorem 4.7.

Remark 4.9. (i) The lower bounds (20) and (21) given in Theorem 4.7 hold for any observed process, no matter what the information patterns for the encoder, decoder and controller are.

(ii) When Theorem 4.7 is applied to controlled sources, then the entropy rate of the outputs of such sources which depends on the control sequence must be used. Nevertheless, from Proposition 4.8 we deduce that the bounds (20) and (21) also hold when the robust entropy rate is replaced by the robust entropy rate of the output process of the uncontrolled analogous sources.

(iii) Uniform observability and robust stability can be defined using single letter criteria as done in [5] rather than for sequences. For single letter criteria the definition of rate distortion should be also replaced by its single letter analogous (i.e., $T = 1$). Subsequently, in (Lemma 4.6, (18)) and Theorem 4.7, the robust entropy of the sequences will be replaced by the robust entropy of the single letter density f_Y over the family $\mathcal{D}_{SU}(g_Y)$ defined in Theorem 2.2, which is related to Rényi entropy of a R.V. Thus, the lower bounds given in Lemma 4.6 and Theorem 4.7 can be computed for a large classes of sources from [21].

(iv) For a class of controlled sources described via (8) with the nominal source model (22) over memoryless AWGN channels, a sufficient condition for reliable communication and control can be defined (by proposing an encoder/decoder and controller) via generalizations of [15], which is done for a single source. The method is based on implementing a source-channel matching technique [26] which results in a joint source channel encoding/decoding scheme, and controller designed.

5. Conclusion

In this paper, for the class of sources described by a constraint on the relative entropy, the explicit solution to robust entropy was found and its connection to Rényi entropy was illustrated. Subsequently, an application of robust entropy rate for uniform observability and robust stability of a control/communication system subject to limited capacity constraint was presented. For future direction, it would be interesting to establish the sufficiency of the conditions found in this paper by constructing actual encoder, decoder and controller that can achieve the obtained lower bounds.

Appendix

Proof of Theorem 2.2. Using ([27], p. 224) it can be shown that the constraint problem (3) is equivalent to the following unconstrained problem

$$H_r(f_Y^{*,s^*}) = \sup_{f_Y \in \mathcal{D}_{SU}(g_Y)} H_S(f_Y) = \min_{s \geq 0} \sup_{f_Y \in \mathcal{D}} L(s, f_Y)$$

$$L(s, f_Y) \triangleq H_S(f_Y) - s(H(f_Y \| g_Y) - R_c)$$

$$= sR_c + (1 + s) \left[\int \log(g_Y(y))^{-\frac{1}{1+s}} f_Y(y) \mathrm{d}y - H(f_Y \| g_Y) \right]. \quad (23)$$

Applying calculus of variation to $L(s, f_Y)$ yields

$$\sup_{f_Y \in \mathcal{D}} \left(\int \log(g_Y(y))^{-\frac{1}{1+s}} f_Y(y) \mathrm{d}y - H(f_Y \| g_Y) \right)$$

$$= \log \int (g_Y(y))^{\frac{s}{1+s}} \mathrm{d}y, \quad (24)$$

where supremum in (24) is attained at $f_Y^{*,s}(y) = \frac{(g_Y(y))^{\frac{s}{1+s}}}{\int (g_Y(y))^{\frac{s}{1+s}} \mathrm{d}y}$. This completes the proof of part i.

(ii) It can be easily shown that

$$\frac{\mathrm{d}}{\mathrm{d}s} L(s, f_Y^{*,s}) = R_c - H(f_Y^{*,s} \| g_Y)$$

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} L(s, f_Y^{*,s}) = \frac{1}{(1+s)^3} \left(\int f_Y^{*,s}(y) \log^2 g_Y(y) \mathrm{d}y \right.$$

$$\left. - \left(\int f_Y^{*,s}(y) \log g_Y(y) \mathrm{d}y \right)^2 \right)$$

$$\triangleq \frac{1}{(1+s)^3} Var_{f_Y^{*,s}} [\log g_Y(y)] \geq 0 \quad \forall s \geq 0. \quad (25)$$

