uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Alireza Farhadi
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Electrical Engineering)
GRADE / DEGREE

School of Information Technology and Engineering
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Stochastic Control Through Capacity Limited Channels

TITRE DE LA THÈSE / TITLE OF THESIS

Charalambos Charalambous
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Namir Ahmed
CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Christoforos Hadjicostis

Sergey Loyka

Florence Danilo-Lemoine

Yongyi Mao

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Stochastic Control Through Capacity Limited Channels

by

Alireza Farhadi

School of Information Technology and Engineering
University of Ottawa
Ottawa, ON, Canada

A dissertation submitted to the Faculty of Graduate and Postdoctoral Studies
in a partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Under supervision of Prof.'s C. D. Charalambous and N. U. Ahmed

Your file Votre référence
ISBN: 978-0-494-50729-2
Our file Notre référence
ISBN: 978-0-494-50729-2

Abstract

In the present thesis, we are concerned with reliable data reconstruction and stability of dynamical systems which are controlled over limited capacity communication channels.
Necessary conditions for reliable data reconstruction and stability of sequences in r-mean and probability are derived which depend on the entropy rate of the input to the encoder and the type of reconstruction and stability criteria. These conditions are given in terms of the Shannon lower bound and they are applicable to linear and nonlinear systems. Using these conditions, some of necessary conditions which are already available in the literature, are obtained as a special case. We discuss some of them throughout the thesis. Moreover, under certain conditions these necessary conditions are also sufficient. The results are applied to a linear stochastic partially observed control system subject to measurement noise when the channel is an Additive White Gaussian Noise (AWGN) channel. Here, we derive an encoder, decoder, and controller for mean square stability and reconstruction using the standard detectability and stabilizability assumptions of Linear Quadratic Gaussian (LQG) theory. From obtained conditions for reliable data reconstruction and stability, it is concluded that the Shannon capacity is still an adequate measure for describing the conditions for moment reliable data reconstruction and stability.
We find the continuous version of the well known eigenvalue rate condition for continuous time systems when they are controlled over continuous time AWGN channels. This condition is described by the summation of the real parts of the unstable eigenvalues of the open loop time-invariant system. This eigenvalue rate condition is obtained by addressing the necessary condition for stability of a fully observed linear continuous time-invariant noiseless plant; and by constructing an encoding scheme and a stability scheme for reliable data reconstruction and stability of a continuous time plant driven by Brownian motion.
Necessary conditions for uniform reliable data reconstruction and robust stability of the uncertain dynamical systems which are controlled over communication channels, are also

derived. These conditions are given in terms of the robust entropy rate of the inputs to the encoder and an additional term which is related to the reconstruction and stability criteria. The uncertainty in the dynamical system is described by a relative entropy constraint. Such uncertainty description is a natural generalization of the sum quadratic uncertainty description. These conditions are applied to specific uncertain systems by computing the robust entropy rate. Moreover, a relation between robust entropy rate and the Algebraic Riccati equation appearing in the H^{∞} estimation and control problem, is established. Under certain conditions the obtained necessary conditions are also sufficient. Furthermore, We study the stability problem of a fully observed controlled Gauss Markov uncertain system which is subject to the sum quadratic uncertainty restriction.

In addition, throughout it is shown that the Shannon lower bound is an adequate measure for describing the conditions for reliable data reconstruction and stability of sequences related to a dynamical system which is controlled over a limited capacity communication channel.

Acknowledgment

I would like to thank my thesis advisors Professor Charlambos D. Charalambous and Professor Nasir Uddin Ahmed for their guidance and support during my Ph.D. research. Professor Charalambous helped me in technical issues related to my research and writing the papers. He also inspired me in some ideas used in the thesis.

Professor Ahmed also helped me in every aspect of my thesis; particularly, in mathematics, system and control theory, and my defense.

Also, I would like to thank University of Cyprus for supporting my research by providing office space and financial support for the period of October 2004 to August 2005 when I spent at the University of Cyprus. It has been a privilege to work with intelligent and friendly people of the Electrical and Computer Engineering Department of the University of Cyprus.

In addition, I would like to thank Dr. Stojan Z. Denic for helpful discussions during my Ph.D. research and collaboration in writing some of the material of Chapter 3.

I would also like to thank my examination committee Professors Christoforos Hadjicostis, Florence Danilo-Lemoine, Sergey Loyka and Yongyi Mao for their valuable comments that improved the quality of the thesis.

Finally, I would like to thank my very special friends in Canada for their supports during my Ph.D. study. Without them, I could not finish my study.

Contents

List of Figures

List of Acronyms

1. a.s. - almost surely
2. AWGN - additive white Gaussian noise
3. DARE - discrete algebraic Riccati equation
4. DMC - discrete memoryless channel
5. CSI - channel state information
6. i.i.d. - independent identically distributed
7. LQG - linear quadratic Gaussian
8. PSD - power spectral density
9. PDF - probability density function
10. RND - Radon Nikodym derivative
11. R.P - random process
12. R.V. - random variable
13. WSS - wide sense stationary

Notation and List of Symbols

Probability Theory

1. $(\Omega, \mathcal{F})$ denotes a measurable space.

2. $(\Omega, \mathcal{F}, P)$ denotes a probability measurable space.

3. $(\Re^d, \mathcal{B}(\Re^d))$ denotes a Borel measurable space.

4. $\sigma\{.\}$ denotes the sigma-algebra.

5. $P(dY)$ denotes the probability measure associated with a R.V. Y.

6. $\mathcal{M}_1(\Re^d)$ denotes the set of all countably additive probability measures defined on the measurable space $(\Re^d, \mathcal{B}(\Re^d))$.

7. $E_{P(dY)}[Y]$, $E_P[Y]$, $E[Y]$, denote the expected value of the R.V. Y with respect to the probability measure $P(dY)$.

8. $\Pr(A)$ denotes the probability of the event A.

9. $Cov(Y)$ denotes the covariance of the R.V. Y.

10. $Var(Y)$ denotes the variance of the scalar R.V. Y.

11. $P(dY; x)$ denotes the stochastic kernel between two R.V.'s Y and X.

12. $E[X|Y], E[X|\sigma\{Y\}]$ denote the conditional expected value.

13. f_Y denotes the density function associated with R.V. Y.

14. $E_{f_Y}[Y]$ denotes the expected value of the R.V. Y with the density function f_Y.

15. $\mathcal{F}_t$ denotes filtration.

16. Y^T denotes a sequence with length $T+1$ of R.V.'s $(Y_0, ..., Y_T)$.

17. $\{Y_t; t \in [0,T]\}$ denotes discrete time R.P.

18. $\{y(t); t \in [0,T]\}$ denotes continuous time R.P.

19. $S_y(w)$ denotes the PSD of continuous time WSS R.P. $\{y(t); t \in [0,T]\}$.

20. $S_Y(e^{jw})$ denotes the PSD of discrete time WSS R.P. $\{Y_t; t \in [0,T]\}$.

21. $Y \sim N(\mu, \Gamma_Y)$ denotes a Gaussian R.V. with mean μ and covariance Γ_Y.

22. $X \rightarrow Y \rightarrow Z$ denotes conditional independence assumption among the R .V.'s X, Y, Z.

Information Theory

1. $H_S(Y) = H_S(f_Y)$ denotes the Shannon entropy associated with the R.V. Y with the density function f_Y.

2. $H_R(Y) = H_R(f_Y)$ denotes the Rényi entropy of the R.V. Y with the density function f_Y.

3. $H_T(Y) = H_T(f_Y)$ denotes the Tsallis entropy of the R.V. Y with the density function f_Y.

4. $H_S(X|Y)$ denotes the conditional Shannon entropy of the R.V. X given Y.

5. $\mathcal{H}_S(\mathcal{Y})$ denotes the Shannon entropy rate of the sequence of R.V.'s $Y^T \triangleq (Y_0, Y_1, ..., Y_T)$.

6. $\bar{\mathcal{H}}_S(\mathcal{Y})$ denotes the conditional Shannon entropy rate of the sequence of R.V.'s Y^T.

7. $I(X;Y)$ denotes the mutual information between R.V. X and R.V. Y.

8. $I(X;Y|Z)$ denotes the conditional mutual information between R.V. X and R.V. Y, given R.V. Z.

9. $I(X^T \rightarrow Y^T)$ denotes the directed information between sequences of R.V.'s $X^T = (X_0, ..., X_T)$ and $Y^T = (Y_0, ..., Y_T)$.

10. $\mathcal{C}_n$ denotes the channel capacity for time horizon n (n channel uses).

11. $\mathcal{C}$ denotes the channel capacity.

12. $\mathcal{R}$ denotes the transmission data rate.

13. $R_T(D_v)$ denotes the rate distortion function of the sequence of R.V.'s $Y^{T-1} = (Y_0, ..., Y_{T-1})$.

14. $R(D_v)$ denotes the rate distortion.

15. $R_S(D_v)$ denotes the Shannon lower bound.

16. $H(P||\Pi)$ denotes the relative entropy between probability measures $P(dY) \in \mathcal{M}_1(\Re^d)$ and $\Pi(dY) \in \mathcal{M}_1(\Re^d)$.

17. $R_{T,r}(D_v)$ denotes the robust rate distortion of the sequence of R.V.'s $Y^{T-1} = (Y_0, ..., Y_{T-1})$.

18. $R_r(D_v)$ denotes the robust rate distortion.

19. $R_{S,r}(D_v)$ denotes the robust Shannon lower bound.

20. $\mathcal{M}_{SU}^T$ denotes the set of admissible probability measures $P(dY^{T-1})$; $Y^{T-1} = (Y_0, ..., Y_{T-1})$.

21. $\mathcal{M}_{SU}$ denotes the set of admissible probability measures $P(dY)$.

22. $\mathcal{D}_{SU}^T$ denotes the set of admissible density functions $f_{Y^{T-1}}$; $Y^{T-1} = (Y_0, ..., Y_{T-1})$.

23. $\mathcal{D}_{SU}$ denotes the set of admissible density functions f_Y.

24. $\mathcal{D}^T$ denotes the set of all density functions associated with a sequence of R.V.'s $Y^{T-1} = (Y_0, ..., Y_{T-1})$.

25. $\mathcal{D}$ denotes the set of all density functions associated with a R.V. Y.

System and Control

1. $\lambda_i(A)$ denotes the eigenvalue of the square matrix A.

2. $S(z)$ denotes sensitivity transfer function of a linear discrete time-invariant system.

3. $S(s)$ denotes sensitivity transfer function of a linear continuous time-invariant system.

4. $L(z)$ denotes the open loop transfer function of a linear discrete time-invariant system.

5. $L(s)$ denotes the open loop transfer function of a linear continuous time-invariant system.

General Notation

1. $\log_e(.)$ denotes the logarithm of natural number.
2. $\log(.)$ denotes the logarithm of base 2.
3. $[x]^+$ denotes x if $x \geq 0$; otherwise $[x]^+ = 0$.
4. $\Re$ denotes the set of real numbers.
5. $\Re^+$ denotes the set of non-negative real numbers.
6. $\mathbf{N}_+ \triangleq \{0, 1, 2, ...\}$ denotes the set of natural numbers.
7. $\Re^d$ denotes the set of d-dimensional real vectors.
8. $||.||$ denotes the Euclidean norm on the space of $\Re^d$.
9. $||.||_\infty$ denotes the infinity norm.
10. I_n denotes the identity matrix with dimension n (i.e., $I_n \in \Re^{n\times n}$).
11. $Re(a)$ denotes the real part of the complex number a.
12. $L_1(X, Y)$ denotes the set of integrable functions (with respect to the Lebesgue measure) defined on the space X which take values in Y.
13. $L_1(P(dY))$ denotes the set of integrable functions with respect to the probability measure $P(dY)$, defined on the measurable measure $(\Re^d, \mathcal{B}(\Re^d))$.
14. M' denotes the transpose of matrix M.
15. M^{-1} denotes the inverse of the square matrix M.
16. $\det(M)$ denotes the determinant of the square matrix M.
17. $trac(A)$ denotes the trac of the square matrix M.

Chapter 1

Introduction

Recent advances in technology have created an increasing demand on networks. Recently, extensive research activity has been devoted to the question of how much capacity must be allocated to each component of a network. Questions of this kind are motivated by applications with limited communication resource that demand one transmits information under minimum possible capacity. In such applications, due to limited capacity constraint, the main assumption is that the source outputs (messages) can not be reproduced exactly as the original messages (i.e., with high precision) at the end of communication; and subsequently a distorted version of source outputs is available. Therefore, for a given distortion criterion it is essential to transmit information under minimum possible capacity, while the reliable data reconstruction of the source messages at the end of communication with respect to the given distortion criterion and/or stability of the controlled source, are guaranteed. Such restrictions on capacity is a common feature of the large scale communication networks such as internet or sensor networks. Such constraint is also a common feature of the tele-operation system of the space exploration devices and bio-instrumentations.

In the present thesis, we try to address above question. We consider a feedback control system which consists of a dynamical system (treated as the information source) and a controller, where the connection from sensors to controller is subject to limited capacity constraint and noise.

This introduction is divided to three sections. In Section 1.1, we describe the problem of reliable data reconstruction and stability subject to limited capacity constraint. In Section 1.2, we review some of the already existing results in the literature which are concerned with reliable data reconstruction and/or stability subject to limited capacity constraint. Finally, in Section 1.3, we discuss the problem considered in this thesis and contributions; and we

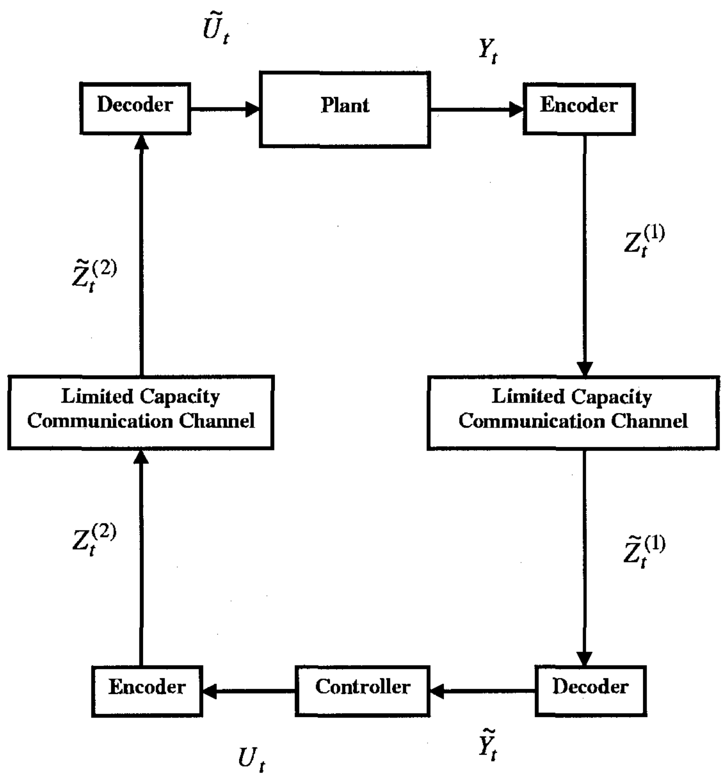

Figure 1.1: A basic block diagram of distributed systems

give a summary of chapters.

1.1 Limited Capacity Problems

Research on reliable data reconstruction and stability subject to limited capacity constraint is concerned with situations involving limited capacity communication links which can be often corrupted by noise.

Figure 1.1 illustrates a basic block diagram for studying reliable data reconstruction and stability subject to limited capacity constraint. Due to capacity constraint, the main assumption is that the source and the controller outputs can not be represented at the end of communication with high precision; and therefore a distorted version of the source and the control outputs are available.

In the literature, for understanding the interaction between control and (noisy) communication channels subject to limited capacity constraint, the focus has been mostly on the simplified model of Figure 1.2, in which throughout the thesis we call it the control/communication

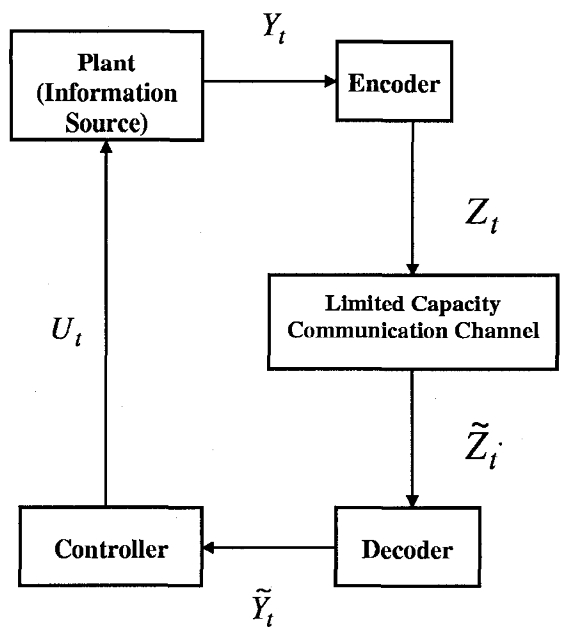

Figure 1.2: A control/communication system

system. In this model, there is a limited communication link from sensors to controller, while the connection from controller to plant is direct.

