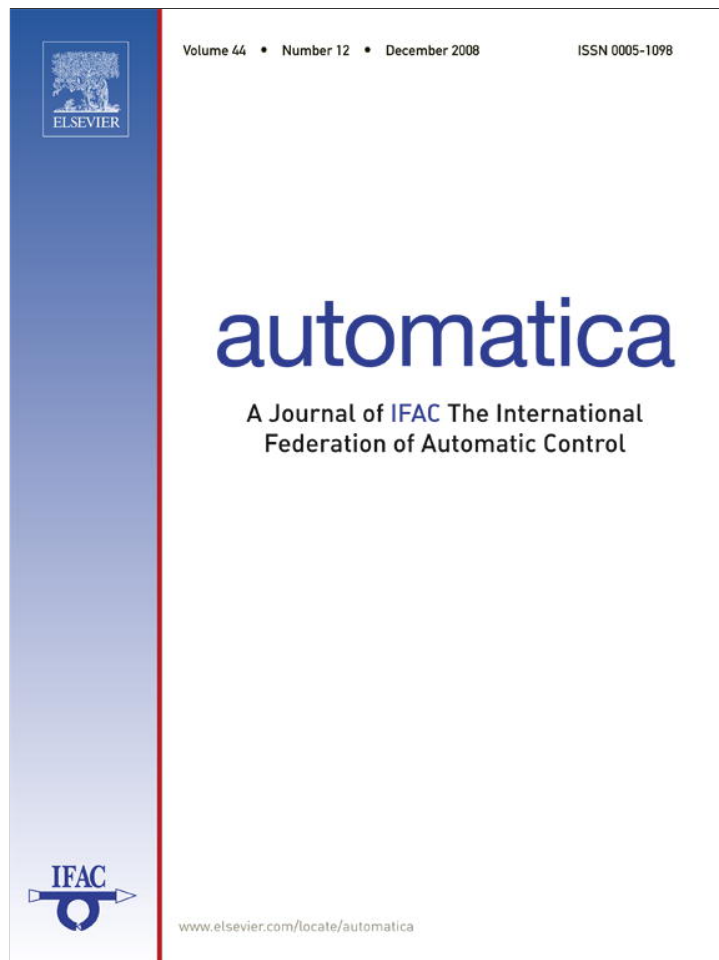


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

LQG optimality and separation principle for general discrete time partially observed stochastic systems over finite capacity communication channels[☆]

Charalambos D. Charalambous^{a,b,*}, Alireza Farhadi^c^a Electrical and Computer Engineering Department, University of Cyprus, 75 Kallipoleos Avenue, Nicosia, Cyprus^b School of Information Technology and Engineering, University of Ottawa, 161 Louis Pasteur, A519, Ottawa, Ontario K1N 6N5, Canada^c School of Information Technology and Engineering, University of Ottawa, 800 King Edward Ave., Ottawa, Ontario, K1N 6N5, Canada

ARTICLE INFO

Article history:

Received 14 June 2007

Received in revised form

16 May 2008

Accepted 16 May 2008

Available online 7 November 2008

Keywords:

Networked

Capacity

Rate distortion

Stability

Separation

ABSTRACT

This paper is concerned with control of stochastic systems subject to finite communication channel capacity. Necessary conditions for reconstruction and stability of system outputs are derived using the Information Transmission theorem and the Shannon lower bound. These conditions are expressed in terms of the Shannon entropy rate and the distortion measure employed to describe reconstruction and stability. The methodology is general, and hence it is applicable to a variety of systems. The results are applied to linear partially observed stochastic Gaussian controlled systems, when the channel is an Additive White Gaussian Noise (AWGN) channel. For such systems and channels, sufficient conditions are also derived by first showing that the Shannon lower bound is exactly equal to the rate distortion function, and then designing the encoder, decoder and controller which achieve the capacity of the channel. The conditions imposed are the standard detectability and stabilizability of Linear Quadratic Gaussian (LQG) theory, while a separation principle is shown between the design of the control and communication systems, without assuming knowledge of the control sequence at the encoder/decoder.

Crown Copyright © 2008 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Control of dynamical systems subject to finite communication channel capacity are often represented by the block diagram of Fig. 1. Due to finite rate constraint of the communication channel, the main underlying assumption is that the plant (dynamical system) output is reproduced at the communication channel output via quantization, hence the input to the controller is a distorted version of the plant output.

Systems described by Fig. 1 are investigated in the literature via a variety of methods. The analysis includes necessary and sufficient conditions for stability and reconstruction (also known as observability in the literature) of unstable control systems subject to limited channel capacity (Elia, 2004; Li & Baillieul, 2004; Liberzon & Hespanha, 2005; Malyavej & Savkin, 2005;

Nair, Dey, & Evans, 2003; Nair, Evans, Mareels, & Moran, 2004; Savkin & Petersen, 2003; Tatikonda & Mitter, 2004a,b; Tatikonda, Sahai, & Mitter, 2004; Tsumura & Maciejowski, 2003). Necessary conditions are often described via the relation between the capacity of the channel and the information rate of the output of the dynamical system. This is reminiscent of the Information Transmission theorem (converse) of information theory, 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), for reliable data reconstruction. On the other hand, sufficient conditions are concerned with the existence and construction of the encoder, decoder and controller for a specific communication channel (e.g., direct part of the Information Transmission theorem).

The objective of this paper is twofold. (1) To derive necessary conditions for reconstruction and stability of sequences in r -mean and probability for general systems, which depend on the entropy rate of the input to the encoder and the type of reconstruction and stability criteria employed. Using these general necessary conditions, some of the necessary conditions found in the literature (Elia, 2004; Li & Baillieul, 2004; Liberzon & Hespanha, 2005; Nair et al., 2004; Tatikonda & Mitter, 2004a,b; Tatikonda et al., 2004; Tsumura & Maciejowski, 2003) are obtained as a special case. (2) To show that under certain conditions, the necessary conditions are also sufficient, when applied to a linear stochastic partially observed controlled systems, and

[☆] This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor George Yin under the direction of Editor Ian R. Petersen. This work is supported by the European Commission under the project Net-Refound and the Cyprus Research Promotion Foundation project ARTEMIS.

* Corresponding author at: Electrical and Computer Engineering Department, University of Cyprus, 75 Kallipoleos Avenue, Nicosia, Cyprus.

E-mail addresses: chadcha@ucy.ac.cy (C.D. Charalambous), farhadialireza@yahoo.com (A. Farhadi).

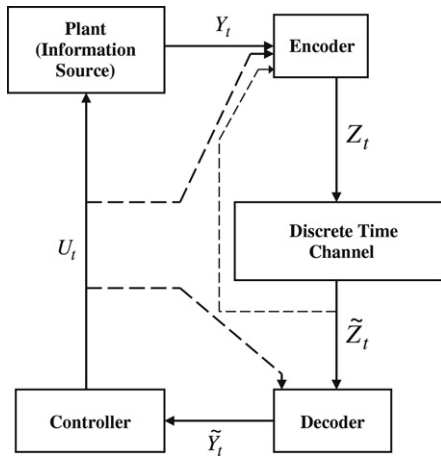


Fig. 1. Control/communication system.

Additive White Gaussian Noise (AWGN) channels. Here, we derive an encoder, decoder, and controller for mean-square stability and reconstruction, using the standard assumptions of detectability and stabilizability of Linear Quadratic Gaussian (LQG) theory (Caines, 1988), without assuming knowledge of the control sequence at the encoder/decoder. Thus, a separation principle is shown to hold between the design of the control and communication systems. This implies that reconstruction of the encoder input at the output of the decoder for a given distortion can be done by assuming the control is zero, while system stability and optimality of the control with respect to Quadratic cost can be achieved using the decoder output, without compromising the optimality of the overall system.

The methodology put forward is information theoretic, while stability and reconstruction are defined for sequences with respect to probability and r -mean. The material discussed under (1) and (2) are new and to the best of our knowledge they have not appeared elsewhere.

The paper is structured as follows. In Section 2, the definitions of various system components are described. In Section 3 some preliminary material on entropy rate and its relations to Kalman filter are introduced, which are used in subsequent sections. In Section 4, the main mathematical tools which are the Information Transmission theorem and the Shannon lower bound are given. Subsequently, general necessary conditions are derived for reconstruction and stability of the control/communication system of Fig. 1. Furthermore, sufficient conditions are derived for linear partially observed controlled Gaussian systems and AWGN channels, subject to the standard LQG assumptions. The importance of the Shannon lower bound is also discussed.

2. Problem formulation

Consider the control/communication system of Fig. 1, where $Y_t \in \mathcal{Y}_t$, $Z_t \in \mathcal{Z}_t$, $\tilde{Z}_t \in \tilde{\mathcal{Z}}_t$, $\tilde{Y}_t \in \tilde{\mathcal{Y}}_t$, $U_t \in \mathcal{U}_t$ are Random Variables (R.V.'s) denoting the source message, channel input, channel output, reconstruction of the source message, and the control input to the source, respectively, at time $t \in \mathbf{N}_+ \triangleq \{0, 1, 2, \dots\}$. It is assumed that \mathcal{Y}_t , \mathcal{Z}_t , $\tilde{\mathcal{Z}}_t$, $\tilde{\mathcal{Y}}_t$, \mathcal{U}_t are finite-dimensional metric spaces. Throughout the paper, for $T \in \mathbf{N}_+$ sequences of R.V.'s are denoted by $\Theta^T \triangleq (\Theta_0, \Theta_1, \dots, \Theta_T)$, while specific realizations by θ^T . $\log(\cdot)$ and $\log_e(\cdot)$ denote logarithm of base 2 and natural logarithm, respectively. Given the measurable spaces $(\bar{\mathcal{A}}, \bar{\mathcal{A}})$, $(\hat{\mathcal{A}}, \hat{\mathcal{A}})$, a stochastic kernel denoted by $P(dF; x)$ is a mapping $P : \hat{\mathcal{A}} \times \bar{\mathcal{A}} \rightarrow [0, 1]$ which satisfies (i) for every $x \in \bar{\mathcal{A}}$, the set function $P(\cdot; x)$ is a probability measure on $\hat{\mathcal{A}}$, and (ii) for every

$F \in \hat{\mathcal{A}}$, the function $P(dF; \cdot)$ is a $\bar{\mathcal{A}}$ -measurable. I_q denotes the identity matrix with dimension $q \times q$, and A^T denotes the transpose of the vector/matrix A , while $\text{Cov}(\cdot)$ is used for the covariance.