Thus, the function $L(s, f_Y^{*,s})$ is a convex function of $s \geq 0$. The minimum over $s \geq 0$ is found by $\frac{\mathrm{d}L(s, f_Y^{*,s})}{\mathrm{d}s} \big|_{s=s^*} = 0$. This yields a minimizing $s^* \geq 0$ which is the solution of $H(f_Y^{*,s} \| g_Y) \big|_{s=s^*} = R_c$. Further, it is evident that $\frac{\mathrm{d}H(f_Y^{*,s} \| g_Y)}{\mathrm{d}s} = -\frac{\mathrm{d}^2 L(s, f_Y^{*,s})}{\mathrm{d}s^2} \leq 0$, which implies that $H(f_Y^{*,s} \| g_Y)$ is a non-increasing function of $s \geq 0$. $\square$

Proof of Lemma 2.5. Eq. (9) is a result of direct substitution of $g_{Y^{T-1}} \sim N(m_T, \Gamma_T)$ into Theorem 2.2. Further, by computing $f_{Y^{T-1}}^{*,s}$ from (5), we have $H(f_Y^{*,s} \| g_Y) = -\frac{Td}{2} \log(\frac{1+s^*}{s^*}) + \frac{Td}{2s^*}$. Consequently, by Theorem 2.2, ii, the minimizing $s^* > 0$ is the solution of $R_c = \frac{1}{T} H(f_Y^{*,s} \| g_Y) = -\frac{d}{2} \log(\frac{1+s^*}{s^*}) + \frac{d}{2s^*}$. Note that above expression can be written into the form $e^{-2R_c} = e^{-\frac{d}{s^*}} (\frac{1+s^*}{s^*})^d \triangleq M(s^*)$. But $\frac{\mathrm{d}M(s^*)}{\mathrm{d}s^*} > 0$. Subsequently, $M(s^*)$ is strictly increasing function of $s^* > 0$. Further, since $M(s^*)$ is a continuous function of $s^* > 0$, the minimum and the maximum of $M(s^*)$ is obtained at $s^* \to 0$ and $s^* \to \infty$, respectively. $\lim_{s^* \to 0} M(s^*) = 0$ and $\lim_{s^* \to \infty} M(s^*) = 1$. Subsequently, $M(s^*) \in (0, 1)$. On the other hand, $\forall R_c \geq 0, e^{-2R_c} \in (0, 1]$. Consequently, as $M(s^*)$ is a continuous and strictly increasing function of $s^* > 0$ which covers the range of $e^{-2R_c}, \forall R_c \geq 0$, there exists a unique $s^* > 0$ such that $e^{-2R_c} = M(s^*)$, or equivalently Eq. (10). $\square$

Proof of Theorem 4.5. Consider the case without feedback channel. If the encoding scheme yields an average distortion $E\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v$ for a class of sources, then from data processing inequality (Theorem 4.5, i) follows that

$$I(Z^{n-1}; \tilde{Z}^{n-1}) \geq I(Y^{T-1}; \tilde{Y}^{T-1}), \quad \forall f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1},$$

$$\sup_{\{f_{Z^{n-1}}: f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}\}} I(Z^{n-1}; \tilde{Z}^{n-1}) \geq \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} I(Y^{T-1}; \tilde{Y}^{T-1}),$$

$$\sup_{f_{Z^{n-1}} \in \mathcal{D}_{CI}} I(Z^{n-1}; \tilde{Z}^{n-1}) \geq \sup_{\{f_{Z^{n-1}}: f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}\}} I(Z^{n-1}; \tilde{Z}^{n-1})$$

$$\geq \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} I(Y^{T-1}; \tilde{Y}^{T-1}),$$

$$\mathcal{C}_n \triangleq \sup_{f_{Z^{n-1}} \in \mathcal{D}_{CI}} I(Z^{n-1}; \tilde{Z}^{n-1})$$

$$\geq \inf_{f_{\tilde{Y}^{T-1}|Y^{T-1}} \in \mathcal{D}_{DC}} \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} I(Y^{T-1}; \tilde{Y}^{T-1})$$

$$\triangleq R_{T,r}(D_v) \quad (26)$$

where the second inequality follows by taking supremum over a class of sources (please note that the channel inputs distribution is affected by the source distribution); and the third inequality follows since the class of channel inputs power constraint must include those channel inputs distribution which are affected by the class of sources; otherwise the reliable communication for the class of sources is not possible.