The control/communication system of Figure 1.2 can correspond to some real life problems. For example, it can correspond to the networked control system of Figure 1.3 where the plant and the corresponding controller are connected through a shared communication media, while there is an unshared or high capacity communication link from controller to plant. In this case, the connection from sensors to controller is subject to limited capacity constraint, while the connection from controller to plant is unconstraint (direct). The control/communication system of Figure 1.2 can also correspond to the tele-operation system of the space exploration devices such as the tele-operation system of Cassini-Huygens of Figure 1.4. Huygens is a space probe landed on Titan (one of the moons of the planet Saturn). The information produced by Huygens is transmitted to Cassini (mother ship), which is orbiting around the Saturn. Huygens is subject to limited power supply. Subsequently, the transmission from Huygens to Cassini is subject to limited capacity constraint. Nevertheless, due to the large power supply of Cassini, the transmission from Cassini to Huygens is not subject to such restriction on capacity. Please note that in addition of above constraint, the tele-operation system of Cassini-Huygens is also subject to long transmission delays, in which analysis of

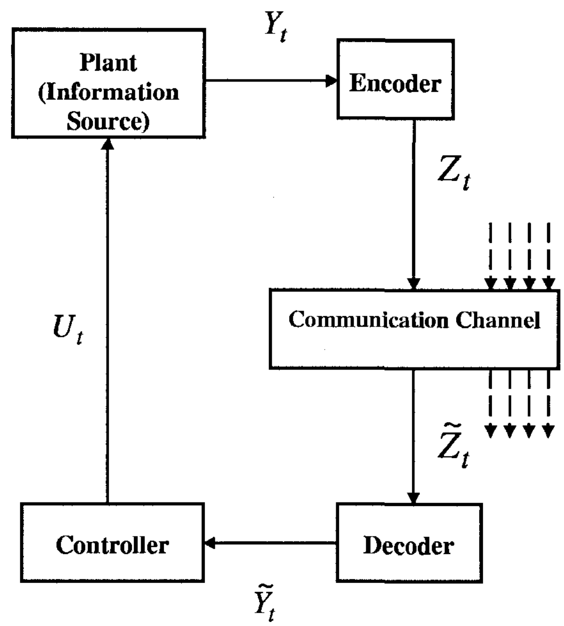

Figure 1.3: Network controlled system

the control/communication systems subject to long transmission delays is beyond the scope of this thesis. The control/communication system of Figure 1.2 can also correspond to bio-instrumentation devices which are used to monitor/regulate the human body/organs (see Figure 1.5). In such applications, it is important that the connection among the different components (sensors, actuators, and monitoring/regulator device) is wireless. Furthermore, in such applications the sensors are small in size; and subsequently they have limited power supply. Consequently, the connection from sensors to remote monitoring/regulator device is subject to limited capacity constraint. Nevertheless, the monitoring/regulator device is normally supported by the large amount of power supply. Subsequently, the connection from this device to the system which is supposed to be regulated is not subject to restriction on capacity. Please note that in the open loop case (i.e., when $U_t = 0$), the control/communication system of Figure 1.2 is also reduced to the communication system of Figure 1.6 subject to limited capacity constraint, which is a basic model for sensors network applications.

Due to broad range of applications of the control/communication system of Figure 1.2, throughout the thesis, we consider the problem of reliable data reconstruction and stability of this control/communication system.

Throughout, it is assumed that plant, communication channel, criteria for reliable data re-

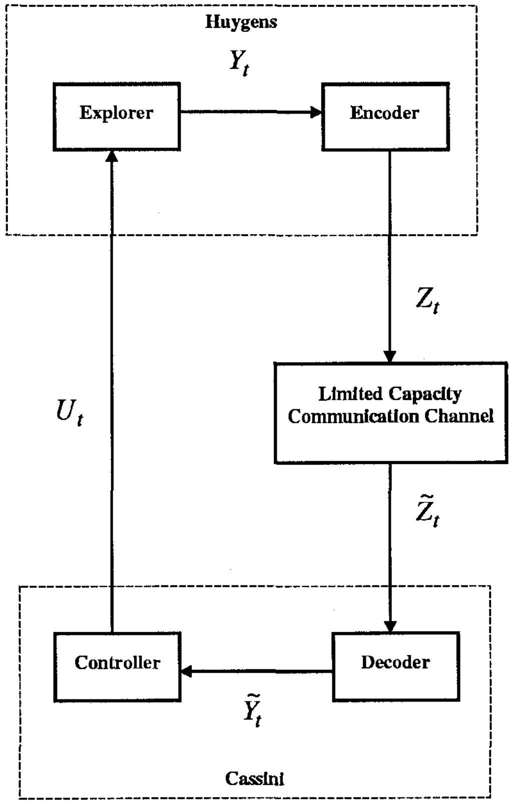

Figure 1.4: Tele-operation of Cassini-Huygens

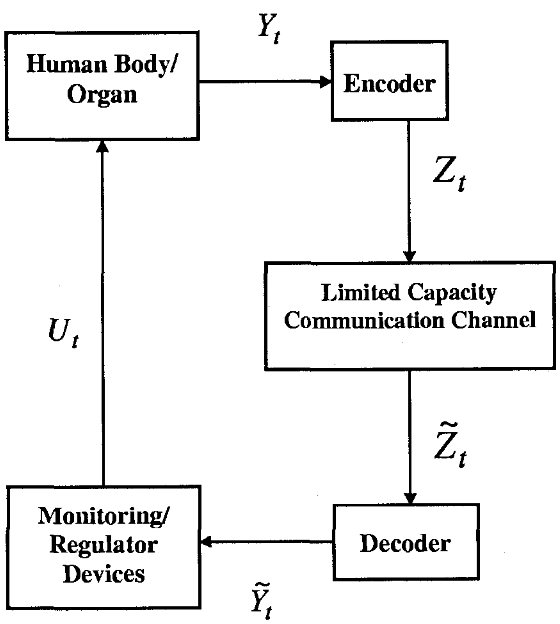

Figure 1.5: Bio-instrumentation

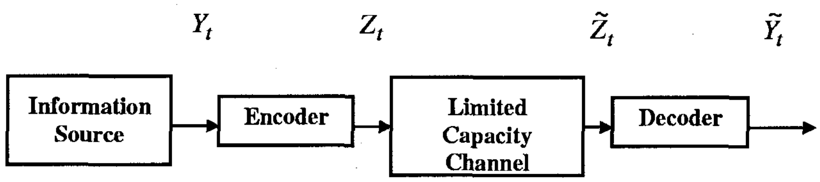

Figure 1.6: A basic block diagram of sensors network

construction and stability are given. The objective is to design an encoder, decoder and controller for reliable data reconstruction and stability when the capacity is the minimum required.

This objective is achieved by taking the following steps. First, we find a necessary condition in the form of a lower bound on the capacity in which if the capacity is less than this lower bound, there is no encoding scheme and stability scheme for reliable data reconstruction and stability. Then, we construct an encoder, decoder and controller which guarantee reliable data reconstruction and stability when the capacity is equivalent to the lower bound found as a necessary condition. Thus, this construction results in a sufficient condition on the capacity for reliable data reconstruction and stability. Following this approach, the minimum capacity under which there are an encoding scheme and a stability scheme for reliable data reconstruction and stability as well as encoder, decoder and controller that guarantee reliability and stability, are found.

Please note that in the literature, reliable data reconstruction is also known as observability.

1.2 Literature Review

There are two distinct research directions in simultaneously addressing communication and control aspects in modern systems and networks. In the first direction, the goal is the reliable data reconstruction and stability over (noisy) communication channels subject to limited capacity constraint [1]-[22]. Intuitively, the results in this direction provide a quantitative understanding of the way in which restriction on the data rate of the exchanged information among components of the system, degrades the performance of the system. This is the direction taken in this thesis. In the second direction, the goal is to stabilize a dynamical system over communication channels subject to noise, delay and loss, while there is no restriction on the transmission data rate [24]-[39].

For most part, research on reliable data reconstruction and stability subject to limited capacity constraint has been focused on the basic problem of Figure 1.2, beginning with [1] and [2] and continuing with [3]-[22]. Various publications have introduced necessary and sufficient conditions for observability and stability of Figure 1.2 [1], [4]-[11], [14], [15], [19]. In most part, these conditions are given in the form of a lower bound on the capacity in terms of rate of change of dynamical system (measured in bits per time). In particular, it is already known that the eigenvalue rate condition (i.e., $C \geq \sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$, where C is the capacity and $\lambda_i(A)$'s are the eigenvalues of the system matrix A of a linear discrete time-invariant system) represents the minimum capacity under which there are an encoding scheme and a stability scheme for observability and stability of linear time-invariant plants [4]-[13]. In other words, the eigenvalue rate condition is a necessary and sufficient condition (i.e., if and only if condition) on the channel capacity under which there are an encoding and/or stability schemes for observability and stability.

In most part, after finding a necessary condition in the form of a lower bound on the capacity, an encoder, decoder and controller have been proposed which guarantee reliable data reconstruction and/or stability when the capacity is equivalent to the lower bound found as a necessary condition. Subsequently, in these publications this lower bound on the capacity represents the necessary and sufficient condition (i.e., if and only if condition) on the channel capacity for observability and/or stability. In other words, this lower bound is the minimum capacity under which there are an encoding and/or stability schemes for observability and/or stability.

Delchamps in [1] addressed the problem of stability of an unstable fully observed linear time-invariant discrete time noiseless plant subject to random initial condition. He used a state feedback subject to the quantized measurements. In [1], it has been concluded that if the underlying system is unstable, there is no control strategy for global asymptotic stability, while if the system is stable, one can implement feedback control law which brings the closed loop system trajectory arbitrary close to zero for an arbitrary long time. Furthermore, Wong and Brockett in [2] investigated the state estimation problem of a partially observed continuous time-invariant system involving finite communication capacity constraint. Unlike classical estimation problems where the observation is a continuous process corrupted by noise, in [2], there is a constraint that the observations must be coded and transmitted over a digital communication channel with finite bit rate (i.e., finite capacity). In [2], it is assumed that the observation data, once obtained, are encoded to form a codeword according to a predefined prefix coding scheme [40]. Since transmission of the coded information imposes a transmission delay, the estimation problem of the corresponding discrete time system over digital noiseless channel with finite bit rate is analyzed. In [2], some recursive coder-estimator algorithms have been developed and various conditions connecting the communication data rate with the rate of change of the underlying dynamics (measured in bits per sampling time) have been established for bounded or asymptotic mean square estimation.

Observability and Stability of Linear Systems

In [3]-[13], authors considered observability and/or stability of linear plants subject to limited capacity constraint.

Wong and Brockett in [3] considered the stability of a partially observed noiseless linear continuous time-invariant plant subject to random initial condition over limited capacity digital communication channels. In [3], the observation is transmitted to a remote decision maker and vice versa from the controller to the plant (see Figure 1.2). Unlike the classical models, the observed information is not transmitted continuously. Hence, the observed information is sampled with varying sampling period. Since this system can not be asymptotically stabilized when the underlying dynamics are unstable (Definition 6.2.38), a weaker stabilizing concept which is called containability condition is introduced. Containability condition requires that for any sphere N centered at the origin there exists an open neighborhood of the origin M and coding and feedback control laws such that any trajectory started in M remains in N for all time. Necessary and sufficient conditions for containability are derived. Sufficient condition is obtained by constructing a coding and feedback control law which guarantees the containability condition under proposed sufficient condition. There is a large gap between the obtained necessary and sufficient conditions. Nevertheless, for the scalar case, it is possible to tighten the sufficient condition to find a condition in the form of a necessary and sufficient condition (i.e., if and only if condition) relating the channel capacity to the rate of change of the underlying dynamic.

Moreover, Sahai in [4] considered observability and stability of a fully observed scalar linear discrete time-invariant plant subject to bounded but arbitrary disturbances over a discrete time noisy channel subject to limited capacity constraint (see Figure 1.2). That is, he considered the following dynamical system $\begin{cases} X_{t+1} = AX_t + NU_t + Z_t, \\ Y_t = X_t, \end{cases}$ where $X_t, U_t, Z_t, Y_t \in \Re$,

X_t is the state process, U_t is the control signal, Z_t is the disturbance such that $||Z_t|| \leq d$ almost surely (d is known, and $||.||$ is the Euclidian norm), and Y_t is the observation process (observed by sensors). In [4], the observability and stability criterion are the moment criteria. That is, $E||X_t - \tilde{X}_t||^r \leq D_v$ and $E||X_t||^r \leq D_v^c$, $\forall t \geq 0$, where $\tilde{X}_t$ is the reproduction of X_t at the end of communication; $r > 0$, $D_v \geq 0$ and $D_v^c \geq 0$ are finite. It has been shown that the eigenvalue rate condition described by Shannon capacity does not represent a tight bound (i.e., the necessary and sufficient condition) for moment observability and stability. Consequently, it has been concluded that a single number like Shannon capacity is not adequate to present the conditions for moment observability and stability. Subsequently, Sahai introduced the notion of anytime capacity (which is a parameterized notion of capacity; and it depends on the order of moment observability and stability criterion), in which the eigenvalue rate condition described by this capacity notion (i.e., $\mathcal{C}_{anytime} \geq \sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$ bits per time step , where $\mathcal{C}_{anytime}$ is the anytime capacity and $\lambda_i(A)$'s are the eigenvalues of the linear time-invariant system) represents a tight bound for moment observability and stability.

Furthermore, Nair and Evans in [5] considered a fully observed linear time-varying plant over a digital noiseless channel with finite data rate (see Figure 1.2). In [5], it has been shown that the optimal coder-controller under a certain finite horizon rth mean power cost, is given by a causally reformulated optimal quantizer for the initial output. Subsequently, by combining this result with asymptotic quantization theory, Nair and Evans derived an optimal coder-controller. Following this construction, a necessary and sufficient condition has been derived for existence of coder-controller that asymptotically stabilizes the plant in the sense that the rth moment output is taken to zero with time. The obtained necessary and sufficient condition relates the minimum channel capacity to the rate of change of the underlying dynamic. For the special case of finite dimensional and time-invariant plants, the results are simplified to the well known eigenvalue rate condition [4]-[13]. Moreover, Nair and Evans in [6] and [7] considered bounded mean square stability of a partially observed discrete time-invariant stochastic linear system subject to both process noise and measurement noise over digital noiseless channels with finite data rate (see Figure 1.2). The objective is to find the minimum capacity and designing a coder-controller for bounded mean square stability. This goal is achieved by finding a lower bound for the mean square of the state variables which is independent of the coder-controller. Following this lower bound, a necessary condition for stability is derived in terms of the eigenvalue rate condition [4]-[13]. Then, by constructing a specific coding scheme and a control scheme, authors establish the bounded mean square stability; and subsequently it is concluded that the eigenvalue rate condition is the minimum capacity for bounded mean square stability. Furthermore, Nair, Dey and Evans in [8] considered the stability of a fully observed noiseless jump linear plant subject to random initial condition over a noiseless digital channel with variable bit rate (see Figure 1.2). The objective is to define the coder-controller for the rth bounded output moment stability. Authors find a condition in the form of necessary and sufficient condition connecting the asymptotic average channel bit rates (i.e., $\mathcal{R} = \liminf_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1}\mathcal{R}_t$, where $\mathcal{R}_t$ is the number of bits transmitted in each time step) to the rate of change of underlying dynamic. The sufficient condition is obtained by constructing the coder-controller that provides stability under the minimum asymptotic average bit rates. In the special case of fully observed linear noiseless plant, the obtained conditions are simplified to the well known eigenvalue rate condition [4]-[13].

In addition, Tatikonda and Mitter in [9] considered almost sure asymptotic observability and stability of a noiseless partially observed linear discrete time-invariant plant subject to random initial condition over discrete time memoryless channels (see Figure 1.2). That is, they considered the following dynamical system $\begin{cases} X_{t+1} = AX_t + NU_t, & X_0 = X \\ Y_t = CX_t, & \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $Y_t \in \Re^d$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$ and $C \in \Re^{d\times q}$. The observability and stability criterion considered in [9] are almost sure asymptotic observability and stability which are defined as follows. Let the error be $E_t = X_t - \hat{X}_t$ where X_t is the state of the plant and $\hat{X}_t$ is the estimated state at the end of communication. The plant is almost surely asymptotically observable if there exists a control sequence $\{U_t\}$ and an encoder and a decoder such that $||E_t|| \to 0$ ($||.||$ is the Euclidian norm) almost surely. Furthermore, the plant is almost surely asymptotically stabilizable if there exists an encoder, a decoder, and a controller such that $||X_t|| \to 0$ almost surely. Using an information theoretic approach by invoking the information transmission theorem and by finding a lower bound for the rate distortion, authors find a necessary condition for the stability of the given plant. Then, by designing an encoding scheme, it is shown that the obtained necessary condition is also a sufficient condition over erasure channels. Further, Tatikonda and Mitter in [10] considered asymptotic observability and stability of a noiseless partially observed linear discrete time-invariant plant subject to random initial condition over digital noiseless channels (see Figure 1.2). Using information theoretic approach, a necessary condition is derived for the given plant. Then, by constructing an encoder, a decoder and a controller, it is shown that the obtained necessary condition is also a sufficient condition. Furthermore, Tatikonda, Sahai, and Mitter in [11] considered the control/communication system of Figure 1.2 described by a fully observed linear discrete time stochastic controlled system driven by Gaussian noise, over digital noiseless and discrete time Additive White Gaussian Noise (AWGN) channels. That is, they considered the following dynamical system $\begin{cases} X_{t+1} = AX_t + NU_t + W_t, \\ Y_t = X_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $W_t \in \Re^q$, $Y_t \in \Re^q$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, and W_t i.i.d. $\sim N(0, \Sigma_W)$ (Example 6.2.2) is the process noise. The objective is to design encoder, decoder and controller to minimize the Linear Quadratic Gaussian (LQG) cost functional. A decoder minimizing the mean square state decoding error is used; and subsequently the conditions under which the classical separation property between the design of the state estimator (decoder) and controller holds and the certainly equivalent controller is optimal, are presented. Furthermore, the sequential rate distortion framework is introduced. In [11], it has been shown that the optimal LQG cost functional is decomposed into two terms: A full knowledge cost (i.e., the optimal LQG cost functional when there is no communication constraint) plus an additional term which is caused by the decoding error. It has been also shown that this decomposition is valid if AWGN channel is matched to the source sequential rate distortion. In this case, the sequential rate distortion function of the fully observed linear discrete time stochastic uncontrolled system is a tight bound for the mean square observability over AWGN channels. That is, $\mathcal{C} \geq R^{seq}(D_v)$ is the necessary and sufficient condition for the mean square observability, where $\mathcal{C}$ is the capacity and $R^{seq}(D_v)$ denotes the sequential rate distortion (Definition 6.2.23).