The different blocks of Fig. 1 are described below.

Information source: The information source is the plant output, defined by the probability measure $P(Y^T \in dy^T)$ on $\times_{k=0}^T \mathcal{Y}_k$, which depends on the control sequence as shown in Fig. 1. When this measure is absolutely continuous with respect to the Lebesgue measure then $P(Y^T \in dy^T) = f(y^T)dy^T$, where $f(\cdot)$ is the probability density function of Y^T .

Communication channel: The communication channels with input Z^n and output \tilde{Z}^n is modeled by a stochastic kernel $\{P(d\tilde{z}_t; z^t, \tilde{z}^{t-1}); t \in \mathbf{N}_+\}$, that is, $P(d\tilde{z}_t; z^t, \tilde{z}^{t-1}) = P(\tilde{Z}_t \in d\tilde{z}_t | Z^t = z^t, \tilde{Z}^{t-1} = \tilde{z}^{t-1})$ are conditional probability measures on $\tilde{\mathcal{Z}}_t$ given $Z^t = z^t$ and $\tilde{Z}^{t-1} = \tilde{z}^{t-1}$.

A communication channel is called memoryless if the channel stochastic kernel satisfies $P(d\tilde{z}_t; z^t, \tilde{z}^{t-1}) = P(d\tilde{z}_t; z_t)$.

A communication channel is used without feedback if the input symbol does not depend on the previous output symbols; probabilistically, it implies $P(dz_t; z^{t-1}, \tilde{z}^{t-1}) = P(dz_t; z^{t-1})$.

A communication channel in which the channel input/output are restricted to finite alphabet sets and the channel is memoryless, is called a Discrete Memoryless Channel (DMC).

An Additive White Gaussian Noise (AWGN) channel is described by $\tilde{Z}_t = Z_t + \tilde{W}_t$, where the orthogonal process $\{\tilde{W}_t\}_{t \in \mathbf{N}_+}$ is a zero mean Gaussian process.

Encoder: For any time $t \in \mathbf{N}_+$, Class A, Class B, and Class C encoders are modeled by stochastic kernels, $P(dz_t; y^t)$, $P(dz_t; y^t, u^{t-1}, \tilde{z}^{t-1})$, and $P(dz_t; y^t, \tilde{z}^{t-1})$, respectively.

Decoder: For any time $t \in \mathbf{N}_+$, Class A and Class B decoders are modeled by stochastic kernels $P(d\tilde{y}_t; \tilde{z}^t)$ and $P(d\tilde{y}_t; \tilde{z}^t, u^{t-1})$, respectively.

Controller: For any time $t \in \mathbf{N}_+$, Class A and Class B control laws are modeled by a stochastic kernels $P(du_t; \tilde{z}^{t-1}, u^{t-1})$ and $P(du_t; \tilde{z}^{t-1})$, respectively.

Reconstruction and stability of sequences associated with the system of Fig. 1 are defined next.

Definition 2.1. Consider the system of Fig. 1.

A (Reconstruction in Probability and r -Mean). (i) A sequence Θ^{T-1} related to the plant is said to be reconstructed in probability with another sequence $\tilde{\Theta}^{T-1}$, if for a given $\delta \geq 0$, there exist (a control sequence and) an encoder and a decoder such that $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E \rho(\Theta_t, \tilde{\Theta}_t) \leq D_v$, where $\rho(\Theta, \tilde{\Theta}) = 1$ if $\|\Theta - \tilde{\Theta}\| > \delta$ and $\rho(\Theta, \tilde{\Theta}) = 0$ if $\|\Theta - \tilde{\Theta}\| \leq \delta$. ($\|\cdot\|$ is the norm on the product space) and $D_v \in [0, 1)$.

(ii) A sequence Θ^{T-1} related to the plant is said to be reconstructed in r -mean with another sequence $\tilde{\Theta}^{T-1}$, if $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E \|\Theta_t - \tilde{\Theta}_t\|^r \leq D_v$, $r > 0$, for a finite $D_v \geq 0$.

B (Stability in Probability and r -Mean). (i) A sequence H^{T-1} related to the plant is said to be stable in probability if for a given $\delta \geq 0$, there exist a controller, encoder, and decoder such that $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E \rho(H_t, 0) \leq D_v$, where $\rho(\cdot, \cdot)$ is defined as above and $D_v \in [0, 1)$.

(ii) A sequence H^{T-1} related to the plant is said to be stable in r -mean if $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E \|H_t - 0\|^r \leq D_v$, $r > 0$, for a finite $D_v \geq 0$.

By Fig. 1, source reconstruction at the output of the decoder corresponds to $\Theta^{T-1} = Y^{T-1}$, $\tilde{\Theta}^{T-1} = \tilde{Y}^{T-1}$. Note that reconstruction in a communication system is analogous to state estimation in a control system.

The main results of the paper are applied throughout the paper to the following system.

Stochastic control system: The system dynamics which are defined on a complete probability space $(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t \geq 0})$ are described by

$$\begin{cases} X_{t+1} = AX_t + NU_t + BW_t, & X_0 = X, \\ Y_t = H_t + DG_t, & H_t = CX_t, \end{cases} \quad (1)$$

where $X : \Omega \times \mathbf{N}_+ \rightarrow \mathfrak{R}^q$ is the unobserved (state) process, $Y : \Omega \times \mathbf{N}_+ \rightarrow \mathfrak{R}^d$ is the observed (measurement) process, $U : \Omega \times \mathbf{N}_+ \rightarrow \mathfrak{R}^n$ is the control signal, $H : \Omega \times \mathbf{N}_+ \rightarrow \mathfrak{R}^d$ is the signal to be controlled, $W : \Omega \times \mathbf{N}_+ \rightarrow \mathfrak{R}^m$, $G : \Omega \times \mathbf{N}_+ \rightarrow \mathfrak{R}^l$ in which $\{W_t; t \in \mathbf{N}_+\}$ is Independent Identically Distributed (i.i.d.) $\sim N(0, I_m)$ and $\{G_t; t \in \mathbf{N}_+\}$ is i.i.d. $\sim N(0, I_l)$. Moreover, $X_0 \sim N(\bar{x}_0, V_0)$ and $\{W_t, G_t, X_0; t \in \mathbf{N}_+\}$ are mutually independent.

3. Shannon entropy rate, Kalman filtering and innovations process

In this section, we introduce Shannon's entropy (Cover & Thomas, 1991) and we discuss its implications in controlling a dynamical system over limited capacity communication channels. Suppose the source sequence Y^T produces messages at a rate of one message every τ_s seconds (e.g., the time between subsequent sampling times Y_j and Y_{j+1} is τ_s seconds). If the information source has joint density $P(dy^T) = f(y^T)dy^T$, then its Shannon entropy is defined by $H_S(Y^T) \triangleq -\int f(y^T) \log f(y^T) dy^T$ bits (Cover & Thomas, 1991). Subsequently, the Shannon entropy rate is given by $\mathcal{H}_S(\mathcal{Y}) \triangleq \lim_{T \rightarrow \infty} \frac{1}{T} H_S(Y^{T-1})$ bits per source message, provided the limit exists; or the source generates $\mathcal{H}_S(\mathcal{Y})/\tau_s$ information bits per second. In Section 4, it will be shown that the Shannon entropy rate is related to the minimum bit rate for reliable data reconstruction and stability of a sequence associated with a controlled source. Here, we discuss connections between the error covariance of the Kalman filter, the innovations process of the source, Shannon entropy rate, and unstable eigenvalues of linear dynamical systems.

Lemma 3.1 (See, Caines (1988, pp. 44)). Let $\{Y_t; t \in \mathbf{N}_+\}$, $Y : \Omega \times \mathbf{N}_+ \rightarrow \mathfrak{R}^d$ be a Gaussian process and define $\Gamma_{Y^{T-1}} \triangleq \text{Cov}[(Y_0^T, Y_1^T, \dots, Y_{T-1}^T)^T]$ and $K_t \triangleq Y_t - E[Y_t | \sigma\{Y^{t-1}\}]$, $A_t \triangleq \text{Cov}(K_t)$, where $\sigma\{Y^{t-1}\}$ denotes the σ -algebra of events generated by the sequence Y^{t-1} . Assume $\Lambda_\infty \triangleq \lim_{T \rightarrow \infty} \Lambda_T$ exists. Then, the Shannon entropy rate of $\{Y_t; t \in \mathbf{N}_+\}$ in bits per source message is

$$\begin{aligned} \mathcal{H}_S(\mathcal{Y}) &\triangleq \lim_{T \rightarrow \infty} \frac{1}{T} H_S(Y^{T-1}) = \frac{d}{2} \log(2\pi e) + \lim_{T \rightarrow \infty} \frac{1}{2T} \log \det \Gamma_{Y^{T-1}} \\ &= \frac{d}{2} \log(2\pi e) + \frac{1}{2} \log \det \Lambda_\infty. \end{aligned} \quad (2)$$

Moreover, if $\{Y_t; t \in \mathbf{N}_+\}$ is an asymptotic stationary Gaussian process with power spectral density $S_Y(e^{j\omega})$ then $\mathcal{H}_S(\mathcal{Y}) = \frac{d}{2} \log(2\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \det S_Y(e^{j\omega}) d\omega$.