Thus $\mathcal{C}_n \geq R_{T,r}(D_v)$ under the assumption that there exists an encoding scheme that yields an average distortion $E\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v$ for a class of sources. This means that $\mathcal{C}_n \geq R_{T,r}(D_v)$ is a necessary condition for existence of such encoding scheme.

For the case of feedback channel, in (26) $I(Z^{n-1}; \tilde{Z}^{n-1})$ must be replaced by $I(Z^{n-1} \to \tilde{Z}^{n-1})$ and supremum must be taken over $\{f_{Z_i|Z^{i-1}, \tilde{z}^{i-1}}\}_{i=0}^{n-1} \in \mathcal{D}_{CI}$, in which for DMC's or memoryless AWGN channels with feedback, the capacity $\sup_{\{\{f_{Z_i|Z^{i-1}, \tilde{z}^{i-1}}\}_{i=0}^{n-1} \in \mathcal{D}_{CI}\}} I(Z^{n-1} \to \tilde{Z}^{n-1})$ is the same as the capacity of those channels without feedback, i.e., $\mathcal{C}_n$ [18]. $\square$

Proof of Lemma 4.6. From [13] it follows that $\forall f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1} \subset \mathcal{D}^{0,T-1}$, we have $\frac{1}{T} R_T(D_v) \triangleq \frac{1}{T} \inf_{f_{\tilde{Y}^{T-1}|Y^{T-1}} \in \mathcal{D}_{DC}} I(Y^{T-1}; \tilde{Y}^{T-1}) \geq \frac{1}{T} H_S(f_{Y^{T-1}}) - \max_{h \in G_D} H_S(h)$, where $R_T(D_v)$ is the rate distortion function for the density function $f_{Y^{T-1}}$; and when $\int_{\Re^d} e^{s\rho(\xi)} \mathrm{d}\xi < \infty, \forall s < 0$, the maximizer $h^*(\xi) \in G_D$ is given by $h^*(\xi) = \frac{e^{s\rho(\xi)}}{\int_{\Re^d} e^{s\rho(\xi)} \mathrm{d}\xi}$, where $s < 0$ satisfies the following $\int_{\Re^d} \rho(\xi) h^*(\xi) \mathrm{d}\xi = D_v$.

Consequently,

$$\sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} \frac{1}{T} R_T(D_v) \geq \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} \frac{1}{T} H_S(f_{Y^{T-1}}) - \max_{h \in G_D} H_S(h). \quad (27)$$

On the other hand, since

$$R_{T,r}(D_v) \triangleq \inf_{f_{\tilde{Y}^{T-1}|Y^{T-1}} \in \mathcal{D}_{DC}} \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} I(Y^{T-1}; \tilde{Y}^{T-1})$$

$$\geq \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^{0,T-1}} \inf_{f_{\tilde{Y}^{T-1}|Y^{T-1}} \in \mathcal{D}_{DC}} I(Y^{T-1}; \tilde{Y}^{T-1}), \quad (28)$$