In another direction, Elia in [12] considered the basic control/communication system of Figure 1.2 described by a partially observed noiseless discrete time-invariant single-input-output unstable noiseless plant subject to random initial condition over a discrete time AWGN channel with memory. The desired stability is the internal stability. It has been shown that the successful stabilization of an unstable linear system over a Gaussian channel corresponds to an actual communication system which uses feedback and transmits at a rate which is equivalent to $\sum_{\{i;\lambda_i(A)\geq 1\}} \log|\lambda_i(A)|$, where $\lambda_i(A)$'s are the eigenvalues of the system matrix A of the open loop system.

In addition, Li and Baillieul in [13] have introduced the Digital Finite Communication Bandwidth (DFCB) control framework which consists of a fully observed unstable linear scalar continuous-time invariant noiseless plant subject to random initial condition over digital noiseless channels. This control/communication system is subject to non-uniform asynchronism of plant sampling and control actions; and subsequently packet loss. Nevertheless, in [13] the objective is mostly to design a control scheme for bounded asymptotic stability around the origin when the state of the system is sampled uniformly and sampling and control actions are simultaneous. In [13], the encoded information is used to produce the control sequence which takes values from a finite set. For a given number of control values, Li and Baillieul propose the so called regular control strategy for stability which achieves the minimum possible bit rate. It has been concluded that the binary control strategy can tolerate the lowest minimum bit rate (i.e., capacity) in which this rate is equivalent to $\max\{0, \log|\lambda(A)|\}$.

Stability of Nonlinear and Uncertain Systems

Nair, Evans, Mareels and Morgan in [14] considered stability of nonlinear systems. The objective is stability of a fully observed noiseless nonlinear time-invariant dynamical system subject to unknown initial condition over a digital noiseless channel. This problem can be modeled by the basic block diagram of Figure 1.2 described by such plant and digital noiseless channel. In [14], authors have developed the notion of topological feedback entropy rate $h^{wi(si)}$ for completely deterministic system which measures the fastest rate at which initial state information can be generated. They have also shown that $\mathcal{C} \geq h^{wi(si)}$ is the necessary and sufficient condition on the channel capacity $\mathcal{C}$ for stability over digital noiseless channels. Furthermore, Liberzon and Hespanha in [15] considered the global asymptotic stability of a fully observed continuous time-invariant nonlinear dynamical systems where the measurements must be received by controller at discrete times; and the data available to the controller is a stream of binaries. That is, they considered the control/communication system of Figure 1.2 described by such nonlinear dynamical system over digital noiseless channels; and they found a sufficient condition for stability relating the channel capacity to parameters describing the nonlinear dynamical system.

In addition, in [4], [9], [10], and [16] authors considered stability of uncertain systems. In [4], [9], and [10] authors considered stability of fully observed linear time-invariant dynamical systems subject to bounded but arbitrary disturbances over erasure [4], [9], and digital noiseless channels [10], respectively. That is, they considered the following dynamical system

$\begin{cases} X_{t+1} = AX_t + NU_t + Z_t, \\ Y_t = X_t, \end{cases}$ where $X_t \in \Re^q, U_t \in \Re^o, Z_t \in \Re^q, Y_t \in \Re^q$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$ and Z_t is the disturbance such that $||Z_t|| \leq d$ almost surely. Moreover, Martins, Dahleh and Elia in [16] considered stability in the asymptotic rth mean sense of the fully observed scalar linear discrete time uncertain stochastic systems over a stochastic digital link. That is, they considered the following dynamical system $\begin{cases} X_{t+1} = A_t(1+\tilde{A}_t)X_t + U_t + G_t(X) + Z_t, \\ Y_t = X_t, \end{cases}$ X_t, $A_t, \tilde{A}_t, U_t, G(.), Z_t, Y_t \in \Re$, A_t is an Independent Identically Distributed (i.i.d.) process with known distribution, while the stochastic process $\tilde{A}_t$, the operator $G_t(X)$, and the process Z_t are such that $|\tilde{A}_t| \leq \bar{a}$, $||G||_\infty \leq \bar{g}$, and $|Z_t| \leq d$, where $\bar{a} \in [0,1)$, $\bar{g} \in [0,1)$ and $\bar{d} \geq 0$ are known constants and $||.||_\infty$ is the infinity norm (see [16], definition 1.2). The stability criterion is the rth mean moment asymptotic stability, that is, $\lim_{t\to\infty} E||X_t||^r \leq D_v^c$, where $r > 0$ and D_v^c are finite. The problem considered in [16] can be described by the basic block diagram of Figure 1.2 subject to uncertainty. In [16], authors studied the case where the communication is performed at a nominal rate and it is affected by an uncertain fluctuation governed by an Independent Identically Distributed (i.i.d) zero mean process. Such case are common in network communications subject to the unpredictable traffic loads. A control scheme which stabilizes the system under a condition relating the nominal bit rate to the dynamical system and additional terms which are caused by uncertainty in the plant and fluctuation on the nominal bit rate, is proposed. Furthermore, using a counting argument when there is no uncertainty effects on the plant, a necessary condition for the existence of the stabilizing control scheme is proposed. Subsequently, it is concluded that for the scalar time-invariant plants when there is no uncertainty effects on the plant, the nominal minimum bit rate for stability is equivalent to $\max\{0, \log|\lambda(A)|\}$ plus an additional term which is caused by the fluctuation on the nominal bit rate.

Stability of Distributed Control Systems

Some of the results for stability of Distributed Control Systems (DCS's) subject to limited capacity constraint are given in [17]-[21]. Tatikonda in [17], Yuksel and Basar in [18] and Matveev and Savkin in [19] considered the case of multiple sensors and centralized decoder where there is communication constraint from sensors to controller, while there is no communication constraint from controller to plant (see Figure 1.2). The dynamical system is a single linear time-invariant noiseless plant, that is, $\begin{cases} X_{t+1} = AX_t + NU_t, \\ Y_t^{(i)} = C^{(i)}X_t, \quad i = 1,2,...,M \end{cases}$ where

$X_t \in \Re^q$, $U_t \in \Re^o$, $Y_t^{(i)} \in \Re^{d_i}$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, and $C^{(i)} \in \Re^{d_i\times q}$. In [17]-[19] authors give the tightest lower bounds on the channel capacity (i.e., the minimum capacity) for which stability is possible. In particular, in [17] and [18] authors considered the stability of multiple senor systems via digital noiseless channel connecting sensors to controller. In [17] and [18], the assumption is that the sensors know the control signals. In addition, in [19] authors considered stability via finite alphabet communication channels subject to delay and loss. The information pattern of sensors is independent of the control signals. Nevertheless, it is assumed that the observation vectors at each of the sensors are sufficient to extract each of the observable modes. Such a restriction reduces the problem to a number of centralized encoding problems.

Furthermore, Nair, Evans and Caines in [20] and Li and Baillieul in [21] studied the stability of distributed linear time invariant feedback control systems subject to limited capacity digital noiseless channels. They propose necessary and sufficient conditions for stability in which under certain conditions they present the tightest lower bounds for stability. In particular, in [20] authors considered stability of distributed linear time invariant noiseless systems under pre-assumption that the system is diagonalizable and there are limited capacity digital channels connecting the sensors to the controllers. That is, they considered the following model $\begin{cases} X_{t+1}^{(i)} = A^{(ii)}X_t^{(i)} + \sum_{j=1}^{U} N^{(ij)}U_t^{(j)}, & 1 \le i \le M \\ Y_t^{(k)} = C^{(k)}[X_t^{(1)'} \ \dots \ X_t^{(M)'}]', & 1 \le k \le Y \end{cases}$ where $'$ denotes transpose, $X_t^{(i)} \in \Re^{q_i}$ is the state of the ith distributed plant, $U_t^{(j)} \in \Re^{o_j}$, $Y_t^{(k)} \in \Re^{d_k}$, $A^{(ii)} \in \Re^{q_i\times q_i}$, $N^{(ij)} \in \Re^{q_i\times o_j}$, and $C^{(k)} \in \Re^{d_k\times q}$; $q = q_1 + \dots + q_M$. They used Slepian-Wolf idea to construct stabilizing scheme. Nevertheless, in [21] authors observed that tiny amount of noise and asynchronism of the distributed sensors dramatically degrades the control performance of the scalar system $\begin{cases} X_{t+1} = AX_t + \sum_{j=1}^{U} N^{(j)}U_t^{(j)}, \\ Y_t^{(k)} = \mathcal{O}_k(X_t + G_t), & 1 \le k \le Y \end{cases}$ where $X_t, U_t^{(j)}, G_t \in \Re$, $A \in \Re$, $N^{(j)} \in \Re$, G_t is bounded and uniformly distributed, $Y_t^{(k)}$ is a binary digit and $\mathcal{O}_k(.) : \Re \rightarrow \{0,1\}$ is an operator. Subsequently, in [21] authors proposed a novel source-coding strategy which is similar but different from the Gray code for stability of such cases.

Stability over Feedforward and Feedback Channels

In addition to [3] and [13], Yuksel and Basar in [22] also considered the problem of remote stability when both feedforward (i.e., from the sensors to the controller) and feedback (i.e., from the controller to the plant) are subject to limited capacity constraint (see Figure 1.1). In [13], authors used a control strategy in which the control values are restricted to a finite set and subsequently, the connection from the controller to the plant is also via a digital noiseless channel with finite bit rate (i.e., finite capacity). In [22] authors considered a fully observed linear time-invariant scalar plant subject to Borwnian motion over finite alphabet memoryless noisy channels. In [22], the objective is bounded mean square stability. By re-writing the stability problem in the form of state estimation problem, a necessary condition is derived for stability over memoryless channels using information transmission results over a degraded relay channel. Then, by constructing a control scheme consisting of encoders, decoders and controller, it is observed that the obtained necessary conditions are far from being sufficient as long as the channel is noisy.

Open Problems

A quick look at existing results in the literature reveals that although some of the basic problems in reliable data reconstruction and stability subject to limited capacity constraint have been addressed, there are still plenty of issues that need to be addressed for the fundamental understanding of the interaction between control and limited capacity communication channels. Most results have been developed for fully observed systems [4], [5], [8], [11], [13], [14], [15] [22], that is, $\begin{cases} X_{t+1} = AX_t + NU_t + BW_t, \\ Y_t = X_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $W_t \in \Re^m$, $Y_t \in \Re^q$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, $B \in \Re^{q\times m}$ ($B = 0$ in [4], [5], [8], [13], [14], [15]) and W_t i.i.d. $\sim N(0, \Sigma_W)$); or partially observed systems without considering the effects of the measurement noise [3], [9], [10], [12], [16]-[20], that is, $\begin{cases} X_{t+1} = AX_t + NU_t, \quad X_0 = X \\ Y_t = CX_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $Y_t \in \Re^d$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$ and $C \in \Re^{d\times q}$. Nevertheless, in practical applications the effects of measurement noise are often present. In fact, in few published results authors considered the partially observed cases subject to the measurement noise [6], [7], that is, $\begin{cases} X_{t+1} = AX_t + NU_t + W_t, \\ Y_t = CX_t + G_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $Y_t \in \Re^d$, $W_t \in \Re^q$ is the process noise, $G_t \in \Re^d$ is the measurement noise, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, $C \in \Re^{d\times q}$, W_t i.i.d $\sim N(0, \Sigma_W)$, and G_t i.i.d. $\sim N(0, \Sigma_G)$. In [6] and [7] authors considered stability over digital noiseless channels.

In addition, most results have been developed for reliable data reconstruction and stability over digital noiseless or erasure channel which are models for noise free or internet like communication links, respectively [3], [5]-[10], [13]-[22]. Nevertheless, in most applications, the effects of communication noise are present. In fact, in few publications authors considered the cases over noisy communication channels in which they considered reliable data reconstruction and stability of fully observed systems over AWGN channel [4], [11], [12].

Although for a quite broad class of channels (i.e., memoryless channels) in [9] and [22] necessary conditions for stability have been presented, these results have been developed for a specific dynamical system and criterion for stability. In addition, in few publications the effects of uncertainty in the dynamical system in the observability and the stability performance have been studied, in which the results of these publications are limited to the fully observed uncertain systems subject to the bounded uncertainty over noise free communication channels [4], [9], [10] and [16]. That is, $\begin{cases} X_{t+1} = AX_t + NU_t + Z_t, \\ Y_t = X_t, \end{cases}$ where $X_t, Z_t, Y_t \in \Re^q$, $U_t \in \Re^o$ and Z_t is the disturbance such that $||Z_t|| \leq d$ almost surely. Furthermore, little effort has been devoted to address the stability question of nonlinear systems [14, 15], and the cases where there are communication constraints in both feedforward and feedback links, in which these results have been developed for stability over digital noiseless channels [3], [13],[22]. In [13] and [22], the plant is fully observed and scalar, while in [3], it is noiseless partially observed.

In addition, since a discrete time model is more consistent with today's digital communication links, in almost all published results in the literature, authors considered the discrete time framework. Nevertheless, in some applications, the analog modulation schemes may be interesting due to simplicity in building such schemes. Furthermore, having a complete theory which deals with continuous time systems will help us gain additional insight and understanding into building control/communuication systems in discrete time framework. Therefore, there is a lack of stability results dealing with cases where all subsystems are continuous time.

1.3 Contributions and Summary of Thesis

The aim of the current thesis is to address some of the mentioned open problems for reliable data reconstruction and stability. Throughout, the observability and stability of sequences (in probability and r-mean) are considered. For such an observability and stability criteria, it is shown that the Shannon lower bound (Lemma 6.2.24) which is a tight lower bound for the rate distortion function, is an adequate measure for describing the conditions for observability and stability.

Since tools from probability theory, control, and information/communication theory are going to be used extensively throughout the thesis to derive new results, in Appendix 6.2, we summarized some of the well known concepts, measures, and results from probability theory, information theory, and systems and control theory that we are going to use them throughout.

The main results and contributions given in the subsequent chapters are described below.

Chapter 2. In this chapter necessary conditions for observability and stability of discrete time systems over discrete time memoryless channels subject to limited capacity constraint, are derived. These conditions are given in Section 2.4.2. The necessary conditions are given in the form of a lower bound on the Shannon capacity in terms of the Shannon lower bound. These conditions are general and they are applicable to linear and nonlinear systems and time and frequency domains, while the already existing conditions have been derived for a specific dynamical system and observability and stability criteria [4]-[16]. These results particularly complements the results of [9], [11], and [22] in which for a specific dynamical system and stability criterion, the conditions for stability have been derived. Using these necessary conditions, some of necessary conditions found in the literature such as the one presented in [41], are obtained as a special case. Furthermore, by implementing the Bode integral formula, we apply the obtained necessary conditions to a dynamical system described in frequency domain. This result is given in Section 2.4.3. The method is then applied to a linear stochastic partially observed Gaussian control system subject to measurement noise. That is, the following dynamical system $\left\{ \begin{array}{l} X_{t+1} = AX_t + NU_t + BW_t, \\ Y_t = CX_t + DG_t, \end{array} \right.$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $Y_t \in \Re^d$, $W_t \in \Re^m$ is the process noise, $G_t \in \Re^l$ is the measurement noise, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, $B \in \Re^{q\times m}$, $C \in \Re^{d\times q}$, $D \in \Re^{d\times l}$, W_t i.i.d $\sim N(0, I_m)$, and G_t i.i.d. $\sim N(0, I_l)$, where $I_m \in \Re^{m\times m}$ and $I_l \in \Re^{l\times l}$ are identity matrices. For such system over Additive White Gaussian Noise (AWGN) channels, we also derive an encoder, decoder, and controller for mean-square stability and observability, using the standard detectability and stabilizability assumptions of LQG theory, without assuming the knowledge of the control sequence at the encoder/decoder. Thus, necessary and sufficient conditions for mean-square stability and observability are derived. These results are given in Section 2.4.4. From these conditions, it is concluded that the Shannon capacity is still an adequate measure for moment observability and stability. Moreover, a separation principle between the design of the control and communication systems, is shown. These results extend the results of [11] and [42] which are concerned with the mean square stability of the following fully observed Gaussian dynamical system over AWGN channels. That is, $\left\{ \begin{array}{l} X_{t+1} = AX_t + NU_t + W_t, \\ Y_t = X_t, \end{array} \right.$

where $X_t \in \Re^q$, $U_t \in \Re^o$, $W_t \in \Re^q$, $Y_t \in \Re^q$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, and W_t i.i.d. $\sim N(0, \Sigma_W)$. In summary, the main contributions of this chapter are the followings. i) Necessary conditions are derived for observability and stability which are applicable to linear and nonlinear systems and time and frequency domains. ii) Mean square observability and stability of a partially observed controlled dynamical system subject to the Gaussian process and measurement noises are addressed when this system is controlled over AWGN channels.

Chapter 3. In this chapter, we consider mean square observability and stability of a linear continuous time controlled system subject to the Brownian motion over flat fading continuous time AWGN channels subject to the limited capacity constraint. This problem is a continuous version of the problem considered in [11] and [42] which are concerned with stability of the control/communication system of Figure 1.2 described by a discrete time-invariant controlled system over a discrete time AWGN channel (i.e., no fading effects). We follow a quite similar methodology developed in Chapter 2. First, we show that the summation of the real parts of the unstable eigenvalues of the open loop system is the required capacity (i.e., a necessary condition on capacity) for mean square stability of the linear continuous time-invariant noiseless plant subject to the random initial condition over continuous time AWGN channels. This result is given in Section 3.3. Then, we propose an encoding scheme which achieves capacity and it provides mean square observability with the corresponding mean square decoding error as minimum as possible. Subsequently, we conclude that the summation of the real parts of the unstable eigenvalues of the open loop continuous time-invariant system is also the minimum capacity for mean square observability over AWGN channels. These results are given in Section 3.4. Next, using this encoding scheme, we propose a control scheme which minimizes an LQG cost functional and stabilizes the time-invariant system in the mean square sense when it is controlled over AWGN channels with a capacity at least equivalent to the summation of the real parts of the unstable eigenvalues of the open loop system. These results are given in Section 3.5.