The application of the above lemma to partially observed systems establishes a relation between the entropy rate and the unstable eigenvalues rate.

3.1. Uncontrolled stochastic dynamical systems

Sufficient conditions for existence of the limit, $\Lambda_\infty = \lim_{T \rightarrow \infty} \Lambda_T$ are given next.

Lemma 3.2. Consider the uncontrolled version of system (1) corresponding to $\{U_t = 0; t \in \mathbf{N}_+\}$, and assume (C, A) is detectable, $(A, (BB^T)^{\frac{1}{2}})$ is stabilizable, and $D \neq 0$.

Then, $\Lambda_\infty = CV_\infty C^T + DD^T$, where $V_t \triangleq E[\tilde{X}_t \tilde{X}_t^T]$, $\tilde{X}_t \triangleq X_t - E[X_t | \sigma\{Y^{t-1}\}]$, $t \in \mathbf{N}_+$, and $V_\infty \triangleq \lim_{T \rightarrow \infty} V_T$ is the unique solution of the following Algebraic Riccati-equation

$$\begin{aligned} V_\infty &= AV_\infty A^T - AV_\infty C^T (CV_\infty C^T + DD^T)^{-1} CV_\infty A^T + BB^T, \\ V_\infty &\geq 0. \end{aligned} \quad (3)$$

Moreover, $\mathcal{H}_S(\mathcal{K}) = \mathcal{H}_S(\mathcal{Y}) = \frac{d}{2} \log(2\pi e) + \frac{1}{2} \log \det(CV_\infty C^T + DD^T)$, where $\mathcal{H}_S(\mathcal{K})$ is the entropy rate of the innovations process.

Proof. Follows from Caines (1988, pp. 156–158) and Lemma 3.1. \square

In the following remark we relate Shannon entropy rate to unstable eigenvalues, e.g., $\lambda_i(A)$'s such that $|\lambda_i(A)| \geq 1$, where $\lambda_i(A)$'s denote the eigenvalues of the matrix A .

Remark 3.3. (i) Consider the scalar version of the uncontrolled system (1) with $q = 1$ and $d = 1$. Then, (3) can be solved explicitly to obtain

$$\begin{aligned} \mathcal{H}_S(\mathcal{Y}) &= \frac{1}{2} \log(2\pi e) + \frac{1}{2} \log \Lambda_\infty \geq \frac{1}{2} \log(2\pi e D^2) \\ &+ \max\{0, \log |A|\}. \end{aligned} \quad (4)$$

Moreover, the inequality in (4) holds with equality when $B = 0$.

(ii) Consider the uncontrolled version of system (1) with scalar observation and measurement noise (i.e., $d = l = 1$) under detectability and stabilizability conditions of Lemma 3.2 when $D \neq 0$. Then, $\mathcal{H}_S(\mathcal{Y}) = \mathcal{H}_S(\mathcal{K}) = \frac{1}{2} \log(2\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log S_K(e^{j\omega}) d\omega$, where $S_K(e^{j\omega})$ is the power spectral density of the asymptotic stationary innovations process $\{K_t = Y_t - E[Y_t | \sigma\{Y^{t-1}\}]; t \in \mathbf{N}_+\}$. From standard Kalman filtering equations (Caines, 1988, pp. 156–158) it follows that $K(z) = M(z)W(z) + S(z)DG(z)$, where $M(z) = C(zI_q - \tilde{A})^{-1}B$, $\tilde{A} = A - \Delta_\infty C$ (it has eigenvalues inside the unit circle), $\Delta_\infty = AV_\infty C^T (CV_\infty C^T + DD^T)^{-1}$, $S(z) = 1 - C(zI_q - \tilde{A})^{-1} \Delta_\infty$ and $K(z), W(z), G(z)$ are the z -transformation of the innovations, process and measurement noises, respectively. Subsequently, $S_K(e^{j\omega}) = M(e^{j\omega})M^T(e^{-j\omega}) + |S(e^{j\omega})|^2 D^2 \geq |S(e^{j\omega})|^2 D^2$. Consequently, $\mathcal{H}_S(\mathcal{Y}) = \mathcal{H}_S(\mathcal{K}) \geq \frac{1}{2} \log(2\pi e) + \frac{1}{2} \log D^2 + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log |S(e^{j\omega})|^2 d\omega$. Next, consider $S(z)$ as the sensitivity transfer function of a stable unit negative feedback system. That is, $S(z) = 1 - C(zI_q - \tilde{A})^{-1} \Delta_\infty = 1 - T(z)$, where $T(z) = C(zI_q - \tilde{A})^{-1} \Delta_\infty$ is the complementary sensitivity function (closed loop transfer function) of this stable feedback system. In state space form, this system is represented by $X_{t+1} = AX_t + \Delta_\infty U_t$, $U_t = -K_t$, $K_t = CX_t + DG_t$. That is, $S(z) \triangleq \frac{K(z)}{DG(z)} = \frac{1}{1+L(z)}$, where $L(z) = C(zI_q - A)^{-1} \Delta_\infty$ is the transfer function of the open loop system $X_{t+1} = AX_t + \Delta_\infty U_t$, $K_t = CX_t$. Using the Bode integral formula (Wu & Jonckheere, 1992), we have $\int_{-\pi}^{\pi} |S(e^{j\omega})|^2 d\omega = 4\pi \sum_{\{i; |\lambda_i(A)| \geq 1\}} \log |\lambda_i(A)|$, and therefore $\mathcal{H}_S(\mathcal{Y}) = \mathcal{H}_S(\mathcal{K}) \geq \frac{1}{2} \log(2\pi e D^2) + \sum_{\{i; |\lambda_i(A)| \geq 1\}} \log |\lambda_i(A)|$. This eigenvalue rate is also found in many of the cited references using different methods.

3.2. Controlled stochastic dynamical systems

Next, we will show that the Shannon entropy rate of the controlled system is bounded below by the Shannon entropy rate of the uncontrolled system.

Corollary 3.4. Consider the controlled system (1) and assume (C, A) is detectable, $(A, (BB^T)^{\frac{1}{2}})$ is stabilizable, and $D \neq 0$. Then, the Shannon entropy rate of the controlled system (1) is bounded below

by that of the uncontrolled system via

$$\begin{aligned} \mathcal{H}_S(\mathcal{Y}) &= \lim_{T \rightarrow \infty} \frac{1}{T} H_S(Y^{T-1}) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=0}^{T-1} H_S(Y_i | Y^{i-1}) \\ &\geq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=0}^{T-1} H_S(Y_i | Y^{i-1}, U^{i-1}) = \frac{d}{2} \log(2\pi e) \\ &\quad + \frac{1}{2} \log \det \Lambda_\infty, \end{aligned} \quad (5)$$

where Λ_∞ is given in Lemma 3.2.

Proof. The second equality follows from the chain of the Shannon entropy (Cover & Thomas, 1991), the inequality follows from the property that conditioning reduces entropy, and the last equality follows from the Gaussianity of the source. \square

4. Information transmission theorem, Shannon lower bound, and conditions for reconstruction and stability

In this section, we first recall the definitions of information channel capacity, rate distortion, and Shannon lower bound (Berger, 1971; Linder & Zamir, 1994). Then we state general necessary conditions for reconstruction and stability in terms of the Shannon entropy rate and the type of fidelity criteria for reconstruction and stability considered. Then we derive sufficient conditions for mean square reconstruction of the innovations process and stability of the controlled system, when the communication channel is an AWGN channel. Moreover, we present an encoder and decoder so that the source is matched to the channel capacity for partially observed systems, and we derive a separation principle between the design of the control and the communication systems.

4.1. Information transmission theorem and the Shannon lower bound

First, we recall the definitions of Shannon information capacity formula and rate distortion function, for feedback channels and sources with feedback, respectively, so that these definitions are applicable to the system of Fig. 1. Throughout we assume that the system of Fig. 1, specifically, $Y^{T-1} \rightarrow Z^{n-1} \rightarrow \tilde{Z}^{n-1} \rightarrow \tilde{Y}^{T-1} \rightarrow U^{T-1}$ forms a Markov chain, which is equivalent to the conditional independence. In particular, for any $T, n \in \mathbf{N}_+$, the Markovian property of the source output sequence Y^{T-1} , the channel input sequence Z^{n-1} and the channel output sequence \tilde{Z}^{n-1} is equivalent to the conditional independence $P(d\tilde{z}_t; z^t, \tilde{z}^{t-1}, y^{t-1}) = P(d\tilde{z}_t; z^t, \tilde{z}^{t-1})$, P -a.s ($t \in \{0, 1, \dots, T-1\}$), and similarly for the rest of the blocks.