we deduce the result. $\square$

Proof of Theorem 4.7 (*Uniform Observability*). Assume there exists an encoder and decoder such that the uniform observability in probability in the sense of Definition 4.1 is obtained. This implies that for a given $\delta \geq 0$ and $D_v \in [0, 1)$, there exist $T(\delta, D_v)$ such that, $\forall t \geq T(\delta, D_v), \sup_{f_{Y^{t-1}} \in \mathcal{D}_{SU}^{0,t-1}} \frac{1}{t} \sum_{k=0}^{t-1} \Pr(\|Y_k - \tilde{Y}_k\| > \delta) \leq D_v$. Subsequently, $\frac{1}{t} \sum_{k=0}^{t-1} \Pr(\|Y_k - \tilde{Y}_k\| > \delta) \leq D_v, \forall f_{Y^{t-1}} \in \mathcal{D}_{SU}^{0,t-1}$. Next, define the following single letter distortion measure $\rho_t(Y^{t-1}, \tilde{Y}^{t-1}) = \frac{1}{t} \sum_{k=0}^{t-1} \rho(Y_k, \tilde{Y}_k)$, where $\rho(., .)$ is given in Definition 4.1. Then, for $t \geq T(\delta, D_v), E\rho_t(Y^{t-1}, \tilde{Y}^{t-1}) =$

$\frac{1}{t} \sum_{k=0}^{t-1} \Pr(\|Y_k - \tilde{Y}_k\| > \delta) \leq D_v, \ \forall f_{Y^{t-1}} \in \mathcal{D}_{SU}^{0,t-1} \subset \mathcal{D}^{0,t-1}$. That is, uniform reconstructability up to the distortion value D_v is obtained for $t \geq T(\delta, D_v)$. Then, by Theorem 4.5 and Lemma 4.6, the capacity and robust rate distortion must for all $t \geq T(\delta, D_v)$ satisfy

$$\lim_{t\to\infty} \frac{1}{t} \mathcal{C}_t \geq \lim_{t\to\infty} \frac{1}{t} H_r(f_{Y^{t-1}}^*) - \max_{h\in G_D} H_S(h)$$

$$\mathcal{C} \geq \mathcal{H}_r(\mathcal{Y}) - \max_{h\in G_D} H_S(h). \tag{29}$$

Since among all distribution with the same covariance, the Gaussian distribution has the biggest entropy, $h^*(\xi) \in G_D$ that maximizes $H_S(h)$ is a Gaussian distribution which satisfies the boundary conditions of G_D. That is, $h^*(\xi) \sim N(0, \Gamma_g)$, in which Γ_g satisfies $\int_{\|\xi\|>\delta} h^*(\xi)\mathrm{d}\xi = D_v$. Consequently, substituting $\max_{h\in G_D} H_S(h) = H_S(h^*) = \frac{1}{2}\log[(2\pi\mathrm{e})^d \det \Gamma_g]$ in (29), the lower bound (20) is obtained. That is, (20) holds under the assumption that there exist an encoder and decoder that yield uniform observability in probability in the sense of Definition 4.1. This means that (20) is a necessary condition for existence of such encoder and decoder.

Necessary condition for uniform observability in r-mean is obtained along the same lines of above proof. The only difference is that from [28] it follows that for this case $\max_{h\in G_D} H_S(h) = \frac{d}{r} - \log(\frac{r}{dV_d\Gamma(\frac{d}{r})}(\frac{d}{rD_v})^{\frac{d}{r}})$ nats per time step.

(Robust stability). Follows similarly by considering the rate distortion between Y^{t-1} and Υ^{t-1}. $\quad\square$