The main contribution of this chapter is the following. The continuous version of the eigenvalue rate condition [4]-[13] for continuous time systems which are controlled over continuous time AWGN channels, are derived. This condition is described by the summation of the real parts of the unstable eigenvalues of the open loop time-invariant system.

Chapter 4. In this chapter by developing and invoking a robust version of the information transmission theorem and the Shannon lower bound, necessary conditions for uniform observability and robust stability of an uncertain dynamical system which is controlled over a limited capacity communication channel, are derived. These conditions relate the channel capacity to the robust Shannon lower bound. These conditions are given in Section 4.4.2. The uncertainty in the dynamical system is described by a relative entropy constraint. Such uncertainty description is a natural generalization of the sum quadratic uncertainty description. The obtained necessary conditions are applied to specific uncertain systems by calculating the corresponding robust entropy rate. In particular, a relation between robust entropy rate and the solution of the Algebraic Riccati equation appearing in the H^∞ control and estimation problem, is shown. These results are given in Section 4.3.2. Furthermore, the robust stability of a fully observed controlled uncertain Gauss Markov system subject to the sum quadratic uncertainty description is also considered; and subsequently it is shown that the obtained necessary condition for uniform observability can be also a sufficient condition. This result is given in Section 4.4.3.

The problem of uniform observability and/or robust stability of fully observed uncertain dynamical systems subject to the bounded uncertainty has been considered in [4], [9], [10], [16]. In [4], [9], [10] authors considered the following uncertain system $\begin{cases} X_{t+1} = AX_t + NU_t + Z_t, \\ Y_t = X_t, \end{cases}$ where $X_t, Z_t, Y_t \in \Re^q$, $U_t \in \Re^o$, and Z_t is the disturbance such that $||Z_t|| \leq d$ almost surely. Furthermore, in [16] authors considered the following dynamical system

$$\begin{cases} X_{t+1} = A_t(1 + \tilde{A}_t)X_t + U_t + G_t(X) + Z_t, \\ Y_t = X_t, \end{cases}$$

where $X_t \in \Re, A_t \in \Re, \tilde{A}_t \in \Re, U_t \in \Re, G(.) \in \Re, Z_t \in \Re, Y_t \in \Re$, A_t is an i.i.d. process with known distribution, while the stochastic process $\tilde{A}_t$, the operator $G_t(X)$ and the process Z_t are such that $|\tilde{A}_t| \leq \bar{a}$, $||G||_\infty \leq \bar{g}$, and $|Z_t| \leq d$, where $\bar{a} \in [0,1)$, $\bar{g} \in [0,1)$, and $d \geq 0$ are known constants. This chapter complements the already existing results in the literature by considering the uncertainty in the dynamical system described by a relative entropy constraint which is a natural generalization of the sum quadratic uncertainty description. Sum quadratic uncertainty description is a more suitable description for modeling uncertain dynamical systems than bounded uncertainty description considered in [4], [9], [10], [16]. We consider the observability and stability problem of the following uncertain system $\begin{cases} X_{t+1} = AX_t + NU_t + BW_t + B\bar{W}_t, \\ Y_t = X_t, \end{cases}$ where $X_t, Y_t \in \Re^q$, $U_t \in \Re^o$, $W_t, \bar{W}_t \in \Re^m$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, $B \in \Re^{q\times m}$, W_t orthogonal $\sim N(0, \Sigma_W)$, $\Sigma_W > 0$, and $\bar{W}_t$ is perturbed noise process such that it satisfies the following sum quadratic con-

straint $\frac{1}{2T}\sum_{t=0}^{T-2} E[\bar{W}_t'\Sigma_W^{-1}\bar{W}_t] \leq R_c + \frac{1}{2T}\sum_{t=0}^{T-1} E[Y_t' M Y_t]$, where R_c is a non-negative scalar, $M = M' \in \Re^{q\times q} \geq 0$ and T can be arbitrary large.
The problem of reliable data reconstruction and stability of above uncertain dynamical system over limited capacity communication channels has not been considered elsewhere; and thus this chapter contains original results.

In addition, from the results of Chapter 2 and Chapter 4, it follows that the Shannon lower bound is a tight bound (i.e., a necessary and sufficient condition) on capacity for controlling purposes (Remark 2.4.7,ii and Remark 4.4.6). Shannon lower bound is given in terms of entropy rate and an additional term which is related to the observability and stability criteria. This bound appears to be an adequate bound for describing the conditions for observability and stability of sequences related to the dynamical system. It can be calculated easily (Sections 2.3 and 4.3.2). For linear time-invariant systems, it results in the well known eigenvalue rate condition (Remark 2.3.3 and Corollary 2.4.5). It is applicable to both time domain (Sections 2.3 and 4.3.2) and frequency domain (Corollary 2.4.5 and 4.3.11). It is also applicable to both linear systems (Section 2.3) and nonlinear systems (Section 4.3.2).

Publications. Some of the contributions of the current thesis have been already published in the conference proceedings; or they are accepted for publication in the refereed journals. Below is the list of these publications.

Articles accepted in refereed journals

1- C. D. Charalambous, Alireza Farhadi, and S. Z. Denic, Control of continuous-time linear Gaussian systems over additive Gaussian wireless fading channels: a separation principle, accepted in the IEEE Transactions on Automatic Control. It was scheduled to be published in the March edition of transactions in 7 pages.

2- C. D. Charalambous and Alireza Farhadi, LQG optimality and separation principle for general discrete time partially observed systems over finite capacity communication channels, accepted in Automatica.

3- Alireza Farhadi and C. D. Charalambous, Robust coding for a class of sources: applications in control and reliable communication over limited capacity channels, accepted (subject to a minor revision) in Systems and Control Letters.

Other contributions (e.g., communications, papers in refereed conference proceedings, etc.)
1- Alireza Farhadi and C. D. Charalambous, Robust stabilizing scheme for uncertain systems controlled over limited capacity Additive White Gaussian Noise channels, in proceedings of the 2008 American Control Conference, June 11-13, 2008.

2- Alireza Farhadi and C. D. Charalambous, Robust control of feedback systems subject to limited capacity constraints, in proceedings of 46th IEEE Conference on Decision and Control, December 2007.

3- C. D. Charalambous and Alireza Farhadi, Control of feedback systems subject to the finite rate constraints via Shannon lower bound, in 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Cyprus, April 16-20, 2007.

4- Alireza Farhadi and C. D. Charalambous, Control of tele-operation systems subject to capacity limited channels and uncertainty, in proceedings of CCECE/CCGEI, Ottawa, May 7-10, 2006, pp. 492-497.

5- C. D. Charalambous and Alireza Farhadi, A mathematical framework for robust control over uncertain communication channels, in proceedings of 44th IEEE Conference on Decision and Control, Seville, Spain, December 12-15, 2005, pp. 2530-2535.

6- C. D. Charalambous, S. Denic and Alireza Farhadi, Control over wireless communication channel for continuous-time systems, in proceedings of 44th IEEE Conference on Decision and Control, Seville, Spain, December 12-15, 2005, pp. 3225-3230.

7- C. D. Charalambous and Alireza Farhadi, Robust entropy rate for uncertain sources and its applications in controlling systems subject to capacity constraints, in proceedings of 43rd Annual Allerton Conference on Communication, Control, and Computation, Allerton House, Chicago, September 28-30, 2005.

8- C. D. Charalambous, Alireza Farhadi, S. Denic and F. Rezaei , Robust control over uncertain communication channels, in proceedings of 13th Mediterranean Conference on Control and Automation, Cyprus, June 27-29, 2005, pp. 737-742.

9- C. D. Charalambous and Alireza Farhadi, Robust entropy rate for uncertain sources: applications to communication and control systems, in proceedings of 9th Canadian Workshop on Information Theory, Montreal, June 5-8, 2005, pp. 307-310.

10- C. D. Charalambous, F. Rezaei, S. Denic, A. Kyprianou, and Alireza Farhadi, Robust information transmission and control subject to uncertainty and power constraints, workshop H-2 (half a day) presented by above authors in 43rd IEEE Conference on Decision and Control, Bahamas, December 2004.

Chapter 2

Control of Discrete Time Systems

2.1 Introduction

Control of dynamical systems subject to finite communication channel capacity are often represented by the block diagram of Figure 2.1. This system can be viewed as a general communication system in which the output of the control 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.

Specific questions which have been addressed in the literature via a variety of methods include necessary and sufficient conditions for stability and observability of unstable control systems subject to limited channel capacity [1], [4]-[22]. For linear time-invariant systems the rate at which information is generated must be bounded below by the summation of the logarithms of the unstable eigenvalues of the source, that is, $\mathcal{C} \geq \sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$, where $\mathcal{C}$ is the channel capacity (measured in bits per time step) and $\lambda_i(A)$'s are the eigenvalues of the open loop system.

Using different methods, including information/communication theory tools, necessary conditions for observability and stability have been derived in [4]-[16]. Nevertheless, already existing conditions in the literature have been developed for specific dynamical systems and observability and stability criteria. One of the main contributions of this chapter is to derive necessary conditions for observability and stability of sequences in r-mean and probability for general systems. These necessary conditions depend on the entropy rate of the input to the encoder and the type of observability and stability criteria. Using these necessary conditions, some of the necessary conditions found in [4], [9]-[16], [41] are obtained as a special case. We discussed some of them throughout this chapter. Moreover, under certain

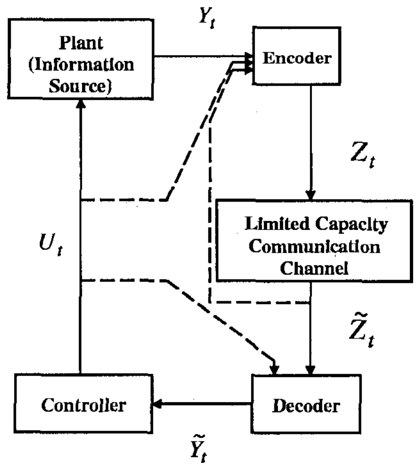

Figure 2.1: Discrete time control/communication system

conditions these necessary conditions are also sufficient. From these general conditions it is concluded that the Shannon capacity is still an adequate measure for describing conditions for moment observability and stability. This overcomes the drawback of using Shannon capacity in describing conditions for moment observability and stability, as discussed in [4]. In [4] author has concluded that the eigenvalue rate (i.e., $\sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$) is not a tight bound (i.e., a necessary and sufficient condition) on Shannon capacity for moment observability and stability of linear time invariant systems over noisy channels; and a bigger bound is required. Consequently, author has concluded that the Shannon capacity is not an adequate measure for describing the conditions for moment observability and stability. Subsequently, he introduced the anytime capacity notion which is a parameterized notion of capacity which depends on the order of moment observability and stability; and he showed that the eigenvalue rate condition described by anytime capacity is the necessary and sufficient condition for moment observability and stability.

The general information theoretic concepts employed to derive necessary and sufficient conditions for observability and stability, and the results discussed above are the followings.

First, the information transmission theorem is employed which states that the (Shannon) channel capacity must be bounded below by the information rate distortion for reliable data reconstruction up to the distortion value D_v, that is, $C \geq R(D_v)$, where C is the capacity and $R(D_v)$ is the rate distortion function. Second, the rate distortion, $R(D_v)$, with a given distortion measure (fidelity criterion) which is related to the observability and stability criteria, is considered. This step may be viewed as a generalization of the almost sure criterion which was first used in [9]. Finally, the Shannon lower bound $R_S(D_v)$ (which under certain conditions is equivalent to the rate distortion function) is employed to obtain the inequalities $\mathcal{C} \geq R(D_v) \geq R_S(D_v)$ as a necessary condition for observability and stability of the control/communication system of Figure 2.1. This result particularly compliments the results found in [9] and [22], in which for a specific dynamical system and stability criterion, the conditions for stability have been derived. The framework developed in this chapter is applicable to both linear and nonlinear systems; and the necessary conditions for observability and stability depend explicitly on the observability and stability criterion, hence different criterion corresponds to different necessary condition. The methodology described above has two practical advantages, namely, i) The Shannon lower bound is given in terms of the Shannon entropy rate of the source output and an additional term which depends on the type of observability and stability criteria (fidelity criteria) employed. ii) For certain fidelity criteria in the limit as $D_v \to 0$, the Shannon lower bound converges (weakly) to the rate distortion function, that is, $\lim_{D_v \to 0}\Big(R(D_v) - R_S(D_v)\Big) = 0$ [43], (Lemma 6.2.24,ii). The first is important in having a simple computation as Shannon entropy rate is often easily computed. The second is also of practical important because the rate distortion function can be rarely computed explicitly. Therefore, the second is particularly important when a tight necessary condition on the capacity is required. In this chapter, the importance of the second is illustrated by showing that for certain sources and channels the Shannon lower bound is a sufficient condition since there exists an encoder, decoder, and controller so that $\mathcal{C} = R(D_v) = R_S(D_v)$.

Throughout the chapter, we apply the obtained necessary conditions to specific systems described in time and frequency domains, and we relate the necessary conditions to the unstable eigenvalues of the open loop system. This method does not require a precise description of the encoder, decoder and controller.

The results of this chapter are applied to a linear stochastic partially observed control system subject to measurement noise when the channel is an Additive White Gaussian Noise (AWGN) channel. That is, we consider the following plant $\begin{cases} X_{t+1} = AX_t + NU_t + BW_t, \\ Y_t = CX_t + DG_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $Y_t \in \Re^d$, $W_t \in \Re^m$ is the process noise, $G_t \in \Re^l$ is the measurement noise, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, $B \in \Re^{q\times m}$, $C \in \Re^{d\times q}$, $D \in \Re^{d\times l}$, W_t i.i.d $\sim N(0, I_m)$, and G_t i.i.d. $\sim N(0, I_l)$, where $I_m \in \Re^{m\times m}$ and $I_l \in \Re^{l\times l}$ are identity matrices. Here, we derive

an encoder, decoder, and controller for mean-square stability and observability using the standard detectability and stabilizability assumptions of Linear Quadratic Gaussian (LQG) theory [23], without assuming knowledge of the control sequence at the encoder/decoder. Thus, necessary and sufficient conditions for mean-square stability and observability are derived. Moreover, a separation principle between the design of the control and communication systems is shown. This result extends the results of [11] which is concerned with the stability of the following fully observed system $\begin{cases} X_{t+1} = AX_t + NU_t + W_t, \\ Y_t = X_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $W_t \in \Re^q$, $Y_t \in \Re^q$, $A \in \Re^{q \times q}$, $N \in \Re^{q \times o}$, and W_t i.i.d. $\sim N(0, \Sigma_W)$, over AWGN channels.
This chapter is organized as follows. In Section 2.2, the problem formulation is given. Since Shannon lower bound is given in terms of Shannon entropy rate and an additional term, the Shannon entropy rate is calculated in Section 2.3. In Section 2.4, necessary conditions are derived for observability and stability of the control/communication system of Figure 2.1. Moreover, sufficient conditions are derived for a linear partially observed control Gaussian system which is controlled over AWGN channels. In this section, it is also shown that Shannon lower bound is the minimum capacity that guarantees mean square observability. Long proofs are given in Appendix 6.1. Since tools from probability theory, system and control, and information/communication theory are used throughout this chapter to derive new results, in Appendix 6.2 we summarized known concepts, measures, and results from probability theory, information theory and systems and control that we are going to use them throughout this chapter.

2.2 Problem Formulation

Throughout, sequences of Random Variables (R.V.'s) are denoted by $Y^T \triangleq (Y_0, Y_1, ..., Y_T)$ for $T \in \mathrm{N}_+ \triangleq \{0, 1, 2, ...\}$. log(.) and $\log_e(.)$ denote the logarithm of base 2 and natural logarithm, respectively.
Consider the control/communication system of Figure 2.1, where $Y_t \in \Re^d$, $Z_t \in \mathcal{Z}_t$, $\tilde{Z}_t \in \tilde{\mathcal{Z}}_t$, $\tilde{Y}_t \in \Re^d$, $U_t \in \Re^o$ are R.V.'s denoting the source message, channel input, channel output, reproduced source message, and the control input to the source, respectively, at time $t \in \mathrm{N}_+$. It is assumed that $\mathcal{Z}_t$, and $\tilde{\mathcal{Z}}_t$ are complete separable metric spaces denoting the channel input and the channel output codeword alphabet sets, respectively; and $(\mathcal{Z}_t, \mathcal{F}(\mathcal{Z}_t))$ and $(\tilde{\mathcal{Z}}_t, \mathcal{F}(\tilde{\mathcal{Z}}_t))$, $\tilde{\mathcal{Z}}_t \subseteq \mathcal{Z}_t$ are measurable spaces (e.g., $\mathcal{F}(\mathcal{Z}_t)$ is an σ-algebra of subsets of the set $\mathcal{Z}_t$ generated by closed set). For $T, n \in \mathrm{N}_+$, sequences of R.V.'s with length T and n of the

source and channel, are identified with the product measurable spaces.

The different blocks of Figure 2.1 are described below.

Information Source: The information source is the plant output (i.e., Y_t). Throughout this chapter, we are particularly interested to the following partially observed stochastic control system with input U_t and output Y_t.

$$\begin{cases} X_{t+1} = AX_t + NU_t + BW_t, & X_0 = X, \\ Y_t = H_t + DG_t, & H_t = CX_t, \end{cases} \tag{2.1}$$

where $X_t \in \Re^q$ is the unobserved (state) process, $Y_t \in \Re^d$ is the observed (measurement) process, $U_t \in \Re^o$ is the control signal, $H_t \in \Re^d$ is the signal to be controlled, $W_t \in \Re^m$, $G_t \in \Re^l$, 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, $X_0 \sim N(\bar{x}_0, \bar{V}_0)$ and $\{W_t, G_t, X_0; t \in \mathbf{N}_+\}$ are mutually independent.