Moreover, a degraded channel connecting Z^{n-1} to \tilde{Z}^{n-1} , denoted by $Z^{n-1} \rightarrow \tilde{Z}^{n-1}$, is defined by $P(d\tilde{z}^{n-1}; z^{n-1}) \triangleq \prod_{i=0}^{n-1} P(d\tilde{z}_i; z^i, \tilde{z}^{i-1})$. This is equivalent to a causal restriction of the conditional distribution; it is a definition and not an assumption. Note that if it were an assumption, in view of Bayes rule, $P(d\tilde{z}^{n-1}; z^{n-1}) = \prod_{i=0}^{n-1} P(d\tilde{z}_i; z^n, \tilde{z}^{i-1}) = \prod_{i=0}^{n-1} P(d\tilde{z}_i; z^i, \tilde{z}^{i-1})$, which implies that \tilde{Z}_i and Z_{i+1}, \dots, Z_{n-1} are conditional independent given Z^i, \tilde{Z}^{i-1} for all $i \in \mathbf{N}_+$. Similarly, a degraded channel connecting \tilde{Z}^{n-1} to Z^{n-1} , denoted by $Z^{n-1} \leftarrow \tilde{Z}^{n-1}$, is defined by $P(dz^{n-1}; \tilde{z}^{n-1}) \triangleq \prod_{i=0}^{n-1} P(dz_i; z^{i-1}, \tilde{z}^{i-1})$. The motivation for introducing these definitions will become obvious shortly.

Recall Shannon's self-mutual information of two sequences defined by $i(z^{n-1}; \tilde{z}^{n-1}) \triangleq \log \frac{P(d\tilde{z}^{n-1}; z^{n-1})}{P(d\tilde{z}^{n-1})}$, and its average,

Shannons mutual information defined by $I(Z^{n-1}; \tilde{Z}^{n-1}) \triangleq E_{P(dz^{n-1}, d\tilde{z}^{n-1})} \{i(z^{n-1}; \tilde{z}^{n-1})\}$ (Cover & Thomas, 1991). It can be easily shown that

$$i(z^{n-1}; \tilde{z}^{n-1}) = \sum_{i=0}^{n-1} \log \frac{P(\tilde{z}_i; \tilde{z}^{i-1}, z^i)}{P(\tilde{z}_i; \tilde{z}^{i-1})} + \sum_{i=0}^{T-1} \log \frac{P(z_i; z^{i-1}, \tilde{z}^{i-1})}{P(z_i; z^{i-1})}$$

and by taking expectation

$$I(Z^{n-1}; \tilde{Z}^{n-1}) = I(Z^{n-1} \rightarrow \tilde{Z}^{n-1}) + I(Z^{n-1} \leftarrow \tilde{Z}^{n-1})$$

where

$$\begin{aligned} I(Z^{n-1} \rightarrow \tilde{Z}^{n-1}) &= \sum_{i=0}^{n-1} I(Z^i; \tilde{Z}_i | \tilde{Z}^{i-1}) \\ &= \sum_{i=0}^{n-1} \int \log \frac{P(d\tilde{z}_i; \tilde{z}^{i-1}, z^i)}{P(d\tilde{z}_i; \tilde{z}^{i-1})} P(d\tilde{z}^i, dz^i) \\ I(Z^{n-1} \leftarrow \tilde{Z}^{n-1}) &= \sum_{i=0}^{n-1} I(\tilde{Z}^{i-1}; Z_i | Z^{i-1}) \\ &= \sum_{i=0}^{n-1} \int \log \frac{P(dz_i; z^{i-1}, \tilde{z}^{i-1})}{P(dz_i; z^{i-1})} P(d\tilde{z}^{i-1}, dz^i). \end{aligned}$$

Note that the term $I(Z^{n-1} \rightarrow \tilde{Z}^{n-1})$ is the directed information from Z^{n-1} to \tilde{Z}^{n-1} discussed in Massey (1990), and corresponds to the Shannon mutual information restricted to the degraded channel $Z^{n-1} \rightarrow \tilde{Z}^{n-1}$, while $I(Z^{n-1} \leftarrow \tilde{Z}^{n-1})$ is the directed information from \tilde{Z}^{n-1} to Z^{n-1} , which corresponds to the Shannon mutual information restricted to the degraded channel $Z^{n-1} \leftarrow \tilde{Z}^{n-1}$.

Note that $I(Z^{n-1} \leftarrow \tilde{Z}^{n-1}) = 0$ if and only if $P(dz_i; z^{i-1}, \tilde{z}^{i-1}) = P(dz_i; z^{i-1})$, P -a.s. for all $i \in \{0, \dots, n-1\}$, which is equivalent to either of the two statements (1) Z_i and \tilde{Z}^{i-1} are conditional independent given Z^{i-1} , for all $i \in \{0, \dots, n-1\}$, (2) \tilde{Z}_i and Z_{i+1}, \dots, Z_{n-1} are conditional independent given Z^i, \tilde{Z}^{i-1} , for all $i \in \{0, \dots, n-1\}$. A channel with feedback always satisfies $I(Z^{n-1} \leftarrow \tilde{Z}^{n-1}) \geq 0$. On the other hand, if a channel has no feedback then by definition of feedback encoding we have $I(Z^{n-1} \leftarrow \tilde{Z}^{n-1}) = 0$, and $I(Z^{n-1}; \tilde{Z}^{n-1}) = I(Z^{n-1} \rightarrow \tilde{Z}^{n-1})$. Hence, for a DMC or a memoryless AWGN channel we have $I(Z^{n-1}; \tilde{Z}^{n-1}) \geq I(Z^{n-1} \rightarrow \tilde{Z}^{n-1}) = \sum_{i=0}^{n-1} \int \log \frac{P(d\tilde{z}_i; z_i)}{P(d\tilde{z}_i; \tilde{z}^{i-1})} P(d\tilde{z}^i, dz^i)$, with equality if and only if Z_i and \tilde{Z}^{i-1} are conditional independent given Z^{i-1} , for all $i \in \{0, \dots, n-1\}$. It can be verified that the channels with feedback discussed in Cover and Thomas (1991) is the degraded channel connecting Z^{n-1} to \tilde{Z}^{n-1} , which does not consider the term $I(Z^{n-1} \leftarrow \tilde{Z}^{n-1})$. The point to be made here is that when the channel has memory and feedback then the work found in the literature (see Cover and Thomas (1991)) consider only the directed information $I(Z^{n-1} \rightarrow \tilde{Z}^{n-1})$, in their general theorems of DMC or AWGN channels with feedback. Similarly, when discussing source reconstruction at the output of the decoder, we use the definition of a degraded stochastic kernel connecting Y^{T-1} to \tilde{Y}^{T-1} , that is, the reconstruction kernel is defined by $P(d\tilde{y}^{T-1}; y^{T-1}) \triangleq \prod_{i=0}^{T-1} P(d\tilde{y}_i; \tilde{y}^{i-1}, y^i)$ to account for the direction $Y^{T-1} \rightarrow \tilde{Y}^{T-1}$.

Definition 4.1 (Information Capacity). Consider a communication channel and let Z^{n-1} and \tilde{Z}^{n-1} be the channel input and output sequences, respectively. Let \mathcal{M}_C denotes the set of channel inputs joint probability measures, $P(dz^{n-1})$, which satisfy certain channel input power constraint. The Shannon information capacity for the time horizon n is defined by

$$\mathcal{C}_n^{Z, \tilde{Z}} \triangleq \sup_{P(dz^{n-1}) \in \mathcal{M}_C} I(Z^{n-1}; \tilde{Z}^{n-1}). \quad (6)$$

The information capacity in bits per channel use is $\mathcal{C}^{Z, \tilde{Z}} \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{C}_n^{Z, \tilde{Z}}$ provided the limit exists.

Note that for channels with memory and feedback the mutual information in (6) should be replaced by $I(Z^{n-1} \rightarrow \tilde{Z}^{n-1})$, which is the mutual information subject to the degraded channel restriction connecting Z^{n-1} to \tilde{Z}^{n-1} . The channel coding theorem states that the information capacity is operational, that is, any rate $R < \mathcal{C}^{Z,\tilde{Z}}$ is achievable, in the sense that there exist a channel encoder and decoder so that the channel can transmit 2^{nR} messages, in n uses of the channel, such that the probability of the decoding error is arbitrary small as the number of channel uses n is chosen arbitrary large. Assuming the channel has capacity $\mathcal{C}^{Z,\tilde{Z}}$ bits per channel use and it is to be used at a rate of one channel symbol in each τ_c seconds, then the channel has capacity $\mathcal{C}^{Z,\tilde{Z}}/\tau_c$ bits per second. For DMC's and AWGN channels without feedback, the information channel capacity of Definition 4.1 represents the operational capacity. Further, the capacity of DMC's or AWGN channels with feedback which is described by the directed information connecting channel inputs to channel outputs, is the same as the capacity of those channels without feedback (Cover & Thomas, 1991).

Example 4.2. AWGN channel: Consider the following AWGN channel without feedback $\tilde{Z}_t = Z_t + \tilde{W}_t$, where $\tilde{W}_t \in \mathfrak{R} \sim N(0, W_c)$ is a zero mean orthogonal Gaussian noise process. At each time instant $t \in \mathbf{N}_+$, this channel transmits the encoded information under transmission power constraint $E[Z_t^2] \leq P_t$. The capacity of this AWGN channel for n channel uses is $\mathcal{C}_n^{Z,\tilde{Z}} = \sum_{t=0}^{n-1} \frac{1}{2} \log(1 + \frac{P_t}{W_c})$ bits in n channel uses; and subsequently when $\lim_{n \rightarrow \infty} P_n = P$, $\mathcal{C}^{Z,\tilde{Z}} = \frac{1}{2} \log(1 + \frac{P}{W_c})$ bits per channel use. Moreover, if at each time instant $t \in \mathbf{N}_+$, this channel transmits a source message, the capacity in bits per source message is also $\mathcal{C}^{Z,\tilde{Z}} = \frac{1}{2} \log(1 + \frac{P}{W_c})$. When the channel has feedback and mutual information is replaced by directed information $I(Z^{n-1} \rightarrow \tilde{Z}^{n-1})$, the capacity is the same.

The next definition gives an information meaning to lossy data reconstruction or quantization, in which the source messages can be represented by another messages up to a given distortion.