References

[1] A.V. Savkin, I.R. Petersen, Set-valued state estimation via a limited capacity communication channel, IEEE Transactions on Automatic Control 48 (4) (2003) 676–680.
[2] V. Malyavej, A.V. Savkin, The problem of optimal robust Kalman state estimation via limited capacity digital communication channels, Systems and Control Letters 54 (3) (2005) 283–292.
[3] G.N. Nair, S. Dey, R.J. Evans, Infimum data rates for stabilizing Markov jump linear systems, in: Proceedings of 42nd IEEE Conference on Decision and Control, Hawaii, December 2003, pp. 1176–1181.
[4] G.N. Nair, R.J. Evans, I.M.Y. Mareels, W. Moran, Topological feedback entropy and nonlinear stabilization, IEEE Transactions on Automatic Control 49 (9) (2004) 1585–1597.
[5] S. Tatikonda, S. Mitter, Control over noisy channels, IEEE Transactions on Automatic Control 49 (7) (2004) 1196–1201.
[6] S. Tatikonda, A. Sahai, S. Mitter, Stochastic linear control over a communication channel, IEEE Transactions on Automatic Control 49 (9) (2004) 1549–1561.
[7] Nicola Elia, When Bode meets Shannon: Control-oriented feedback communication schemes, IEEE Transactions on Automatic Control 49 (9) (2004) 1477–1488.
[8] K. Li, J. Baillieul, Robust quantization for digital finite communication bandwidth (DFCB) control, IEEE Transactions on Automatic Control 49 (9) (2004) 1573–1584.
[9] D. Liberzon, J.P. Hespanha, Stabilization of nonlinear systems with limited information feedback, IEEE Transactions on Automatic Control 50 (6) (2005) 910–915.
[10] C.D. Charalambous, A. Farhadi, Robust entropy rate for uncertain sources: Applications to communication and control systems, in: Proceedings of 9th Canadian Workshop on Information Theory, Montreal, June 2005, pp. 307–310.
[11] A. Farhadi, C.D. Charalambous, Robust control of a class of feedback systems subject to limited capacity constraints, in: Proceedings of 46th IEEE Conference on Decision and Control, New Orleans, December 2007, pp. 3970–3975.
[12] N.C. Martins, A. Dahleh, N. Elia, Feedback stabilization of uncertain systems in the presence of a direct link, IEEE Transactions on Automatic Control 51 (3) (2006) 438–447.
[13] Toby Berger, Rate Distortion Theory: A Mathematical Basis for Data Compression, Printice-Hall, 1971.
[14] S. Yuksel, T. Basar, Coding and control over discrete noisy forward and feedback channels, in: Proceedings of 44th IEEE Conference on Decision and Control, and the European Control Conference, Seville, December 12–15, 2005, pp. 2517–2522.
[15] C.D. Charalambous, Alireza Farhadi, LQG optimality and separation principle for general discrete time partially observed stochastic systems over finite capacity communication channels, Automatica, in 10 pages (in press).
[16] T. Basar, P. Bernhard, H-infinity Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach, 2nd edition, Birkhauser, 1995.
[17] C.D. Charalambous, F. Rezaei, Stochastic uncertain systems subject to relative entropy constraints: Induced norms and monotonicity properties of minimax games, IEEE Transactions on Automatic Control 54 (4) (2007) 647–663.
[18] T.M Cover, Joy A. Thomas, Elements of Information Theory, John Wiley and Sons, 1991.
[19] F. Rezaei, C.D. Charalambous, Robust coding for uncertain sources: A minimax approach, in: Proceedings of the 2005 IEEE International Symposium on Information Theory, Adelaide, Australia, September 4–9, 2005, pp. 1539–1548.
[20] A. Rényi, On measures of entropy and information, in: Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability, vol. I, Berkeley, CA, 1961, pp. 547–561.
[21] K. Zografos, S. Nadarajah, Survival exponential entropies, IEEE Transactions on Information Theory 51 (3) (2005).
[22] R.G. Baraniuk, P. Flandrin, A.J.E.M. Janssen, O.J.J. Michel, Measuring time-frequency information content using the Rényi entropies, IEEE Transactions on Information Theory 47 (4) (2001) 1391–1409.
[23] P.E. Caines, Linear Stochastic Systems, John Wiley and Sons, 1988.
[24] J. Massey, Causality, feedback and directed information, in: Proceedings of the 1990 Symposium on Information Theory and its Applications, 1990, pp. 303–305.
[25] David J. Sakrison, The rate distortion function for a class of sources, Information and Control 15 (1969) 165–195.
[26] M. Gastper, B. Rimoldi, V. Vetterli, To code, or not to code: Lossy source-channel communication revisited, IEEE Transactions on Information Theory 49 (5) (2003) 1147–1158.
[27] D.G. Luenberger, Optimization by Vector Space Methods, John Wiley, 1969.
[28] T. Linder, R. Zamir, On the asymptotic tightness of the Shannon lower bound, IEEE Transactions on Information Theory 40 (6) (1994) 2026–2031.