Models similar to (2.1) have been considered in many places [4], [9]- [14], [16], [41] to derive various type of results, via information theoretic as well as non-information theoretic concepts. In [9] authors addressed the question of almost sure asymptotic observability and stability of a linear time-invariant noiseless system subject to the random initial condition. That is, they considered the following plant. $\begin{cases} X_{t+1} = AX_t + NU_t, & X_0 = X \\ Y_t = CX_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $Y_t \in \Re^d$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$ and $C \in \Re^{d\times q}$. In [11] authors studied stability of a fully observed Gauss-Markov system, that is, $\begin{cases} X_{t+1} = AX_t + NU_t + W_t, \\ Y_t = X_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $W_t \in \Re^q$, $Y_t \in \Re^q$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, and W_t i.i.d. $\sim N(0, \Sigma_W)$. In [41] authors studied bounded input, bounded output stability of a linear time-invariant system subject to the scalar Gaussian measurement noise, that is, $\begin{cases} X_{t+1} = AX_t + NU_t, \\ Y_t = CX_t + G_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re$, $Y_t \in \Re$, $G_t \in \Re$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times 1}$, and G_t i.i.d. $\sim N(0,1)$.

Communication Channel: The communication channels with input Z^n and output $\tilde{Z}^n$ is modeled by a stochastic kernel (Definition 6.2.1) $\{P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1}); t \in \mathbf{N}_+\}$.

Encoder: We define and discuss the following types of encoders.

Class A) The encoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(dZ_t; y^t)$.

Class B) The encoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(dZ_t; y^t, u^{t-1}, \tilde{z}^{t-1})$.
Class C) The encoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(dZ_t; y^t, \tilde{z}^{t-1})$.

As an example, consider an encoder of Class B. This encoder corresponds to an encoding law which by the knowledge of the control signals and the outputs of the channel up to time $t-1$; also by the knowledge of the source messages up to time t, it can produce the encoded information at time t. A deterministic encoder of Class B is described by a Dirac measure; and similarly for the other encoders in different classes.

Decoder: We define and discuss the following types of decoders.
Class A) The decoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(d\tilde{Y}_t; \tilde{z}^t)$.
Class B) The decoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(d\tilde{Y}_t; \tilde{z}^t, u^{t-1})$.

Controller: The control law at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(dU_t; \tilde{z}^t, u^{t-1})$.

Please note that the descriptions presented above for channel, encoder, decoder and controller, are similar to the descriptions presented in [11].
Observability and stability of the control/communication system of Figure 2.1 are defined next.

Definition 2.2.1 *(Observability in Probability and r-Mean). Consider the system of Figure 2.1.*
i) For a given $\delta \geq 0$ and $D_v \in [0,1]$, the plant is called (δ, D_v)- observable in probability if there exist an encoder and a decoder such that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \Pr(||Y_t - \tilde{Y}_t|| > \delta) \leq D_v, \tag{2.2}$$

where $||.||$ is the Euclidean norm on the space of $\Re^d$.
ii) For a given $r > 0$ and finite $D_v \geq 0$, the plant is called (r, D_v)-observable if $\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} E||Y_t - \tilde{Y}_t||^r \leq D_v$.

Definition 2.2.2 *(Stability in Probability and r-Mean). Consider the system of Figure 2.1 in which $Y_t = H_t + \Upsilon_t$, where H_t is the signal to be controlled and Υ_t represents the effects* *of the measurement noise.*

i) For a given $\delta \geq 0$ *and* $D_v \in [0,1]$*, the plant is called* (δ, D_v)*- stabilizable in probability if there exist a controller, encoder, and decoder such that*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \Pr(||H_t|| > \delta) \leq D_v. \tag{2.3}$$

ii) For a given $r > 0$ *and finite* $D_v \geq 0$*, the plant is* (r, D_v)*- stabilizable if* $\lim_{T\to 0} \frac{1}{T}\sum_{t=0}^{T-1} E|| H_t||^r \leq D_v$.

The above definitions of observability and stability which use sequences, are variants of the almost sure observability and stability criterion introduced in [9]. In [9] the following plant has been considered. $\begin{cases} X_{t+1} = AX_t + NU_t + W_t, \\ Y_t = X_t, \end{cases}$ where $X_t \in \Re^q$, $U_t \in \Re^o$, $W_t \in \Re^q$, $Y_t \in \Re^q$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, and W_t i.i.d. $\sim N(0, \Sigma_W)$. The observability and stability criterion considered in [9] are almost sure criteria described as follows $\Pr(\lim_{t\to\infty} ||Y_t - \tilde{Y}_t|| = 0) = 1$ (for observability) and $\Pr(\lim_{t\to\infty} ||X_t|| = 0) = 1$ (for stability). Although the proposed encoding/decoding scheme and the controller in [9] guarantee observability and stability in asymptotic sense, but the transition behavior can be poor since the criteria to address observability and stability are defined in asymptotic sense. More appropriate criteria for observability and stability are the ones defined over sequences as we defined in Definitions 2.2.1 and 2.2.2. The encoding scheme and stability scheme for observability and stability in the sense of Definitions 2.2.1 and 2.2.2, not only guarantee observability and stability in asymptotic sense, but also they guarantee smooth transition behavior.

Throughout this chapter, it is assumed that plant, communication channel, the types of encoder, decoder and controller are given. The objective is to design encoder, decoder and controller for observability and stability in probability and r-mean as defined in Definitions 2.2.1 and 2.2.2, when the capacity is as minimum as possible.

Throughout this chapter, we are particularly interested to the control/communication system of Figure 2.2 described by the controlled dynamical system (2.1) and the AWGN channel (i.e., $\tilde{Z}_t = Z_t + \tilde{W}_t$, where Z_t is the channel input, $\tilde{Z}_t$ is the channel output, and $\tilde{W}_t$ i.i.d. $\sim N(0, W_c)$ is the channel noise). In this control/communication system the encoder and the decoder are of Class B. Please note that the control/communication system of Figure 2.2 is a special case of the general control/communication system of Figure 2.1.

Next, consider the control/communication system of Figure 2.2. The observations process

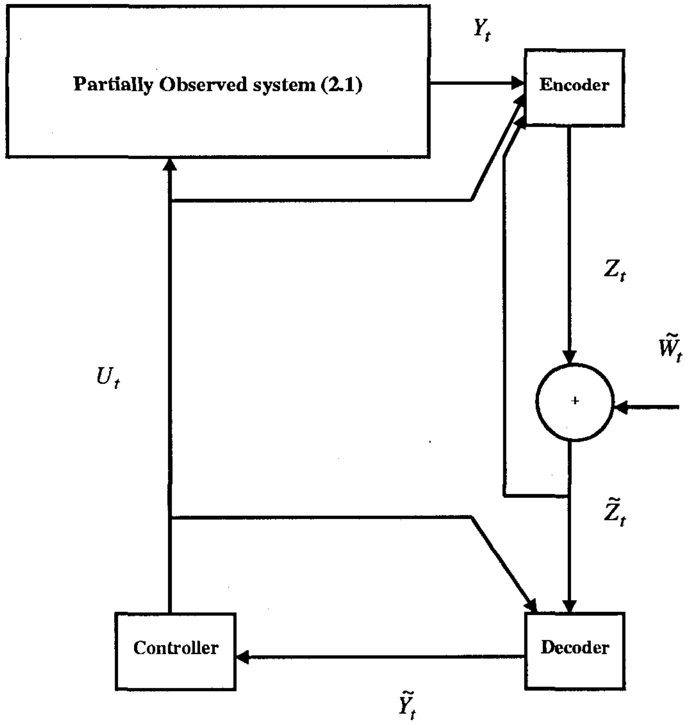

Figure 2.2: Discrete time control/communication system

of system (2.1) can be written as follow.

$$Y_t = CA^tX_0 + \sum_{i=0}^{t-1} CA^{t-1-i}BW_i + DG_t + \sum_{i=0}^{t-1} CA^{t-1-i}NU_i. \tag{2.4}$$

By the knowledge of the control sequence U^{t-1} at the encoder and the communication end, the last term in (2.4) (i.e., $\sum_{i=0}^{t-1} CA^{t-1-i}NU_i$) is reconstructed perfectly at the encoder and at the communication end. Thus, in the system of Figure 2.2, the problem of reliable data reconstruction of Y_t is equivalent to the problem of reliable data reconstruction of the remaining terms in (2.4). That is, $\bar{Y}_t$ given as follow.

$$\bar{Y}_t = CA^tX_0 + \sum_{i=0}^{t-1} CA^{t-1-i}BW_i + DG_t, \tag{2.5}$$

where, $\bar{Y}_t$ is the output of the following dynamical system

$$\begin{cases} X_{t+1} = AX_t + BW_t, & X_0 = X, \\ \bar{Y}_t = H_t + DG_t, & H_t = CX_t, \end{cases} \tag{2.6}$$

which is the uncontrolled analogous of system (2.1). Thus, in addressing the observability question of the control/communication system of Figure 2.2, without loss of generality, we can consider the uncontrolled dynamical system (2.6).

It is evident that, if there are no encoder and decoder that guarantee reliable data reconstruction of the observations process, then there are no encoder, decoder and stabilizing controller that stabilize the controlled dynamical system (2.1) because the stabilizing controller in system of Figure 2.2 is a feedback stabilizing controller, in which it can stabilize the unstable controlled dynamical system if only there is a reliable feedback loop. In other words, any necessary condition for observability of the control/communication system of Figure 2.2 is also a necessary condition for stability of this system. Thus, in addressing the necessary conditions for stability of the control/communication system of Figure 2.2, without loss of generality, we can also consider the uncontrolled dynamical system (2.6).

Following above discussion, in the control/communication system of Figure 2.2 by the knowledge of the control sequence in encoder and decoder, the observability problem of the observations process is equivalent to the observability problem of the observations process of the uncontrolled analogous system. Thus, when the observability Definition 2.2.1 is applied to this control/communication system, the observability in probability and r-mean involve the existence of the control sequence, encoder and decoder. Furthermore, the distributions of the control sequence does not contribute to the expectation (used in r-mean observability Definition 2.2.1, ii) and the probability criterion (2.2).

2.3 Shannon Entropy Rate and Kalman Filtering

The observability and stability of sequences as defined in Definitions 2.2.1 and 2.2.2 are addressed by finding the entropy rate of the observations process.

In this section, we calculate the Shannon entropy rate of the observations process of the uncontrolled version of system (2.1) (i.e., $\{U_t = 0; t \in \mathbb{N}_+\}$). Here, a connection is established between the error covariance of the Kalman filter, the innovations process of the source and Shannon entropy rate. The results of this section will be used in the next section to address the conditions for observability and stability of the control/communication system of Figure 2.2, defined by the dynamical system (2.1). The main results of this section which are concerned with the calculation of the Shannon entropy rate of the partially observed systems, are given in Lemma 2.3.2, Remark 2.3.3 and Corollary 2.3.4, in which this lemma, remark and corollary are new contributions of this chapter. We shall need the following Lemma which can be found in ([23], pp. 44).

Lemma 2.3.1 *Let* $\{Y_t; t \in \mathbf{N}_+\}$, $Y : \Omega \times \mathbf{N}_+ \rightarrow \Re^d$ *be a Gaussian process. Let* $\Gamma_{Y^{T-1}} \triangleq Cov[(Y_0' \ Y_1' \ ... \ Y_{T-1}')'] \neq 0$ *and* $K_t \triangleq Y_t - E[Y_t|\sigma\{Y^{t-1}\}]$, $\Lambda_t \triangleq Cov(K_t)$ *and* $\Lambda_\infty \triangleq \lim_{T\rightarrow\infty} \Lambda_T \neq 0$, *where* $\sigma\{Y^{t-1}\}-$ *denotes the* σ*-algebra of events generated by the sequence* Y^{t-1}. *Then, the Shannon entropy rate of* $\{Y_t; t \in \mathbf{N}_+\}$ *in bits per time step denoted by* $\mathcal{H}_S(\mathcal{Y})$ *is given by the following*

$$\begin{aligned}\mathcal{H}_S(\mathcal{Y}) &\triangleq \lim_{T\rightarrow\infty} \frac{1}{T} H_S(Y^{T-1}) = \frac{d}{2}\log(2\pi e) + \lim_{T\rightarrow\infty} \frac{1}{2T} \log\det \Gamma_{Y^{T-1}} \\ &= \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log\det \Lambda_\infty, \end{aligned} \tag{2.7}$$

where $H_S(Y^{T-1})$ *denotes the Shannon entropy of* Y^{T-1} *(Definition 6.2.9). Moreover, if* $\{Y_t; t \in \mathbf{N}_+\}$ *is an (asymptotic) stationary Gaussian process with power spectral density* $S_Y(e^{jw})$ *then* $\mathcal{H}_S(\mathcal{Y}) = \frac{d}{2}\log(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \log\det S_Y(e^{jw})dw$.

Proof: Follows for Szego limit formula and Cholesky decomposition (see, [23], pp. 44).

2.3.1 Uncontrolled Stochastic Dynamical Systems

Consider the uncontrolled version of system (2.1) (i.e., system (2.6)). Sufficient Conditions for existence of the limiting R.V. are given in terms of stabilizability (Definition 6.2.39) and detectability (Definition 6.2.40) of (2.6).

Lemma 2.3.2 *Consider the uncontrolled version of system (2.1) corresponding to* $\{U_t = 0; t \in \mathbf{N}_+\}$ *(i.e., system (2.6)), and assume* (C, A) *is detectable (Definition 6.2.40),* $(A, (BB')^{\frac{1}{2}}$ *) is stabilizable (Definition 6.2.39), and* $D \neq 0$.
Then, $\Lambda_\infty = CV_\infty C' + DD'$, *where* $V_t \triangleq E[\tilde{X}_t \tilde{X}_t']$, $\tilde{X}_t \triangleq X_t - E[X_t|\sigma\{\bar{Y}^{t-1}\}]$, $t \in \mathbf{N}_+$, *and* $V_\infty \triangleq \lim_{t\rightarrow\infty} V_t$ *is the unique positive semi-definite solution of the following Algebraic Riccati-equation*

$$V_\infty = AV_\infty A' - AV_\infty C'(CV_\infty C' + DD')^{-1}CV_\infty A' + BB'. \tag{2.8}$$

Moreover, $\mathcal{H}_S(\bar{\mathcal{Y}}) = \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log\det(CV_\infty C' + DD')$, *where* $\mathcal{H}_S(\bar{\mathcal{Y}})$ *is the Shannon entropy rate of the observations process of the uncontrolled system (2.6).*

Proof: Follows from Definition 6.2.44 and Lemma 2.3.1.
We have the following remark regarding the results of Lemma 2.3.2.

Remark 2.3.3 *i) Consider the scalar version of the uncontrolled system (2.6) with* $q = 1$ *and* $d = 1$*. Then, (2.8) can be solved explicitly and then substituted into the Shannon entropy rate* $\mathcal{H}_S(\bar{\mathcal{Y}})$ *found in Lemma 2.3.2 to obtain*

$$\mathcal{H}_S(\bar{\mathcal{Y}}) = \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log \Lambda_\infty \geq \frac{1}{2}\log(2\pi e D^2) + \max\{0, \log|A|\} \tag{2.9}$$

Moreover, the inequality in (2.9) holds with equality when $B = 0$*. Notice that (2.9) relates Shannon entropy rate and the system matrix* A*, via* $\max\{0, \log|A|\}$*.*

ii) Consider the uncontrolled system (2.6) with scalar observation and measurement noise (i.e., $d = l = 1$*) under detectability and stabilizability conditions of Lemma 2.3.2 when* $D \neq 0$*. Then,* $\mathcal{H}_S(\bar{\mathcal{Y}}) = \frac{1}{2}\log(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi}\log S_K(e^{jw})dw$*, where* $S_K(e^{jw})$ *is the power spectral density of the asymptotic stationary innovations process* $\{K_t = \bar{Y}_t - E[\bar{Y}_t|\sigma\{\bar{Y}^{t-1}\}]; t \in \mathbb{N}_+\}$*. From standard Kalman filtering equations (Definition 6.2.44) 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),* $\Delta_\infty = AV_\infty C'\Lambda_\infty^{-1}$ *and* $S(z) = 1 - C(zI_q - \tilde{A})^{-1}\Delta_\infty$ *(*$K(z)$*,* $W(z)$*,* $G(z)$ *are the z-transform of the innovations, process and measurement noises, respectively). Subsequently,* $S_K(e^{jw}) = M(e^{jw})M'(e^{-jw}) + |S(e^{jw})|^2D^2 \geq |S(e^{jw})|^2D^2$*; and* $\mathcal{H}_S(\bar{\mathcal{Y}}) =\geq \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log D^2 + \frac{1}{4\pi}\int_{-\pi}^{\pi}\log|S(e^{jw})|^2dw$*. 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 the state space form this system is represented by* $\begin{cases} S_{t+1} = AS_t + \Delta_\infty U_t, & U_t = -K_t \\ K_t = CS_t + DG_t \end{cases}$ *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* $\begin{cases} S_{t+1} = AS_t + \Delta_\infty U_t, \\ K_t = CS_t \end{cases}$ *Subsequently, from Bode integral formula [44] (Theorem 6.2.41), we have* $\int_{-\pi}^{\pi}|S(e^{jw})|^2dw = 4\pi\sum_{\{i;|\lambda_i(A)|\geq 1\}}\log|\lambda_i(A)|$ *, where* $\lambda_i(A)$*'s are the eigenvalues of the system matrix* A*. Thus,*

$$\mathcal{H}_S(\bar{\mathcal{Y}}) \geq \frac{1}{2}\log(2\pi e D^2) + \sum_{\{i;|\lambda_i(A)|\geq 1\}}\log|\lambda_i(A)| \tag{2.10}$$

Notice that (2.10) relates Shannon entropy rate and the unstable eigenvalues of the system matrix A*, via* $\sum_{\{i;\lambda_i(A)\geq 1\}}\log|\lambda_i(A)|$*.*

2.3.2 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 2.3.4 *Consider the controlled system (2.1) and assume (C, A) is detectable, $(A, (BB')^{\frac{1}{2}})$ is stabilizable, and $D \neq 0$. Then, the Shannon entropy rate of the controlled system (2.1) is bounded below by that of the uncontrolled system via*

$$\mathcal{H}_S(\mathcal{Y}) = \lim_{T\to\infty}\frac{1}{T}H_S(Y^{T-1}) = \lim_{T\to\infty}\frac{1}{T}\sum_{i=0}^{T-1}H_S(Y_i|Y^{i-1}) \geq \lim_{T\to\infty}\frac{1}{T}\sum_{i=0}^{T-1}H_S(Y_i|Y^{i-1}, U^{i-1})$$
$$= \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log\det\Lambda_\infty = \mathcal{H}_S(\bar{\mathcal{Y}}), \tag{2.11}$$

where, Y^{T-1} is a sequence with length T of the observations process associated with the controlled system (2.1), Λ_∞ is given in Lemma 2.3.2, $\mathcal{H}_S(\mathcal{Y})$ is the Shannon entropy rate of the observations process of system (2.1) and $\mathcal{H}_S(\bar{\mathcal{Y}})$ is the Shannon entropy rate of the observations process of the uncontrolled system (2.6).