Definition 4.3 (Information Rate Distortion). Let Y^{T-1} and \tilde{Y}^{T-1} be sequences of length T of the source and the reconstruction of the source messages, respectively, and $\mathcal{M}_{DC} \triangleq \{P(d\tilde{y}^{T-1}; y^{T-1}); E_{\rho_T}(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v\}$ denote the set of distortion constraints in which $D_v \geq 0$ is the distortion level and $\rho_T \in [0, \infty)$ is the distortion measure. The information rate distortion for time horizon T is defined by:

$$R_T^{Y,\tilde{Y}}(D_v) \triangleq \inf_{P(d\tilde{y}^{T-1}; y^{T-1}) \in \mathcal{M}_{DC}} I(Y^{T-1}; \tilde{Y}^{T-1}). \quad (7)$$

The information rate distortion in bits per source message is $R^{Y,\tilde{Y}}(D_v) \triangleq \lim_{T \rightarrow \infty} \frac{1}{T} R_T^{Y,\tilde{Y}}(D_v)$ provided the limit exists.

Similar to capacity, for sources with memory and feedback, (7) should be replaced by $R_{T,SRD}^{Y,\tilde{Y}}(D_v) \triangleq \inf_{\{P(d\tilde{y}_i; \tilde{y}^{i-1}, y^i)\}_{i=0}^{T-1} \in \mathcal{M}_{DC}} I(Y^{T-1} \rightarrow \tilde{Y}^{T-1})$, that is, the infimum is taken with respect to the causal kernels, and mutual information is replaced by directed information. The source coding theorem is concerned with the operational meaning of information rate distortion (Cover & Thomas, 1991) (in one direction when the source has feedback). In the rest of the paper we shall assume, without loss of generality that $\tau_s = \tau_c$. For memoryless sources the information rate distortion represents the operational definition (Cover & Thomas, 1991).

In the next remark we identify a sufficient condition so that the minimizing stochastic kernel in (7) gives an optimal

reconstruction kernel which is a causal operation on the source sequence (Tatikonda et al., 2004).

Remark 4.4. Suppose the distortion measure is given by $\rho_T(k^{T-1}, \tilde{k}^{T-1}) = \frac{1}{T} \sum_{t=0}^{T-1} \rho(k_t, \tilde{k}_t)$, where $\rho(k_t, \tilde{k}_t) : \mathcal{K}_t \times \tilde{\mathcal{K}}_t \rightarrow [0, \infty)$ is continuous and non-negative and assume the source K^{T-1} is an independent process, without feedback. The minimizing stochastic kernel in (7) is

$$P^*(d\tilde{k}^{T-1}; k^{T-1}) = \prod_{t=0}^{T-1} P^*(d\tilde{k}_t; k_t) = \prod_{t=0}^{T-1} \frac{e^{\frac{1}{T} s(D_v) \rho(k_t, \tilde{k}_t)} P^*(d\tilde{k}_t)}{\int_{\tilde{\mathcal{K}}_t} e^{\frac{1}{T} s(D_v) \rho(k_t, \tilde{k}_t)} P^*(d\tilde{k}_t)}, \quad s(D_v) = \frac{dR_T^{K,\tilde{K}}(D_v)}{dD_v}. \quad (8)$$

Since $P^*(d\tilde{k}^{T-1}; k^{T-1}) = \prod_{t=0}^{T-1} P^*(d\tilde{k}_t; k_t)$ the reproduction kernel is indeed causal for the given source. When the encoder has feedback, the kernel which minimizes $I(K^{T-1} \rightarrow \tilde{K}^{T-1})$ is the same.

Next, we present a necessary condition for end to end transmission up to distortion level $D_v \geq 0$. This is the converse of information transmission theorem.

Theorem 4.5 (Information Transmission Theorem). Consider the control/communication system of Fig. 1 under the conditional independence assumptions introduced earlier. Then

- (i) $I(Z^n; \tilde{Z}^n) \geq I(Z^n \rightarrow \tilde{Z}^n) \geq I(Y^t; \tilde{Y}^t) \geq I(Y^t \rightarrow \tilde{Y}^t)$, $\forall t, n \in \mathbf{N}_+$.
- (ii) A necessary condition for reproducing a sequence of source messages Y^{T-1} up to distortion level D_v by \tilde{Y}^{T-1} at the output of the decoder (e.g., $E_{\rho_T}(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v$) using a sequence of the channel inputs and outputs with length n ($T \leq n$) is

$$\mathcal{C}_n^{Z,\tilde{Z}} \geq R_T^{Y,\tilde{Y}}(D_v). \quad (9)$$

Proof. (i) This is similar to Massey (1990). (ii) Follows from the data processing inequalities. \square

Note that when feedback is assumed, then in the definition of (9), mutual information should be replaced by directed information and the reproduction kernel by causal kernels as discussed earlier. Next, we present the Shannon lower bound which is used to obtain necessary conditions for reconstruction and stability. This lower bound, in view of data processing inequalities of Theorem 4.5, is practical in terms of providing a tight necessary condition.

Lemma 4.6 (Shannon Lower Bound). Let $Y^{T-1}, Y_t \in \mathfrak{R}^d, 0 \leq t \leq T-1$ be a sequence with length T produced by the source $P(\tilde{Y}^{T-1} \in d\tilde{y}^{T-1}) = f(y^{T-1})d\tilde{y}^{T-1}$. Consider the following form of 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 $\rho(y_t, \tilde{y}_t) = \rho(y_t - \tilde{y}_t) : \mathfrak{R}^d \rightarrow [0, \infty)$ is continuous. Then,

- (i) A lower bound for $\frac{1}{T} R_T^{Y,\tilde{Y}}(D_v)$ is given by

$$\frac{1}{T} R_T^{Y,\tilde{Y}}(D_v) \geq \frac{1}{T} H_S(Y^{T-1}) - \max_{h \in G_D} H_S(h), \quad (10)$$

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

$$h^*(\xi) = \frac{e^{s\rho(\xi)}}{\int_{\mathfrak{R}^d} e^{s\rho(\xi)} d\xi}, \quad \int_{\mathfrak{R}^d} \rho(\xi) h^*(\xi) d\xi = D_v. \quad (11)$$

Subsequently, when $R^{Y,\tilde{Y}}(D_v)$ and $\mathcal{H}_S(\mathcal{Y})$ exist, the Shannon lower bound denoted by $R_S^{Y,\tilde{Y}}(D_v)$ is given by

$$R^{Y, \bar{Y}}(D_v) \geq \mathcal{H}_S(\mathcal{Y}) - \max_{h \in \mathcal{G}_D} H_S(h) \triangleq R_S^{Y, \bar{Y}}(D_v). \quad (12)$$

(ii) Suppose the difference distortion measure $\rho(\cdot)$ satisfies the conditions a, b, d of Linder and Zamir (1994, pp. 2029), $\int_{\mathfrak{R}^d} e^{s\rho(\xi)} d\xi < \infty$ for all $s < 0$, $\mathcal{H}_S(\mathcal{Y}) > -\infty$ and there exists an $y^* \in \mathfrak{R}^d$ such that $E\rho(y - y^*) < \infty, \forall y \in \mathfrak{R}^d$.

Then, in the limit as $D_v \rightarrow 0$, the lower bound is asymptotically exact. That is, for the case when $R^{Y, \bar{Y}}(D_v)$ and $\mathcal{H}_S(\mathcal{Y})$ exist, $\lim_{D_v \rightarrow 0} [R^{Y, \bar{Y}}(D_v) - (\mathcal{H}_S(\mathcal{Y}) - H_S(h^*))] = 0$.

Proof. Follows from Linder and Zamir (1994) by considering the method proposed in Berger (1971, pp. 140). \square

Remark 4.7. Note that the above Shannon lower bound is derived by considering rate distortion defined via mutual information, and without assuming causal reproduction kernel. (i) A sufficient condition for the existence of $R^{Y, \bar{Y}}(D_v)$ is stationarity of the source (Berger, 1971).

(ii) For distortion measure $\rho(y, \bar{y}) = \|y - \bar{y}\|^r$, in the limit as $D_v \rightarrow 0$, the Shannon lower bound is equal to the rate distortion function (Linder & Zamir, 1994). Furthermore, for this distortion measure, $\max_{h \in \mathcal{G}_D} H_S(h) = \log e^{\frac{d}{r}} - \log\left(\frac{r}{dV_d \Gamma(\frac{d}{r})} \left(\frac{d}{rD_v}\right)^{\frac{d}{r}}\right)$ bits per source message, where V_d is the volume of the unit sphere and $\Gamma(\cdot)$ is the gamma function (Linder & Zamir, 1994).

4.2. Necessary conditions for reconstruction and stability of general systems

For the general control/communication system of Fig. 1, the main theorem which connects reliable communication (e.g., reconstruction) and stability for general systems is given next.

Theorem 4.8. Consider the system of Fig. 1 under the Markov chain assumption in which $Y^{T-1}, Y_t \in \mathfrak{R}^d$ is the observed process to be reconstructed at the output of the decoder. Assume the Shannon entropy rate corresponding to the Y^{T-1} process exists and it is finite.

For the reconstruction of Y^{T-1} in probability, a necessary condition on the channel capacity is

$$C^{Z, \bar{Z}} \geq \mathcal{H}_S(\mathcal{Y}) - \frac{1}{2} \log[(2\pi e)^d \det \Gamma_g] = R_S^{Y, \bar{Y}}(D_v), \quad (13)$$

where $\mathcal{H}_S(\mathcal{Y})$ is the Shannon entropy rate of the Y^{T-1} process 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 $\int_{\|\xi\| > \delta} h^*(\xi) d\xi = D_v$. Moreover, a necessary condition on the channel capacity for the reconstruction of $Y^{T-1}, Y_t \in \mathfrak{R}^d$ in the r -mean is

$$C^{Z, \bar{Z}} \geq \mathcal{H}_S(\mathcal{Y}) - \log e^{\frac{d}{r}} + \log\left(\frac{r}{dV_d \Gamma(\frac{d}{r})} \left(\frac{d}{rD_v}\right)^{\frac{d}{r}}\right) = R_S^{Y, \bar{Y}}(D_v), \quad (14)$$

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

Furthermore, for the case when the observed process, $Y^{T-1}, Y_t \in \mathfrak{R}^d$, and the signal to be controlled, $H^{T-1}, H_t \in \mathfrak{R}^d$, are related by $Y_t = H_t + \Upsilon_t$, (e.g., $\rho(H_t, 0) = \rho(Y_t - \Upsilon_t, 0)$), then (13) and (14) are also necessary conditions for stability of the sequence $H^{T-1}, H_t \in \mathfrak{R}^d$ in probability and r -mean.

Proof. See Appendix. \square

Remark 4.9. We have the following remarks regarding the results of Theorem 4.8.

(i) First, we note that the above necessary conditions have not appeared elsewhere.

(ii) The lower bounds (13) and (14) given in Theorem 4.8 hold for any observed process, and they depend on the type of reconstruction and stability criteria employed. Hence, when Theorem 4.8 is applied to the controlled system (1), then the Shannon entropy rate is that of the controlled output of system (1). However, under assumption that (C, A) is detectable, $(A, (BB^T)^{\frac{1}{2}})$ is stabilizable, and $D \neq 0$ when the encoder and decoder are of Class A, by Corollary 3.4 we deduce that bounds (13) and (14) also hold when the Shannon entropy rate is replaced by the Shannon entropy rate of the output process of the uncontrolled version of system (1). This is also true when the encoder and decoder are of Class B.

(iii) The condition (13) and (14) are given in terms of the Shannon entropy rate which can be easily computed. Further, by Remark 3.3, these conditions imply the eigenvalue rate condition which appeared in the literature.

(iv) For the case of $d = 1$, condition $\int_{\|\xi\| > \delta} h^*(\xi) d\xi = D_v$ is reduced to $2\Phi\left(-\frac{\delta}{\sqrt{\Gamma_g}}\right) = D_v, \Phi(t) \triangleq \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$. Using a table for this integral, we notice that for a given $\delta \geq 0, \Gamma_g \leq \frac{\delta^2}{16}$ gives a small value for D_v . On the other hand, using a Γ_g smaller than $\frac{\delta^2}{16}$ does not yield significantly different result for reconstruction and stability performance, so for a small quantity of $D_v, \Gamma_g = \frac{\delta^2}{16}$ can be used in (13). Further, for $d = 1$ and $r = 2$, the extra term $-\log e^{\frac{d}{r}} + \log\left(\frac{r}{dV_d \Gamma(\frac{d}{r})} \left(\frac{d}{rD_v}\right)^{\frac{d}{r}}\right)$ in (14) is given by $-\frac{1}{2} \log(2\pi e D_v)$ which implies that smaller D_v requires bigger channel capacity for mean square reconstruction and stability.

4.3. Design of communication system for controlled systems

In this section, we design an encoder of Class B and C, and a decoder of Class B and A for the control system (1), when $Y_t \in \mathfrak{X}$ (the general case is similar). Moreover, we show a separation principle between the design of the control and communication systems. The design of the encoder and decoder is done as follows. First, we compute the rate distortion function. Second, we match the feedback innovations process to the AWGN channel

$$\tilde{Z}_t = Z_t + \tilde{W}_t, \quad \tilde{W}_t \text{ orthogonal } \sim N(0, W_c), \quad (15)$$

by identifying a specific encoder/decoder and controller. This leads to a joint source-channel coding theorem. First, note that if the encoder is an innovations encoder that does not use channel feedback (e.g. of Class A), then it is not possible to stabilize an unstable controlled system. Thus, consider the control/communication system of Fig. 1 described by the stochastic control system (1) and encoder and decoder of Class B (see Fig. 2). The encoder consists of a pre-encoder which produces the orthogonal Gaussian innovations process $\{K_t; t \in \mathbf{N}_+\}; K_t \triangleq Y_t - E[Y_t | \sigma\{\tilde{K}^{t-1}, U^{t-1}\}] = Y_t - CE[X_t | \sigma\{\tilde{K}^{t-1}, U^{t-1}\}]$ using the feedback channel information \tilde{K}^{t-1} of the decoder output and the previous control sequence U^{t-1} . We assume an encoder $Z_t = \alpha_t K_t$ and a decoder $\tilde{K}_t = \gamma_t \tilde{Z}_t$, where α_t, γ_t are non-negative scalars to be determined so that the link from K_t to \tilde{K}_t is matched to the minimizing stochastic kernel. Note that

$$\begin{aligned} \tilde{K}_t &= \gamma_t \tilde{Z}_t = \gamma_t (Z_t + \tilde{W}_t) \\ &= \gamma_t \alpha_t \left(Y_t - CE[X_t | \sigma\{\tilde{K}^{t-1}, U^{t-1}\}] \right) + \gamma_t \tilde{W}_t, E[Z_t^2] = P_t. \end{aligned} \quad (16)$$

Define the mean square state estimator $\hat{X}_t \triangleq E[X_t | \sigma\{\tilde{K}^{t-1}, U^{t-1}\}]$. Then an application of least square estimation when the measurements admit output feedback specified by (16) gives the

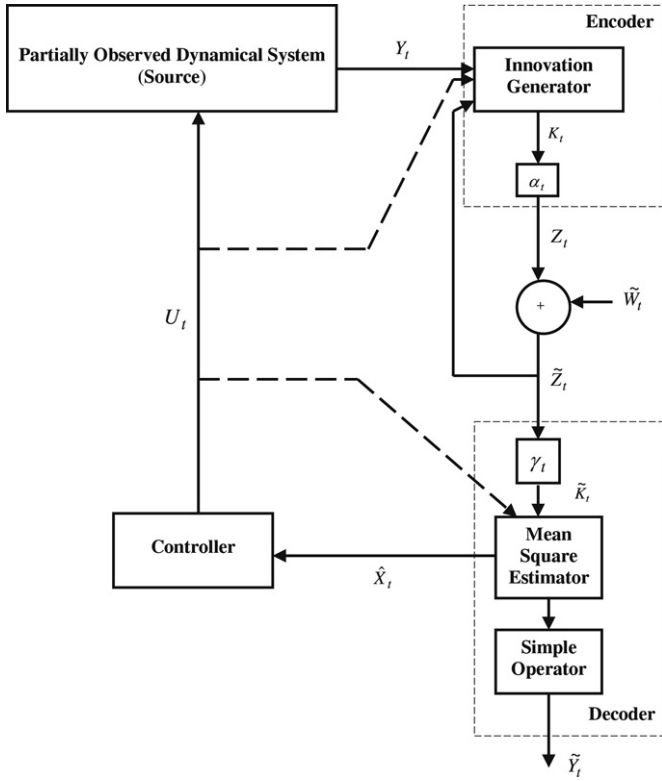


Fig. 2. Control/communication system described by stochastic control system (1).

following recursive Kalman filter estimate. $\hat{X}_{t+1} = A\hat{X}_t + \frac{1}{\alpha_t \gamma_t} A \Pi_t C^T (C \Pi_t C^T + DD^T + \frac{W_c}{\alpha_t^2})^{-1} \tilde{K}_t + NU_t$, $\hat{X}_0 = \bar{x}_0 = EX_0$, where the control is of the form $U_t = \mu(t, \tilde{K}^{t-1}, U^{t-1})$ ($\mu(\cdot)$ will be defined shortly to minimize a quadratic cost) and Π_t is the mean square state estimation error given by

$$\begin{aligned} \Pi_{t+1} &= A \Pi_t A^T - A \Pi_t C^T \left(C \Pi_t C^T + DD^T + \frac{W_c}{\alpha_t^2} \right)^{-1} \\ &\quad \times C \Pi_t A^T + BB^T, \quad \Pi_0 = \bar{V}_0. \end{aligned} \quad (17)$$

Rate distortion computation: Consider the orthogonal Gaussian feedback innovations process $K_t \triangleq Y_t - E[Y_t | \sigma\{\tilde{K}^{t-1}, U^{t-1}\}]$, $K_t \sim N(0, \Psi_t)$, $t \in \mathbf{N}_+$, where $\Psi_t \triangleq C \Pi_t C^T + DD^T$ in which Π_t is the solution of the Riccati equation (17), corresponding to the control system (1) with $U_t = \mu(t, \tilde{K}^{t-1}, U^{t-1})$. Also consider a mean square distortion measure $\rho_T(k^{T-1}, \tilde{k}^{T-1}) = \frac{1}{T} \sum_{t=0}^{T-1} (k_t - \tilde{k}_t)^2$.