Proof: The proof is original and it is given in Appendix.

2.4 Conditions for Observability and Stability

In this section, we state necessary conditions for observability and stability in terms of Shannon entropy rate and the observability and stability criteria considered, which can be applied to variety of sources and channels. We then derive sufficient conditions for mean square observability and stability of the controlled system when the communication channel is an AWGN channel. Here, we illustrate the Shannon lower bound is equivalent to the rate distortion, verifying the tightness of the bound. Moreover, we present an encoder and a decoder so that the source is matched to the channel, generalizing the results found in [11] to the partially observed system.

2.4.1 Rate Distortion Solution and Information Transmission Theorem

In order to address sufficient conditions for observability and stability, we need to compute the rate distortion function for a given source distribution and distortion measure. On the other hand, the necessary conditions are obtained by applying the information transmission theorem which relates the channel capacity (per source message) to the rate distortion function. The next remark identifies a sufficient condition so that the minimizing stochastic kernel of the rate distortion function $R_T(D_v)$ (Definition 6.2.22) is a causal operation on the source sequence, and subsequently the rate distortion is sequential (causal) rate distortion (Definition 6.2.23), as introduced in [11], [45]. This sufficient condition which is given in the following lemma, is a new contribution of this chapter.

Lemma 2.4.1 *Let $\{K_t \in \Re^d; t \in \mathrm{N}_+\}$ be an independent stochastic process. Consider the single letter mean square distortion measure $\rho_T(k^{T-1}, \tilde{k}^{T-1}) = \frac{1}{T}\sum_{t=0}^{T-1} \rho(k_t, \tilde{k}_t)$, where $\tilde{K}_t \in \Re^d$ is the reconstruction of K_t and $\rho(k_t, \tilde{k}_t) : \Re^d \times \Re^d \to [0, \infty)$ is continuous and non-negative. Then, the minimizing stochastic kernel $P^*(d\tilde{K}^{T-1}; k^{T-1})$ of the rate distortion function $R_T(D_v)$ (Definition 6.2.22) is a causal operation on the source sequence and it is given by*

$$P^*(d\tilde{K}^{T-1}; k^{T-1}) = \prod_{t=0}^{T-1} P^*(d\tilde{K}_t; k_t) = \prod_{t=0}^{T-1} \frac{e^{\frac{1}{T}s(D_v)\rho(k_t,\tilde{k}_t)} P(d\tilde{K}_t)}{\int_{\Re^d} e^{\frac{1}{T}s(D_v)\rho(k_t,\tilde{k}_t)} P(d\tilde{K}_t)}, \tag{2.12}$$

where $s(D_v) = \frac{dR_T(D_v)}{dD_v}$.

Proof: The proof follows from (6.112) and by taking into account that K_t is an independent process.

Next, in the following theorem we present a necessary condition for end to end transmission up to distortion value $D_v \geq 0$. Here, we assume the Markovian property which is equivalent to the conditional independence of the various blocks of system of Figure 2.1. 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 $P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1}, y^{T-1}) = P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1}), P - a.s$ $(t \in \{0, 1, ..., n-1\})$, and similarly for the rest of the blocks. In the following theorem, it is also assumed that the channel is a Discrete Memoryless Channel (DMC) or AWGN channel (Definition 6.2.17, v). For DMC's or AWGN channel, without feedback, the information channel capacity C (Definition 6.2.18) represents the operational capacity. Furthermore, the capacity of those channels with feedback which is described by the directed information [46] is the same as the capacity without feedback [40], [47]. Examples in which feedback increases the capacity are given in [48]. However, our source model (2.1) and Gaussian channel considered in subsequent sections does not give rise to feedback that increases the capacity. Rather, in our context feedback is used to achieve the capacity of the channel and to stabilize the plant.

Theorem 2.4.2 *(Information Transmission Theorem) Consider the control/communication system of Figure 2.1 under the conditional independence assumption. A necessary condition on the capacity for n channel uses, i.e.,* $\mathcal{C}_n$ *(Definition 6.2.18) for reproducing a sequence of source messages* Y^{T-1} *up to distortion value* D_v *by* $\tilde{Y}^{T-1}$ *at the output of the decoder (i.e.,* $E\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v$*, where* $\rho_T(.,.)$ *is the distortion measure) using a sequence of the channel inputs and channel outputs with length n* $(T \leq n)$*, is given as follow*

$$\mathcal{C}_n \geq R_T(D_v), \tag{2.13}$$

where $R_T(D_v)$ *is the rate distortion function (see Definition 6.2.22).*

Proof: The proof is the variant of the one presented for the classical information transmission theorem in ([49], pp. 72) and it follows from the data processing inequality (Remark 6.2.14, ii). The detail of the proof is given in Appendix.

2.4.2 Necessary Conditions for Observability and Stability for General Systems

For the general control/communication system of Figure 2.1, the main theorem which connects capacity, observability and stability is given next. This is applied to the system of Figure 2.2 described by system (2.1). The following theorem is a direct result of Theorem 2.4.2; and it is obtained following the same methodology used in [9] and [11] by implementing a slight modification.

Theorem 2.4.3 *Consider the system of Figure 2.1 under conditional independence assumption in which* $Y_t \in \Re^d$ *is the observed process. Assume the Shannon entropy rate corresponding to the observed process exists and it is finite.*

Then, for observability in probability (Definition 2.2.1, i), a necessary condition on the channel capacity is

$$\mathcal{C} \;\geq\; \mathcal{H}_S(\mathcal{Y}) - \frac{1}{2}\log[(2\pi e)^d \det \Gamma_g] \triangleq R_S(D_v), \tag{2.14}$$

where $\mathcal{H}_S(\mathcal{Y})$ *is the Shannon entropy rate of the observed process,* $R_S(D_v)$ *is the Shannon lower bound, and* Γ_g *is the covariance matrix of the Gaussian distribution* $h^*(\xi) \sim$

$N(0,\Gamma_g), (\xi \in \Re^d)$ *which satisfies*

$$\int_{||\xi||>\delta} h^*(\xi)d\xi = D_v. \tag{2.15}$$

Moreover, a necessary condition on the channel capacity for r-mean observability (Definition 2.2.1, ii) is

$$\mathcal{C} \geq \mathcal{H}_S(\mathcal{Y}) - \log e^{\frac{d}{r}} + \log(\frac{r}{dV_d\Gamma(\frac{d}{r})}(\frac{d}{rD_v})^{\frac{d}{r}}) \triangleq R_S(D_v), \tag{2.16}$$

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

Furthermore, for the case where the observed process Y_t *and the signal to be controlled* H_t *are related by* $Y_t = H_t + \Upsilon_t$*, (2.14) and (2.16) are also necessary conditions for stability in probability and r-mean (Definition 2.2.2).*

Proof: See Appendix.

Remark 2.4.4 *We have the followings remarks regarding the results of Theorem 2.4.3.*

i) The capacity used in (2.14) and (2.16) are measured in bits per source message; or bits per time (since each message is generated in each time step). Consequently, this capacity is related to the transmission data rate, $\mathcal{R}$*. For AWGN, erasure, and digital noiseless channels, the direct relationship between this capacity and transmission bit rate has been illustrated in Example 6.2.19.*

ii) The conditions (2.14) and (2.16) are new. If the capacity is less than the bounds (2.14) and (2.16), there are no encoding and/or stabilizing schemes for observability and/or stability of sequences related to the dynamical system. That is, bounds (2.14) and (2.16) are the fundamental limits on the Shannon capacity for observability and/or stability of sequences.

iii) The lower bounds (2.14) and (2.16) given in Theorem 2.4.3 hold for any observed process no matter what the information patterns for encoder, decoder and controller are. Hence, when Theorem 2.4.3 is applied to the controlled system (2.1), then the Shannon entropy rate of the controlled output of system (2.1) must be used. However, under assumption that (C,A) *is detectable,* $(A,(BB')^{\frac{1}{2}})$ *is stabilizable, and* $D \neq 0$ *when the encoder and decoder are of Class A, by Corollary 2.3.4 we deduce that bounds (2.14) and (2.16) also hold when the Shannon entropy rate is replaced by the Shannon entropy rate of the observations process of the uncontrolled version of system (2.1), which has been calculated in Lemma 2.3.2. Following* *discussion of Section 2.2, for the control/communication system of Figure 2.2 described by the encoder/decoder of type B, the Shannon entropy rate of the observations process of the uncontrolled version of system (2.1) which has been calculated in Lemma 2.3.2, can be used in the lower bounds (2.14) and (2.16) to address a necessary condition for observability and stability.*

iv) The condition (2.14) and (2.16) are given in terms of the Shannon entropy rate which can be easily computed. Further, by Remark 2.3.3, these conditions result in a lower bound on the channel capacity in terms of the eigenvalue rate (i.e., $\sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$*) which has appeared in the literature [4]-[13].*

v) For the case of $d = 1$*, condition (2.15) is reduced to*

$$2\Phi(-\frac{\delta}{\sqrt{\Gamma_g}}) = D_v, \quad \Phi(t) \triangleq \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du. \tag{2.17}$$

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*. Furthermore, using a* Γ_g *smaller than* $\frac{\delta^2}{16}$ *does not yield significantly different result for observability and stability performance. So, for a small quantity of* D_v, $\Gamma_g = \frac{\delta^2}{16}$ *can be used in (2.14). Further, for* $d = 1$ *and* $r = 2$*, the extra term* $-\log e^{\frac{d}{r}} + \log(\frac{r}{dV_d\Gamma(\frac{d}{r})}(\frac{d}{rD_v})^{\frac{d}{r}})$ *in (2.16) 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 observability and stability.*

vi) As it will be shown in Section 2.4.4 for the case of $r = 2$*, the lower bound (2.16) is a tight bound (i.e., a necessary and sufficient condition on channel capacity) for observability of the innovations as well as the observations process over AWGN channels. For the fully observed analogous system of (2.1), under similar conditions, a tight bound is given in [11].*

vii) In [4], it has been concluded that a single number like Shannon capacity is not an adequate measure to present conditions for moment observability and stability of linear time-invariant systems controlled over noisy channels. In fact, as it has been shown in [16] for moment stability of linear stochastic systems over noisy channels, a tight bound on Shannon capacity for stability in addition to the eigenvalue rate also involves an additional term which depends on the order of moment stability. Similarly, the lower bound (2.16) (in which, it can be a tight bound as discussed in Remark 2.4.4, vi), in addition to the eigenvalue rate (which is resulted from the entropy rate) also involves an additional term dependent on the order of moment stability and observability. The tightness of the bound (2.16) under certain conditions (as discussed in Remark 2.4.4, vi) overcomes the drawback of using Shannon capacity *in describing conditions for moment observability.*

viii) (2.14) and (2.16) are similar to the condition presented in [14], in which a condition for stability of deterministic nonlinear systems in terms of topological entropy rate has been found. However, since here we deal with stochastic dynamical systems, the statistical information theoretic measures such as Shannon entropy rate have been used.

2.4.3 Necessary Conditions for Stability of Partially Observed Linear Gaussian Systems

In this section, we employ the conclusion of Theorem 2.4.3 to find necessary conditions for stability of the following special case of system (2.1)

$$\begin{cases} X_{t+1} = AX_t + NU_t + BW_t, & X_0 \sim N(\bar{x}_0, \bar{V}_0), \\ Y_t = CX_t + DG_t, \end{cases} \tag{2.18}$$

in which, $X_t \in \Re^q$, $W_t \in \Re^m$, $Y_t \in \Re, U_t \in \Re$, and $G_t \in \Re^l$. This result is given in the following corollary, in which it presents a new necessary condition for stability of the control/communication system of Figure 2.3.

Corollary 2.4.5 *Consider the system (2.18) over a linear time-invariant, single-input and single-output AWGN channel (see Figure 2.3). That is, in compact notation, $\tilde{Y}(z) = Y(z) + \tilde{W}(z)$, where $\tilde{W}(z)$ is the z- transform of the channel noise $\{\tilde{W}(t) \in \Re; t \in \mathbf{N}_+\}$ which is a Gaussian process noise with mean zero and variance W_c and it is mutually independent of $\{W_t, G_t, X_0; t \in \mathbf{N}_+\}$. Also, assume the controller is stable linear time-invariant (e.g., the controller transfer function $-K_c(z)$ has poles inside the unit circle) and the open loop transfer function $L(z) = P(z)K_c(z)$, where $P(z) = C(zI_q - A)^{-1}N$, is strictly proper (i.e., $P(z) = \frac{N(z)}{D(z)}$, where the degree of polynomial $D(z)$ is greater than the degree of polynomial $N(z)$).*

Then, a necessary condition for stability in r-mean for the observed sequence $\{Y_t; t \in \mathbf{N}_+\}$ (e.g., $\lim_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} E||X_k||^r_{C'C} \leq D_v$, $r > 0$) is

$$\begin{aligned} \mathcal{C} \geq \mathcal{H}_S(\mathcal{Y}) - \Delta = & \sum_{\{i; |\lambda_i(A)| \geq 1\}} \log|\lambda_i(A)| + \frac{1}{2}\log(2\pi e) \\ & + \frac{1}{4\pi}\int_{-\pi}^{\pi} \log(F(e^{jw})F'(e^{-jw}) + DD' + L(e^{jw})L'(e^{-jw})W_c)dw - \Delta, \end{aligned} \tag{2.19}$$

where $F(e^{jw}) = C(e^{jw}I_q - A)^{-1}B$, $L(e^{jw}) = P(e^{jw})K_c(e^{jw})$, and $\Delta = \log e^{\frac{1}{r}} - \log\left(\frac{r}{2\Gamma(\frac{1}{r})}(\frac{1}{rD_v})^{\frac{1}{r}}\right)$ bits per time step.

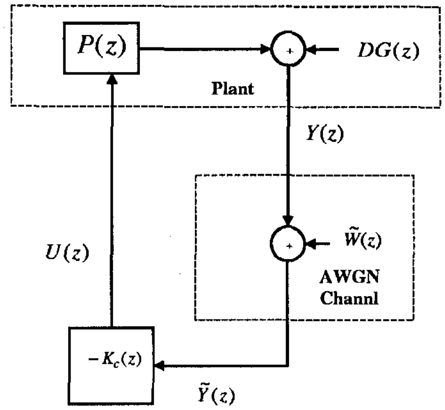

Figure 2.3: Control/commmunication system

Proof: The proof is original and it is given in Appendix.

Remark 2.4.6 *i) We can view the observed process* $\{Y_t; t \in \mathbf{N}_+\}$ *as a consequence of passing the Gaussian process* $\{E_t = \mathcal{Z}^{-1}[F(z)W(z) + DG(z) - L(z)\tilde{W}(z)]; t \in \mathbf{N}_+\}$ *(*$\mathcal{Z}^{-1}[.]$ *denotes the inverse z-transform) through a stable linear filter with transfer function* $S(z) = \frac{1}{1+L(z)}$. *Subsequently, from Lemma 6.2.43, it follows that*

$$\mathcal{H}_S(\mathcal{Y}) = \bar{\mathcal{H}}_S(\mathcal{E}) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|S(e^{jw})|dw = \bar{\mathcal{H}}_S(\mathcal{E}) + \sum_{\{i;\lambda_i(A)\geq 1\}} \log|\lambda_i(A)|, \tag{2.20}$$

where $\bar{\mathcal{H}}_S(\mathcal{E})$ *is the Shannon conditional entropy rate of the process* $\{E_t; t \in \mathbf{N}_+\}$ *(Definition 6.2.11). Subsequently, comparing this result with (2.19), it follows that*

$$\bar{\mathcal{H}}_S(\mathcal{E}) = \frac{1}{2}\log(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \log(F(e^{jw})F^{'}(e^{-jw}) + DD^{'} + L(e^{jw})L^{'}(e^{-jw})W_c)dw. \tag{2.21}$$

Thus, the lower bound (2.19) is also given by $C \geq \sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)| + \bar{\mathcal{H}}_S(\mathcal{E}) - \Delta$.
ii) Condition (2.19) gives as special case the result derived in [41] in which a digital noiseless channel (e.g., $C = \mathcal{R}$*), with sufficiently large transmission data rate* $\mathcal{R}$ *is used. This follows from the result of Corollary 2.4.5, condition (2.19) by letting* $r = 2$ *and setting* $D = 1$ *and* $B = 0$ *in the controlled system (2.18) and* $W_c = 0$ *for the channel. Substituting these values in (2.19), we have*

$$\mathcal{R} \geq \frac{1}{2}\log(2\pi e) + \sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)| - \frac{1}{2}\log(2\pi e D_v), \quad \textit{(bits/time step)}. \tag{2.22}$$

Clearly, (2.22) is the condition derived in [41] in which a uniform quantizer is used.

2.4.4 Design of Communication System for the Controlled System

One of the main goals of this section is to show that under certain conditions, the necessary conditions of Theorem 2.4.3 can be also sufficient conditions. Here, we consider a single letter mean square distortion criterion with distortion value D_v (i.e., $E\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v$, $\rho_T(y^{T-1}, \tilde{y}^{T-1}) = \frac{1}{T}\sum_{t=0}^{T-1} ||y_t - \tilde{y}||^2$) for reliable data reconstruction of the controlled system (2.1) over AWGN channels. We design an encoder of Class B (resp. Class C) and decoder of Class B (resp. Class A) to guarantee reliable data reconstruction when the capacity (measured in bits per time step) is as minimum as possible for such reliable data reconstruction. Moreover, we show a separation principle between the design of the control and communication systems. The problem of reliable data reconstruction and stability of the partially observed system (2.1) over AWGN channels has not been considered elsewhere. So, in this section new encoding/decoding scheme and stability method are proposed for stability of the control/communication systems. Please note that in [11], the problem of reliable data reconstruction and stability of the fully observed version of system (2.1) has been addressed. We start with the case of $Y_t \in \Re$. The general case is treated similarly.

The case of $Y_t \in \Re$

Consider the control/communication system of Figure 2.4 which is the system of Figure 2.1 described by the stochastic control system (2.1) when $Y_t \in \Re$ and the encoder and the decoder are of Class B, and the communication channel is the following AWGN channel.

$$\tilde{Z}_t = Z_t + \tilde{W}_t, \quad \tilde{W}_t \text{ orthogonal} \sim N(0, W_c), \; Z_t \in \Re, \; E[Z_t^2] \leq P_t < \infty, \tag{2.23}$$

where $Z_t \in \Re$ is the channel input, $\tilde{Z}_t \in \Re$ is the channel output, $\tilde{W}_t \in \Re$ is the channel noise and P_t represents the power constraint. In this control/communication system α_t and γ_t (see Figure 2.4) are non-negative scalars.

Encoder and Decoder: The encoder consists of a pre-encoder which produces the orthogonal Gaussian innovations process $\{K_t; t \in \mathbf{N}_+\}$; $K_t \triangleq Y_t - E[Y_t|\sigma\{\tilde{K}^{t-1}, U^{t-1}\}] = Y_t - CE[X_t|\sigma\{\tilde{K}^{t-1}, U^{t-1}\}] = Y_t - C\hat{X}_t$ ($\hat{X}_t \triangleq E[X_t|\sigma\{\tilde{K}^{t-1}, U^{t-1}\}]$) using feedback channel information $\tilde{K}^{t-1}$ ($\tilde{K}_t = \gamma_t\tilde{Z}_t$) and the previous control sequence U^{t-1} (see Figure 2.4).

By the knowledge of the control sequence and the channel outputs at the decoder, $\hat{X}_t$ can be produced at the decoder. That is, at each instance of time t, $\hat{X}_t$ is known for the decoder. Therefore, since at each time instant t, $\hat{X}_t$ is also known at the encoder, from relation

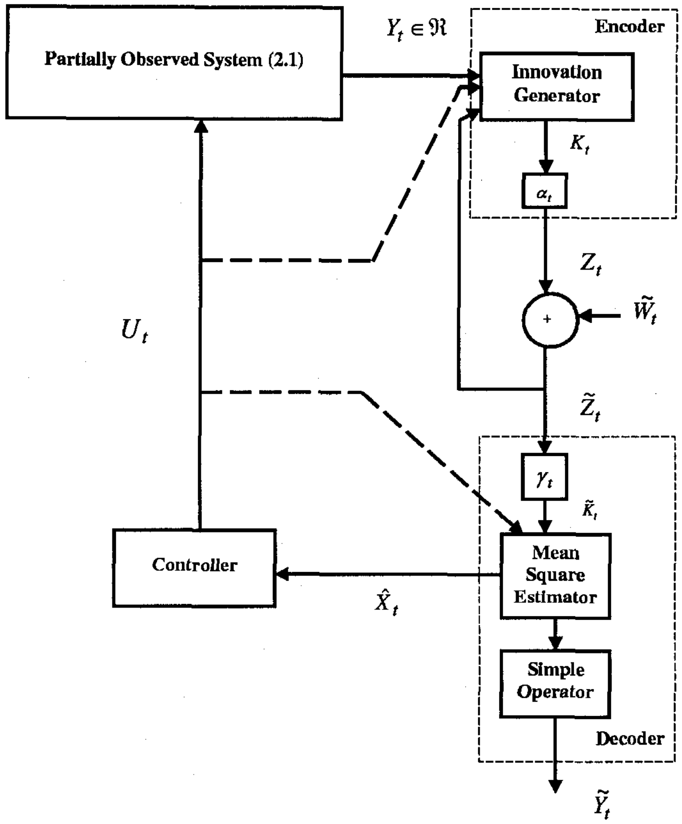

Figure 2.4: Discrete time control/communication system

$K_t = Y_t - C\hat{X}_t$ (or equivalently $Y_t = K_t + C\hat{X}_t$) follows that the problem of reliable data reconstruction of Y_t is equivalent to the problem of reliable data reconstruction of K_t which is easier to work with. Therefore, we design an encoder and a decoder that guarantee mean square observability of the innovations process K_t.

Here, we use the source-channel matching technique [50] which is a well developed technique in the information theory literature. This technique allows us to design an encoding scheme for reliable data reconstruction in the absence of complexity and delay which is essential in the control applications. Source channel matching technique states that when a memoryless source (i.e., a source producing i.i.d. messages) is probabilistically matched to a memoryless channel, that is, the source message to reconstruction message behaves like the minimizing kernel (of the rate distortion function of the memoryless source), then the optimal cost-distortion trade-off, is achieved in the absence of complexity and delay.
Roughly speaking, the source-channel matching technique requires that the stochastic kernel describing message to reconstruction message to be the same as the minimizing kernel of the rate distortion function. As discussed in [50], implementing this technique re-

sults in a joint source-channel encoding/decoding scheme which is different from common used encoding/decoding schemes. Common used encoding/decoding schemes are resulted from separation principle, i.e., the strategy of splitting the encoding into two stages, source encoder/decoder (e.g., quantizer/de-quantizer) and channel encoder/decoder (e.g., turbo codes). As discussed in [50] such strategy results in complexity and delay which are not suitable for control applications.

The encoder and decoder for the mean square observability of the innovations process K_t which is an independent process are $Z_t = \alpha_t K_t$ and $\tilde{K}_t = \gamma_t \tilde{Z}_t$, respectively, where α_t and γ_t are non-negative scalars to be determined so that the link from K_t to $\tilde{K}_t$ (i.e., the reconstruction of K_t) is matched to the minimizing kernel.

Define the mean square state estimator by $\hat{X}_t \triangleq E[X_t|\sigma\{\tilde{K}^{t-1}, U^{t-1}\}]$; then it is given by the following recursive Kalman filter

$$\hat{X}_{t+1} = A\hat{X}_t + \frac{1}{\alpha_t \gamma_t} A\Pi_t C'(C\Pi_t C' + DD' + \frac{W_c}{\alpha_t^2})^{-1}\tilde{K}_t + NU_t, \ \hat{X}_0 = \bar{x}_0 = E[X_0], \quad (2.24)$$

where the control is of the form $U_t = \mu(t, \tilde{K}^{t-1}, U^{t-1})$ ($\mu(\cdot)$ will be defined shortly) and Π_t is the mean square state estimation error (i.e., $\Pi_t \triangleq E(X_t - \hat{X}_t)^2$) which satisfies the following recursive equation.

$$\Pi_{t+1} = A\Pi_t A' - A\Pi_t C'(C\Pi_t C' + DD' + \frac{W_c}{\alpha_t^2})^{-1} C\Pi_t A' + BB' \ \ \Pi_0 = \bar{V}_0. \quad (2.25)$$

Next, consider the independent innovations process $K_t \triangleq Y_t - E[Y_t|\sigma\{\tilde{K}^{t-1}, U^{t-1}\}], t \in \mathbf{N}_+$ corresponding to the control system (2.1) with $U_t = \mu(t, \tilde{K}^{t-1}, U^{t-1})$. At each time instant t, K_t is a Gaussian R.V. with mean zero and variance $\Psi_t = C\Pi_t C' + DD'$. That is, $K_t \sim N(0, \Psi_t)$. The minimizing stochastic kernel of the rate distortion $R_T(D_v)$ (Definition 6.2.22) with single letter mean square distortion measure for a sequence with length T of the innovations process when the distortion value D_v satisfies $D_v < \min_{t \in \mathbf{N}_+} \Psi_t$, is given by the following.

$$P^*(d\tilde{K}^{T-1}; k^{T-1}) = \prod_{t=0}^{T-1} q^*_{\tilde{K}_t|K_t} d\tilde{K}^{T-1}, \ \ q^*_{\tilde{K}_t|K_t} \sim N(\eta_t k_t, \eta_t D_v), \eta_t \triangleq 1 - \frac{D_v}{\Psi_t}. \quad (2.26)$$

Then, the corresponding information rate distortion function is the causal rate distortion function and it is given by $R_T(D_v) = \frac{1}{2}\sum_{t=0}^{T-1} \log \frac{\Psi_t}{D_v}$, $D_v < \min_{t \in \mathbf{N}_+} \Psi_t$.

Next, consider the control/communication system of Figure 2.4. The stochastic kernel describing K^{T-1} and $\tilde{K}^{T-1}$ is given by the following.

$$P(d\tilde{K}^{T-1}; k^{T-1}) = \prod_{t=0}^{T-1} q_{\tilde{K}_t|K_t} d\tilde{K}^{T-1}, \ \ q_{\tilde{K}_t|K_t} \sim N(\alpha_t \gamma_t k_t, \gamma_t^2 W_c). \quad (2.27)$$

Subsequently, a matched communication link (i.e., a communication channel equipped with an encoder and decoder such that the innovations process-to-reconstruction behaves like the rate distortion minimizing stochastic kernel) is obtained if $\alpha_t = \sqrt{\frac{\eta_t W_c}{D_v}}$ and $\gamma_t = \sqrt{\frac{D_v \eta_t}{W_c}}$ (please note that for this α_t and γ_t, $q^*_{\tilde{K}_t|K_t} = q_{\tilde{K}_t|K_t}$). The power constraint corresponding to this encoding scheme is $E[Z_t^2] = \alpha_t^2 \Psi_t = \frac{\eta_t W_c}{D_v}\Psi_t \triangleq P_t < \infty$. Subsequently, for the distortion value $D_v < \min_{t\in\mathbf{N}_+} \Psi_t$, we have the followings

$$\begin{aligned}
1 + \frac{P_t}{W_c} = 1 + \frac{\eta_t \Psi_t}{D_v} &= 1 + \frac{(1 - \frac{D_v}{\Psi_t})\Psi_t}{D_v} \\
&= 1 + \frac{\Psi_t - D_v}{D_v} \\
&= \frac{\Psi_t}{D_v} \\
\log(1 + \frac{P_t}{W_c}) &= \log \frac{\Psi_t}{D_v} \qquad (2.28)
\end{aligned}$$

Consequently for $\tilde{Z}_t = Z_t + \tilde{W}_t$, $Z_t = \alpha_t K_t$, the capacity for T channel uses i.e., $\mathcal{C}_T$ (see Definition 6.2.18) corresponding to T time steps; and subsequently the capacity $\mathcal{C}$ (measured in bits per time step) are given as follows.

$$\begin{aligned}
\mathcal{C}_T \triangleq \sup_{\{P(dZ^{T-1}); E[Z_t^2] \leq P_t, t\in\{0,1,\ldots,T-1\}\}} I(Z^{T-1}; \tilde{Z}^{T-1}) &= \frac{1}{2}\sum_{t=0}^{T-1} \log(1 + \frac{P_t}{W_c}) \\
&= \frac{1}{2}\sum_{t=0}^{T-1} \log \frac{\Psi_t}{D_v} = R_T(D_v) \\
\mathcal{C} \triangleq \lim_{T\to\infty} \frac{1}{T}\mathcal{C}_T &= \frac{1}{2}\log\frac{\Psi_\infty}{D_v} = R(D_v),
\end{aligned} \qquad (2.29)$$

where $R(D_v)$ is the rate distortion function (see Definition 6.2.22) of the innovation process and $\Psi_\infty = C\Pi_\infty C' + DD'$ where under the assumption that (C, A) is detectable (Definition 6.2.40) and $(A, (BB')^{\frac{1}{2}})$ is stabilizable (Definition 6.2.39), then $\Pi_\infty \triangleq \lim_{T\to\infty} \Pi_T$ exists and it is given as the solution of the following equation

$$\begin{aligned}
\Pi_\infty &= A\Pi_\infty A' - A\Pi_\infty C'(C\Pi_\infty C' + DD' + \frac{D_v}{\eta_\infty})^{-1} C\Pi_\infty A' + BB', \\
\eta_\infty &= 1 - \frac{D_v}{C\Pi_\infty C' + DD'}.
\end{aligned} \qquad (2.30)$$

Furthermore, corresponding to this encoding/decoding scheme, we have an end to end transmission with distortion

$$E(K_t - \tilde{K}_t)^2 = E(K_t - \eta_t K_t - \gamma_t \tilde{W}_t)^2 = (1-\eta_t)^2 E[K_t^2] + \gamma_t^2 E[\tilde{W}_t^2]$$

$$\begin{aligned}
&= (1-1+\frac{D_v}{\Psi_t})^2\Psi_t + \frac{D_v\eta_t}{W_c}W_c \\
&= \frac{D_v^2}{\Psi_t^2}\Psi_t + \eta_t D_v \\
&= \frac{D_v^2}{\Psi_t} + (1-\frac{D_v}{\Psi_t})D_v \\
&= \frac{D_v^2}{\Psi_t} + D_v - \frac{D_v^2}{\Psi_t} \\
&= D_v. \qquad (2.31)
\end{aligned}$$

Subsequently, using this encoding scheme for the distortion value $D_v < \min_{t\in\mathbf{N}_+}\Psi_t$, we have mean square reliable data reconstruction of K_t by $\tilde{K}_t$, that is, $E(K_t-\tilde{K}_t)^2 = D_v$, $\forall t\in\mathbf{N}_+$ and thus we have a mean square observability of the innovations process in the form of $\frac{1}{T}\sum_{t=0}^{T-1}E(K_t-\tilde{K}_t)^2 = D_v$, $\forall T\geq 1$, over the communication channel (2.23), when the capacity is $\mathcal{C}=R(D_v)$ bits in each time step.

Furthermore, 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 $E(Y_t-\tilde{Y}_t)^2 = E(K_t-\tilde{K}_t)^2 = D_v$, $\forall t\in\mathbf{N}_+$; and subsequently we have mean square observability of the observations process (i.e., $\frac{1}{T}\sum_{t=0}^{T-1}E(Y_t-\tilde{Y}_t)^2 = D_v$, $\forall T\geq 1$).

Separation Principle: Here, we shall show that the proposed 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 $\tilde{K}_t = \gamma_t\tilde{Z}_t, Z_t = \alpha_t K_t$ when $U_t = 0$, and $K_t^0 \triangleq Y_t - C\hat{X}_t$ is the innovations process when $U_t = 0$. 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 ([23], 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_t. Consequently, the encoder can be of Class C, while the decoder of Class A, and thus in Figure 2.4, 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.

Computation of Shannon Lower Bound: Following the minimizing kernel (2.26), it has been shown that the rate distortion function of the innovations process is given by $R(D_v) =$

$\frac{1}{2}\log\frac{\Psi_\infty}{D_v}$, when the distortion value D_v satisfies $D_v < \min_{t\in\mathbf{N}_+}\Psi_t$. For the same distortion measure, as above, the Shannon lower bound (i.e., $R_S(D_v)$, see Lemma 6.2.24) of the innovations process is given by

$$R(D_v) \geq R_S(D_v) = \mathcal{H}_S(\mathcal{K}) - \max_{h\in G_D} H_S(h) = \lim_{T\to\infty}\frac{1}{T}\sum_{i=0}^{T-1} H_S(K_i|K^{i-1}) - \max_{h\in G_D} H_S(h)$$
$$= \frac{1}{2}\log\Psi_\infty - \frac{1}{2}\log D_v, \tag{2.32}$$

where $\mathcal{H}_S(\mathcal{K})$ is the Shannon entropy rate of the innovations process and $\max_{h\in G_D} H_S(h)$ has been described in Lemma 6.2.24.

Subsequently, for the distortion value $D_v < \min_{t\in\mathbf{N}_+}\Psi_t$, the Shannon lower bound is equivalent to rate distortion $R(D_v)$.

Control Law: Suppose in addition to the previous assumptions for existence of the rate distortion function, $((C'C)^{\frac{1}{2}}, A)$ is detectable and (A, N) is stabilizable. Then the control law which minimizes the following LQG cost functional $\lim_{T\to\infty}\frac{1}{T}E\sum_{t=0}^{T-1}(||X_t||^2_{C'C} + ||U_t||^2_H)$ ($H \in \Re^{o\times o} > 0$) is given by $U_t = -\Delta\hat{X}_t$, where $\Delta = (H + N'P_\infty N)^{-1}N'P_\infty A$ and P_∞ is the unique positive semi-definite solution of the following Algebraic Riccati equation $P_\infty = A'P_\infty A - A'P_\infty N(H + N'P_\infty N)^{-1}N'P_\infty A + C'C$ (Definition 6.2.45). This shows that via the above feedback channel information the sequence $H^{T-1} = (H_0, ..., H_t, ..., H_{T-1})$, $H_t = CX_t$, is mean-square stable.

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.