The minimizing stochastic kernel of $R_T^{K, \tilde{K}}(D_v)$, for $D_v < \min_{t \in \mathbf{N}_+} \Psi_t$ is given by $P^*(d\tilde{k}^{T-1}; k^{T-1}) = \prod_{t=0}^{T-1} q^*(\tilde{k}_t | k_t) d\tilde{k}_t$, $q^*(\tilde{k}_t | k_t) \sim N(\eta_t k_t, \eta_t D_v)$, $\eta_t \triangleq 1 - \frac{D_v}{\Psi_t}$. By Remark 4.4, we have $R_T^{K, \tilde{K}}(D_v) = R_{T, SRD}^{K, \tilde{K}}(D_v) = \frac{1}{2} \sum_{t=0}^{T-1} \log \frac{\Psi_t}{D_v}$, $D_v < \min_{t \in \mathbf{N}_+} \Psi_t$. On the other hand, from the expression of the pre-encoding scheme it follows that $\tilde{Y}_t = \tilde{K}_t + C\hat{X}_t$ is the reconstruction of Y_t at the communication end. For this reconstruction, we have $E(Y_t - \tilde{Y}_t)^2 = E(K_t - \tilde{K}_t)^2$, $\forall t \in \mathbf{N}_+$. Under the assumption that (C, A) is detectable and $(A, (BB^T)^{\frac{1}{2}})$ is stabilizable, then $\Pi_\infty \triangleq \lim_{T \rightarrow \infty} \Pi_T$ exists and it is given by the Algebraic Riccati equation

$$\begin{aligned} \Pi_\infty &= A \Pi_\infty A^T - A \Pi_\infty C^T \left(C \Pi_\infty C^T + DD^T + \frac{D_v}{\eta_\infty} \right)^{-1} \\ &\quad \times C \Pi_\infty A^T + BB^T, \end{aligned} \quad (18)$$

where $\eta_\infty \triangleq 1 - \frac{D_v}{\Psi_\infty}$ and $\Psi_\infty = C \Pi_\infty C^T + DD^T$. Subsequently, for $D_v < \min_{t \in \mathbf{N}_+} \Psi_t$, then $R^{K, \tilde{K}}(D_v) = \lim_{T \rightarrow \infty} \frac{1}{T} R_T^{K, \tilde{K}}(D_v) = \frac{1}{2T} \sum_{t=0}^{T-1} \log \frac{\Psi_t}{D_v} = \frac{1}{2} \log \frac{\Psi_\infty}{D_v}$.

Separation principle: Here, we shall show that the encoder/decoder is independent of the control sequence U^{T-1} , and hence the encoder and decoder do not need access to the control sequence.

Let $\mathcal{G}^{t,u} \triangleq \sigma\{\tilde{K}^t, U^t\}$ and $\mathcal{G}^{t,0} \triangleq \sigma\{\tilde{K}^{t,0}\}$, $t \in \mathbf{N}_+$, where the superscript u denotes dependence on the control sequence U^t , $\tilde{K}^{t,0}$ is the decoder output arising from $\tilde{K}_t = \gamma_t \tilde{Z}_t$, $Z_t = \alpha_t K_t$, with $U_t = 0$, and $K_t^0 \triangleq Y_t - C\hat{X}_t$ is the innovations process with $U_t = 0$. First, note that $U_t \in \mathcal{G}^{t-1,u} = \sigma\{\tilde{Z}^{t-1}, U^{t-1}\}$, and $\hat{X}_t \in \mathcal{G}^{t-1,u}$. Following an induction method as in Caines (1988, pages 688, 689), we deduce that, $Y_t - C\hat{X}_t = K_t^0$ and $\tilde{K}_t - E[\tilde{K}_t | \mathcal{G}^{t-1,u}] = \alpha_t \gamma_t (Y_t - C\hat{X}_t) + \gamma_t \tilde{W}_t - 0 = \tilde{K}_t^0$, and that $\mathcal{G}^{t,u} = \mathcal{G}^{t,0}$, $\hat{X}_t \in \mathcal{G}^{t,0}$, $t \in \mathbf{N}_+$. Hence, the encoder, decoder and feedback information provided to the decoder is independent of the control U . Consequently, the encoder can be of Class C, while the decoder of Class A, and thus in Fig. 2, the link between the output of the controller and the input to the encoder/decoder can be removed. Thus, separation holds between the design of the communication system (encoder, decoder) and the controller.

Furthermore, given a sequence of the mean square state estimation $\hat{X}^{T-1} = (\hat{X}_0, \dots, \hat{X}_{T-1})$ (at each instance of time, \hat{X}_t is known to the transmitter and receiver), we have $I(Y^{T-1} \rightarrow \tilde{Y}^{T-1}) = \sum_{t=0}^{T-1} H_S(\tilde{Y}_t | \tilde{Y}^{t-1}) - \sum_{t=0}^{T-1} H_S(\tilde{Y}_t | \tilde{Y}^{t-1}, Y^t) = \sum_{t=0}^{T-1} H_S(\tilde{K}_t^0 + C\hat{X}_t | \tilde{K}_t^0 + C\hat{X}_{t-1}, \dots, \tilde{K}_{t-1}^0 + C\hat{X}_{t-1}) - \sum_{t=0}^{T-1} H_S(\tilde{K}_t^0 + C\hat{X}_t | \tilde{K}_t^0 + C\hat{X}_{t-1}, K_0^0 + C\hat{X}_0, \dots, \tilde{K}_{t-1}^0 + C\hat{X}_{t-1}, K_t^0 + C\hat{X}_t) = \sum_{t=0}^{T-1} (H_S(\tilde{K}_t^0 | \tilde{K}^{t-1,0}) - H_S(\tilde{K}_t^0 | \tilde{K}^{t-1,0}, K^{t,0})) = I(K^{T-1,0} \rightarrow \tilde{K}^{T-1,0}) = I(K^{T-1} \rightarrow \tilde{K}^{T-1}) = I(K^{T-1,0}; \tilde{K}^{T-1,0}) = I(Y^{T-1}; \tilde{Y}^{T-1})$, where the first equality follows from the definition, the second equality follows by substituting the expression, and separation principle, the third by a variant of Cover and Thomas (1991, Theorem 9.6.3), the fourth by definition, the fifth by separation principle, the sixth by conditional independence $P(dk_t; k^{t-1}, \tilde{k}^{t-1}) = P(dk_t; k^{t-1})$, $\forall t \in \mathbf{N}_+$, P -a.s., the seventh because $P(dy_t; y^{t-1}, \tilde{y}^{t-1}) = P(dy_t; y^{t-1})$, $\forall t \in \mathbf{N}_+$, P -a.s., which implies $I(Y^{T-1} \leftarrow \tilde{Y}^{T-1}) = 0$ (note that \tilde{K}_0^0 can be taken to be zero, and $\hat{X}_{-1} = \hat{X}_0$).

Next, for a given sequence \hat{X}^{T-1} , consider the rate distortion functions $R_{T, SRD}^{Y, \tilde{Y}}(D_v)$ and $R_T^{K, \tilde{K}}(D_v)$ with a mean square distortion measure $\rho_T(y^{T-1}, \tilde{y}^{T-1}) = \frac{1}{T} \sum_{t=0}^{T-1} (y_t - \tilde{y}_t)^2$ and $\rho_T(k^{T-1}, \tilde{k}^{T-1}) = \frac{1}{T} \sum_{t=0}^{T-1} (k_t - \tilde{k}_t)^2$, respectively. From above analysis, since the distortion constraints are the same and $I(Y^{T-1} \rightarrow \tilde{Y}^{T-1}) = I(K^{T-1}; \tilde{K}^{T-1})$, then, $R_{T, SRD}^{Y, \tilde{Y}}(D_v) = R_T^{K, \tilde{K}}(D_v)$.

Computation of Shannon lower bound: The Shannon lower bound is given by $R^{K, \tilde{K}}(D_v) \geq R_S^{K, \tilde{K}}(D_v) = \mathcal{H}_S(\mathcal{K}) - \max_{h \in \mathcal{G}_D} H_S(h) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=0}^{T-1} H_S(K_i | K^{i-1}) - \max_{h \in \mathcal{G}_D} H_S(h) = \frac{1}{2} \log \Psi_\infty - \frac{1}{2} \log D_v$. Subsequently, for $D_v < \min_{t \in \mathbf{N}_+} \Psi_t$, the rate distortion function $R^{K, \tilde{K}}(D_v)$ and Shannon lower bound are the same, hence in this case the lower bound is tight.

Realization of a communication link matched to the source: Next, for the AWGN channel (15) a matched communication link such that the innovations process-to-reconstruction behaves like the rate distortion minimizing stochastic kernel is obtained if $\alpha_t = \sqrt{\frac{\eta_t W_c}{D_v}}$, $\eta_t \triangleq 1 - \frac{D_v}{\Psi_t}$ and $\gamma_t = \sqrt{\frac{D_v \eta_t}{W_c}}$, for $D_v < \min_{t \in \mathbf{N}_+} \Psi_t$. The power constraint corresponding to this encoding scheme is $E[Z_t^2] = \alpha_t^2 \Psi_t = \frac{\eta_t W_c}{D_v} \Psi_t \triangleq P_t$.

Moreover, under the assumption that (C, A) is detectable and $(A, (BB^T)^{\frac{1}{2}})$ is stabilizable, then $\Pi_\infty \triangleq \lim_{T \rightarrow \infty} \Pi_T$ exists and it is given by (18). Subsequently, for $D_v < \min_{t \in \mathbf{N}_+} \Psi_t$, $\mathcal{C}^{z, \tilde{z}} = R^{K, \tilde{K}}(D_v) = \frac{1}{2} \log \frac{\Psi_\infty}{D_v} = R_{S^{\tilde{K}}}^{K, \tilde{K}}(D_v)$. Therefore, using feedback channel information to communicate \tilde{Z}^{t-1} , the encoder for the output process Y^{T-1} is $K_t \triangleq Y_t - C\hat{X}_t$, and its reconstructed version is $\tilde{Y}_t = \tilde{K}_t + C\hat{X}_t$, having distortion $E(Y_t - \tilde{Y}_t)^2 = E(K_t - \tilde{K}_t)^2 = D_v$, $\forall t \in \mathbf{N}_+$ by transmitting $\mathcal{C}^{z, \tilde{z}} = R_{S^{\tilde{K}}}^{Y, \tilde{Y}}(D_v) = R^{K, \tilde{K}}(D_v) = R_{S^{\tilde{K}}}^{K, \tilde{K}}(D_v)$ ($D_v < \min_{t \in \mathbf{N}_+} \Psi_t$) bits per source message. This shows that given control sequence, for the specific encoder/decoder and for mean square distortion measure the lower bound in Theorem 4.8 is also a sufficient condition for reconstructing the processes K^T and Y^T .

Control law: Suppose in addition to the previous assumptions for existence of the rate distortion function, $((C^T C)^{\frac{1}{2}}, A)$ is detectable and (A, N) is stabilizable. Then the control law which minimizes the LQG cost functional (Caines, 1988) $\lim_{T \rightarrow \infty} \frac{1}{T} E \sum_{t=0}^{T-1} (\|X_t\|_{C^T C}^2 + \|U_t\|_H^2)$ ($H > 0$) is given by $U_t = -\Delta \hat{X}_t$, where $\Delta = (H + N^T P_\infty N)^{-1} N^T P_\infty A$ and P_∞ is the unique positive semi-definite solution of the following Algebraic Riccati equation $P_\infty = A^T P_\infty A - A^T P_\infty N (H + N^T P_\infty N)^{-1} N^T P_\infty A + C^T C$. This shows that via the above feedback channel information the sequence H^{T-1} is mean-square stable (follows from standard LQG theory).

Moreover, from the above construction it is evident that the design of the controller (stability) is independent of the design of the communication system (reconstruction), hence a separation principle holds, and the control is a certainty equivalence control law (Caines, 1988).

Remark 4.10. Notice that $\alpha_t \gamma_t = \eta_t$. In the limit, as $D_v \rightarrow 0$, then $\eta_t \rightarrow 1$, $\alpha_t \rightarrow \infty$. Thus the recursive estimator becomes the standard Kalman filter and the proposed certainly equivalent controller is reduced to the standard LQG controller (Caines, 1988).

Appendix

Proof of Theorem 4.8 (Reconstruction). Assume there exists an encoder/decoder pair such that reconstruction in probability is obtained. Then for a given $\delta \geq 0$ and $D_v \in [0, 1)$, there exists $T(\delta, D_v)$ such that $\forall t \geq T(\delta, D_v)$, $\frac{1}{t} \sum_{k=0}^{t-1} \Pr(\|Y_k - \tilde{Y}_k\| > \delta) \leq D_v$. Define $\rho_t(Y^{t-1}, \tilde{Y}^{t-1}) \triangleq \frac{1}{t} \sum_{k=0}^{t-1} \rho(Y_k, \tilde{Y}_k)$, where $\rho(\cdot, \cdot)$ is defined in Definition 2.1. Then, for $t \geq T(\delta, D_v)$, $E \rho_t(Y^{t-1}, \tilde{Y}^{t-1}) = \frac{1}{t} \sum_{k=0}^{t-1} E \rho(Y_k, \tilde{Y}_k) = \frac{1}{t} \sum_{k=0}^{t-1} \Pr(\|Y_k - \tilde{Y}_k\| > \delta) \leq D_v$. That is, a rate distortion with distortion level D_v is obtained for $t \geq T(\delta, D_v)$. Then by Theorem 4.5 the channel capacity and rate distortion must for all $t \geq T(\delta, D_v)$ satisfy $\frac{1}{t} \mathcal{C}_t^{z, \tilde{z}} \geq \frac{1}{t} R_t^{Y, \tilde{Y}}(D_v) \Leftrightarrow \lim_{t \rightarrow \infty} \frac{1}{t} \mathcal{C}_t^{z, \tilde{z}} \geq \lim_{t \rightarrow \infty} \frac{1}{t} H_S(Y^{t-1}) - \max_{h \in \mathcal{G}_D} H_S(h)$. Hence, $\mathcal{C}^{z, \tilde{z}} \geq \mathcal{H}_S(\mathcal{Y}) - \max_{h \in \mathcal{G}_D} H_S(h)$. Since among all distribution with the same covariance, the Gaussian distribution has the biggest entropy (Cover & Thomas, 1991), $h^*(\xi) \in \mathcal{G}_D$ that maximizes $H_S(h)$ is a Gaussian distribution which occurs on the boundary of \mathcal{G}_D . That is, $h^*(\xi) \sim N(0, \Gamma_g)$ in which Γ_g satisfies the equation of Theorem. Consequently, letting $H_S(h^*) = \frac{1}{2} \log[(2\pi e)^d \det \Gamma_g]$ (Cover & Thomas, 1991), the result is obtained. A necessary condition for reconstruction in r -mean is obtained similarly. The only difference is that from Linder and Zamir (1994), it follows that for this case, $\max_{h \in \mathcal{G}_D} H_S(h) = \log e^{\frac{d}{r}} - \log(\frac{r}{dV_d \Gamma(\frac{d}{r})} (\frac{d}{rD_v})^{\frac{d}{r}})$ bits per source message.

(Stability). Follows similarly by considering the rate distortion between Y^{t-1} and \mathcal{Y}^{t-1} . \square

References

Berger, T. (1971). *Rate distortion theory: A mathematical basis for data compression*. Prentice-Hall.

Caines, P. E. (1988). *Linear stochastic systems*. John Wiley and Sons.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. John Wiley and Sons.

Elia, N. (2004). When Bode meets Shannon: Control-oriented feedback communication schemes. *IEEE Transactions on Automatic Control*, 49(9), 1477–1488.

Li, K., & Baillieul, J. (2004). Robust quantization for digital finite communication bandwidth (DFCB) control. *IEEE Transactions on Automatic Control*, 49(9), 1573–1584.

Liberzon, D., & Hespanha, J. P. (2005). Stabilization of nonlinear systems with limited information feedback. *IEEE Transactions on Automatic Control*, 50(6), 910–915.

Linder, T., & Zamir, R. (1994). On the asymptotic tightness of the Shannon lower bound. *IEEE Transactions on Information Theory*, 40(6), 2026–2031.

Massey, J. (1990). Causality, feedback and directed information. In *Proceedings of the 1990 symposium on information theory and its applications* (pp. 303–305).

Malyavej, V., & Savkin, A. V. (2005). The problem of optimal robust Kalman state estimation via limited capacity digital communication channels. *Systems and Control Letters*, 54(3), 283–292.

Nair, G. N., Dey, S., & Evans, R. J. (2003). Infimum data rates for stabilizing Markov jump linear systems. In *Proceedings of the 42nd IEEE conference on decision and control* (pp. 1176–1181).

Nair, G. N., Evans, R. J., Mareels, I. M. Y., & Moran, W. (2004). Topological feedback entropy and nonlinear stabilization. *IEEE Transactions on Automatic Control*, 49(9), 1585–1597.

Savkin, A. V., & Petersen, I. R. (2003). Set valued state estimation via a limited capacity communication channel. *IEEE Transactions on Automatic Control*, 48(4), 676–680.

Tatikonda, S., & Mitter, S. (2004a). Control over noisy channels. *IEEE Transactions on Automatic Control*, 49(7), 1196–1201.

Tatikonda, S., & Mitter, S. (2004b). Control under communication constraints. *IEEE Transactions on Automatic Control*, 49(7), 1056–1068.

Tatikonda, S., Sahai, A., & Mitter, S. (2004). Stochastic linear control over a communication channel. *IEEE Transactions on Automatic Control*, 49(9), 1549–1561.

Tsumura, K., & Maciejowski, J. (2003). Stabilizability of SISO control systems under constraints of channel capacities. In *Proceedings of 42nd IEEE conference on decision and control* (pp. 193–197).

Wu, B. F., & Jonckheere, E. A. (1992). A simplified approach to Bode's theorem for continuous-time and discrete-time systems. *IEEE Transactions on Automatic Control*, 37(11), 1797–1802.



Charalambos D. Charalambous received the Electrical Engineering B.S. degree in 1987, the M.E. degree in 1988, and the Ph.D. in 1992, all from Department of Electrical Engineering from Old Dominion University, Virginia, USA. In 2003 he joined the Department of Electrical and Computer Engineering, University of Cyprus, where he is currently Associate Professor and Associate Dean of the School of Engineering. He was an Associate Professor at University of Ottawa, School of Information Technology and Engineering from 1999 to 2003. He has served on the faculty of McGill University, Department of Electrical and Computer Engineering, as a non-tenure faculty member, from 1995 to 1999. From 1993 to 1995 he was a Post-Doctoral Fellow at Idaho State University, Engineering Department. He is currently an associate editor of the IEEE Communications Letters, and from 2002 to 2004 he served as an Associate Editor of the IEEE Transactions on Automatic Control. He was a member of the Canadian Centers of Excellence through MITACS (the mathematics of information technology and complex systems), from 1998 to 2001. In 2001 he received the Premier's Research Excellence Award of the Ontario Province of Canada. Charalambous's research group ICCSSystems, Information, Communication and Control of Complex Systems is interested in theoretical and technological developments concerning large scale distributed communication and control systems and networks in science and engineering. These include theory and applications of stochastic processes and systems subject to uncertainty, communication and control systems and networks, large deviations, information theory, robustness and their connections to statistical mechanics.



Alireza Farhadi received a M.Sc. degree in Electrical Engineering from Iran University of Science and Technology, Tehran, Iran, in 2000. He then pursued his study towards a Ph.D. degree in Electrical Engineering at the University of Ottawa, Ontario, Canada (2002–2007). Dr. Farhadi's research interests include information theoretic aspects of estimation and control over noisy channels. He is currently a Post Doctoral Fellow at the School of Information Technology and Engineering of the University of Ottawa.