Remark 2.4.7 *i) In this section we used feedback channel and source-channel matching technique for reliable data reconstruction and subsequently stability of a controlled dynamical system over a limited capacity communication channel. Using feedback channel we can achieve the capacity by implementing a codeword with finite length. Furthermore, using source-channel matching technique we can achieve the rate distortion function when the reconstruction is instantaneous (i.e.,* $E(Y_t - \tilde{Y}_t)^2 = E(K_t - \tilde{K}_t)^2 = D_v$, $\forall t \in \mathbf{N}_+$*).*

ii) In this section, for a given single letter mean square distortion criterion with distortion value D_v*, we have proposed an encoding scheme and a stability scheme for mean square reliable data reconstruction and stability by transmitting* $\mathcal{C} = R_S(D_v)$ *bits per time step, where* $\mathcal{C}$ *is the capacity and* $R_S(D_v)$ *is the Shannon lower bound of the innovations process. Following the necessary condition (2.16) (applied to the innovations process), the proposed* *encoding/decoding scheme guarantees reliable data reconstruction of the innovations process by transmitting minimum capacity* $\mathcal{C} = R_S(D_v)$ *where the distortion value* D_v *satisfies* $D_v < \min_{t\in\mathbf{N}_+} \Psi_t$. *From (2.30) and (2.8) follows that for* D_v *sufficiently small (e.g.,* D_v *is mush smaller than* $C\Pi_\infty C' + DD'$*) we have* $\Pi_\infty \approx V_\infty$*; and thus* $\mathcal{H}_S(\mathcal{K}) = \mathcal{H}_S(\bar{\mathcal{Y}})$, *where* $\mathcal{H}_S(\mathcal{K}) = \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log\det(C\Pi_\infty C' + DD')$ *is the Shannon entropy rate of the innovations process and* $\mathcal{H}_S(\bar{\mathcal{Y}}) = \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log(CV_\infty C' + DD')$ *is the Shannon entropy rate of the observations process of the uncontrolled analogous system. Subsequently, the Shannon lower bound of these two processes are the same (for* D_v *sufficiently small); and thus following Remark 2.4.4,iii, the rate* $\mathcal{C} = R_S(D_v)$ *is the minimum capacity for mean square reliable data reconstruction of the innovations as well as observations process (when* D_v *is sufficiently small). Therefore, for the partially observed controlled system (2.1) over AWGN channels subject to the single letter mean square distortion criterion with distortion value* D_v *(sufficiently small), using the proposed encoding scheme and stability scheme, the mean square stability in the form of* $\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1} E||X_t||^2_{C'C} \leq D_v^c$, $D_v^c \geq D_v^*$, *(*D_v^* *is the value of the cost functional when* $U_t = -\Delta\hat{X}_t$ *and* H *is small) is achieved by transmitting minimum capacity which reliably reconstructs the observations process.*

iii) The case of $D_v \approx 0$ *corresponds to the case where the transmission power* P_t *and subsequently capacity are unlimited. For the distortion value* $D_v \approx 0$, $Y_t = \tilde{Y}_t$ *and thus the encoder, channel and decoder in Figure 2.4 can be replaced by a straight line. Subsequently, the recursive estimator (2.24) and the proposed certainly equivalent controller are reduced to the standard LQG results (Definition 6.2.45) [23]. Thus, our proposed stability scheme and encoding scheme extend the standard LQG results to the cases where the observed process* Y_t *is transmitted over an AWGN channel subject to the limited transmission power and subsequently limited capacity.*

iv) The problem of mean square state estimation over communication channels has been also considered in [24], in which the goal is the successful state estimation over a noisy channel subject to loss when there are no restriction on the capacity. In this section, one of the goals is to have a successful state estimation over noisy channel (2.23) subject to capacity constraint. Therefore, in the presence of limited capacity constraint, the design of the mean square state estimator must be done based on the Kalman filter (2.24) and (2.25). Regardless, of the difference in the problem formulation considered in this chapter (i.e., restriction on capacity) and [24] (i.e., no restriction on capacity), the conditions relating the parameters *of the communication channel to the dynamical system for successful state estimation, are given in terms of the unstable eigenvalues of the open loop system.*

v) The mean square observability scheme and stability scheme proposed in this section can be used to address observability and stability questions of the systems described in Section 1.1 when these systems are described by the partially observed system (2.1) and AWGN channels. In these applications, it is assumed that the desired stability criterion (i.e., D_v^c *in Remark 2.4.7, ii) is given. Desired stability criterion determines the acceptable distortion value* D_v *for reconstruction of the observations process* Y_t *(following the relation* $D_v^c = D_v^*$*, where* D_v^* *is the value of the cost functional when* $U_t = -\Delta\hat{X}_t$ *and* H *is small). Subsequently, by implementing the proposed encoding scheme the mean square observability up to the admissible distortion value* D_v *is achieved, while the capacity is as minimum as possible and the stability is being guaranteed using the proposed controller. Please note that following our discussion in Remark 2.4.7, ii, for a given partially observed system with a given admissible distortion value* D_v*, if we have* $\mathcal{H}_S(\mathcal{K}) \approx \mathcal{H}_S(\bar{\mathcal{Y}})$ *(where* $\mathcal{H}_S(\mathcal{K})$ *is the Shannon entropy rate of the innovations process and* $\mathcal{H}_S(\bar{\mathcal{Y}})$ *is the Shannon entropy rate of the observations process of the uncontrolled analogous system), then the reliability of the observations process is guaranteed, while the capacity is minimum for this reliable reconstruction. Nevertheless, if the entropy rate of the innovations process is much bigger than the entropy rate of the observations process of the uncontrolled analogous system, then reliability of the observations process is guaranteed by using (possibly) more capacity than the minimum capacity required for such reconstruction. However, for this case the capacity is minimum for reliable reconstruction of the innovations process.*

The Case of $Y_t \in \Re^d$

In the previous section, we have proposed an encoding scheme and a control scheme which guarantee mean square observability and stability of system (2.1) when $Y_t \in \Re$. In this section, we intend to extend these results to the case where $Y_t \in \Re^d$.

Here, we are concerned with the following AWGN channel

$$\tilde{Z}_t = Z_t + \tilde{W}_t, \quad \tilde{W}_t \text{ orthogonal} \sim N(0, W_c),$$
$$Z_t = [Z_{t1} \ \dots \ Z_{td}]' \in \Re^d \quad E[Z_{ti}^2] \le P_{ti} < \infty, \quad 1 \le i \le d \tag{2.33}$$

where $W_c \in \Re^{d \times d}$ is diagonal (i.e., $W_c = diag(W_1 \ \dots \ W_d)$). The encoder is of Class B (resp. Class C) and the decoder is of Class B (resp. Class A). The encoder consists of a pre-encoder which produces the innovations process $K_t = Y_t - C\hat{X}_t \in \Re^d \sim N(0, \Psi_t)$, where $\hat{X}_t = E[X_t|\sigma\{\tilde{K}^{t-1}, U^{t-1}\}]$ and $\Psi_t = C\Pi_t C' + DD'$ ($\Pi_t \in \Re^{d\times d}$ will be defined shortly). Then, the encoder pre-processes the innovations process by applying E_t' ($E_t \in \Re^{d\times d}$ and $'$ denotes matrix transpose) and $\mathcal{A}_t = diag(\sqrt{\frac{\eta_{t1}W_1}{\frac{D_v}{d}}} \ldots \sqrt{\frac{\eta_{td}W_d}{\frac{D_v}{d}}})$ to produce $Z_t = \mathcal{A}_t E_t' K_t$, where $diag(\ldots)$ denotes the diagonal matrix, D_v is the distortion value and it is such that $\frac{D_v}{d} < \min_{i\in\{1,2,\ldots,d\}} \lambda_{ti}$, $\forall t \in \mathbf{N}_+$, $\eta_{ti} = 1 - \frac{D_v}{d\lambda_{ti}}$ and E_t is the unitary matrix that diagonalizes Ψ_t, i.e., $\Psi_t = E_t\Sigma_t E_t'$, $\Sigma_t = diag(\lambda_{t1} \ldots \lambda_{td})$. The power constraint associated with this encoding scheme is $E[Z_{ti}^2] = \frac{\eta_{ti}W_i}{\frac{D_v}{d}}\lambda_{ti} = P_{ti}$, $\forall 1 \leq i \leq d$. On the other hand, the decoder multiplies the channel outputs by $\mathcal{B}_t = diag(\sqrt{\frac{D_v}{d}\frac{\eta_{t1}}{W_1}} \ldots \sqrt{\frac{D_v}{d}\frac{\eta_{td}}{W_d}})$ and then by E_t to produce the corresponding reproduced innovation process $\tilde{K}_t = E_t\mathcal{B}_t\tilde{Z}_t$. Subsequently, using this encoding scheme, it is shown that (for $\frac{D_v}{d} < \min_{i\{1,2,\ldots,d\}} \lambda_{ti}$, $\forall t \in \mathbf{N}_+$), the capacity for T channel uses is given by the following.

$$\begin{aligned} \mathcal{C}_T &= \sup_{\{P(dZ^{T-1});E[Z_{ti}^2]\leq P_{ti},\forall 0\leq t\leq T-1,\ 1\leq i\leq d\}} I(Z^{T-1};\tilde{Z}^{T-1}) \\ &= \frac{1}{2}\sum_{t=0}^{T-1}\sum_{i=1}^{d}\log(1+\frac{P_{ti}}{W_i}) = \frac{1}{2}\sum_{t=0}^{T-1}\log\frac{\det\Sigma_t}{\frac{D_v}{d}} = R_T(D_v), \end{aligned} \tag{2.34}$$

where $R_T(D_v)$ is the rate distortion function of the innovations process K^{T-1} with single letter mean square distortion measure. Furthermore,

$$\begin{aligned} E||K_t - \tilde{K}_t||^2 &= E||(I_d - \mathcal{A}_t\mathcal{B}_t)E_t'K_t - \mathcal{B}_t\tilde{W}_t||^2 \\ &= \sum_{i=1}^{d}\frac{D_v^2}{d^2\lambda_{ti}} + \sum_{i=1}^{d}(1-\frac{D_v}{d\lambda_{ti}})\frac{D_v}{d} \\ &= D_v, \end{aligned} \tag{2.35}$$

where $I_d \in \Re^{d\times d}$ is the identity matrix.

Consequently $\frac{1}{T}\sum_{t=0}^{T-1} E||K_t - \tilde{K}_t||^2 = D_v$, $\forall T \geq 1$, and thus $\frac{1}{T}\sum_{t=0}^{T-1} E||Y_t - \tilde{Y}_t||^2 = D_v$, $\forall T \geq 1$, where $\tilde{Y}_t = \tilde{K}_t + C\hat{X}_t$ in which $\hat{X}_t$ is given by the following recursive Kalman filter

$$\hat{X}_{t+1} = A\hat{X}_t + A\Pi_tC'(C\Pi_tC' + DD' + (E_t\mathcal{A}_t^{-1})W_c(E_t\mathcal{A}_t^{-1})')^{-1}(E_t\mathcal{B}_t\mathcal{A}_tE_t')^{-1}\tilde{K}_t + NU_t, \tag{2.36}$$

where $\hat{X}_0 = \bar{x}_0$ and Π_t is the mean square state estimation error which can be determined from the following recursive equation

$$\Pi_{t+1} = A\Pi_tA' - A\Pi_tC'(C\Pi_tC' + DD' + (E_t\mathcal{A}_t^{-1})W_c(E_t\mathcal{A}_t^{-1})')^{-1}C\Pi_tA' + BB', \tag{2.37}$$

where $\Pi_0 = \bar{V}_0$.

Next, along the same lines of previous part it is shown that separation holds and the Shannon lower bound of the innovations process and the observations process of the uncontrolled analogous system, are the same (for sufficiently small distortion value D_v).

From Lemma 6.2.24, ii follows that the Shannon lower bound is equivalent to the rate distortion as the distortion value D_v tends to zero. Therefore, for sufficiently small D_v under assumption that (C, A) is detectable and $(A, (BB^{\frac{1}{2}}))$ is stabilizable, we have the following result for channel capacity $\mathcal{C}$.

$$\mathcal{C} = \lim_{T\to\infty} \frac{1}{2T} \sum_{t=0}^{T-1} \log \frac{\det \Sigma_t}{\frac{D_v}{d}} = \frac{1}{2} \log \frac{\det \Sigma_\infty}{\frac{D_v}{d}} = R(D_v) = R_S(D_v), \tag{2.38}$$

where $R(D_v)$ and $R_S(D_v)$ are the rate distortion and the Shannon lower bound of the innovations process, respectively, Σ_∞ is the diagonalized version of $\Psi_\infty = C\Pi_\infty C' + DD'$ (i.e., $\Psi_\infty = E_\infty \Sigma_\infty E'_\infty$), and Π_∞ is the solution of the Algebraic Riccati equation associated with the Riccati equation (2.37).

Furthermore, under the assumption that $((C'C)^{\frac{1}{2}}, A)$ is detectable and (A, N) is stabilizable, the mean-square stabilizing controller in which it minimizes the cost functional $\lim_{T\to\infty} \frac{1}{T} E \sum_{t=0}^{T-1}$
$(||X_t||^2_{C'C} + ||U_t||^2_H)$ $(H > 0)$, is given by $U_t = -\Delta \hat{X}_t$ where $\Delta = (H + N'P_\infty N)^{-1} N'P_\infty A$ and P_∞ is the unique positive semi-definite solution of the following Algebraic Riccati equation $P_\infty = A'P_\infty A - A'P_\infty N(H + N'P_\infty N)^{-1} N'P_\infty A + C'C$.

2.5 Conclusion

In this chapter, using the information theoretic tools and by invoking the information transmission theorem and the Shannon lower bound, we introduced a set of necessary conditions for observability and stability of sequences. The necessary conditions were given in the form of a lower bound on the Shannon capacity in terms of the Shannon lower bound. Under certain conditions these necessary conditions were also sufficient. Shannon lower bound are given in terms of Shannon entropy rate of the inputs to the encoder and an additional term which is related to the observability and stability criteria. Subsequently, Shannon lower bound depends on the observability and stability criteria; and thus it has been concluded that the Shannon capacity is still an adequate measure for describing conditions for moment observability and stability. Shannon lower bound appears to be quite nice to use. In Remark

2.3.3, it has been shown that for linear time-invariant systems, Shannon entropy rate and subsequently Shannon lower bound is related to $\sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$. Consequently, conditions on channel capacity in terms of Shannon lower bound result in the well known eigenvalue rate condition for linear-time invariant systems (i.e., $\mathcal{C} \geq \sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)| + F$; $F \geq 0$). Moreover, Shannon entropy rate is applicable to time domain (Section 2.3) and frequency domain (Section 2.4.3); and subsequently, Shannon lower bound is also applicable to time domain and frequency domain. Furthermore, Shannon lower bound can be easily computed. In addition, since it is given in terms of the Shannon entropy rate which can be calculated for both linear and nonlinear systems, Shannon lower bound can be also used to address the observability and stability question of nonlinear dynamical systems.

Chapter 3

Control of Continuous Time Systems

3.1 Introduction

Since a discrete time model is more consistent with today's digital communication links, the focus on the observability and stability subject to limited capacity constraint has been on the observability and/or stability of discrete time systems which are controlled over a discrete time communication channel with finite capacity [1]-[22], [42]. Nevertheless, in some applications, analog modulation schemes may be interesting due to simplicity in building such schemes. Furthermore, having a complete theory which deals with continuous time systems will help us gain additional insight and understanding into building reliable data reconstruction and controlling schemes when the subsystems are discrete time.

In this chapter, we are concerned with the mean square observability and stability of the control/communication system of Figure 3.1 described by a linear continuous time system driven by Brownian motion (Definition 6.2.5) over a flat fading continuous time AWGN channel. The results of this chapter is the continuous version of [11] and [42]. In [11] and [42] authors considered the following dynamical system $\begin{cases} X_{t+1} = AX_t + NU_t + W_t, \\ Y_t = X_t, \end{cases}$ (where $X_t \in \Re^q$, $U_t \in \Re^o$, $W_t \in \Re^q$, $Y_t \in \Re^q$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, and W_t i.i.d. $\sim N(0, \Sigma_W)$) which is controlled over the following AWGN channel $\tilde{Z}_t = Z_t + \tilde{W}_t$; $\tilde{W}_t$ orthogonal $\sim N(0, W_c)$. In [11] and [42] authors have proposed a stabilizing scheme which minimizes a Linear Quadratic Gaussian (LQG) cost functional and guarantees the mean square stability, by transmitting $\mathcal{C} = \sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$ where $\mathcal{C}$ is the capacity of the AWGN channel and $\lambda_i(A)$'s are the eigenvalues of the system matrix A.

The main contribution of this chapter is the following. The continuous version of the eigenvalue rate condition (i.e., $\mathcal{C} \geq \sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)|$ [4]-[13]) is derived for continuous time systems which are controlled over continuous time AWGN channels. The obtained condition is described by the summation of the real parts of the unstable eigenvalues of the open loop time-invariant system (i.e., $\mathcal{C} \geq \sum_{\{i;Re(\lambda_i(A))\geq 0\}} Re(\lambda_i(A))$, where $Re(.)$ is an operator which returns the real parts of the complex numbers).

The steps taken to obtain above contribution are described below.

First we consider the problem of stability of the control/communication system of Figure 3.1 described by a fully observed linear continuous time-invariant noiseless plant which is controlled over a continuous time AWGN channel. By implementing Bode integral formulae, a necessary condition for existence of a stabilizing controller is found in terms of the summation of the real parts of the unstable eigenvalues of the open loop system. Then, we consider a fully observed linear continuous time-varying stochastic Gaussian system driven by Brownian motion which is controlled over a flat fading continuous time AWGN channel. Here, we assume complete knowledge of the channel throughout the transmission, at the transmitter and the receiver ends. That is, the encoder and decoder know the Channel State Information (CSI) (i.e., $\theta(t) \in \Re^g$), in which at each instant of time t is a vector including the information about transmission such as amount of attenuation on the amplitude of the received signal, Doppler shift, and phase difference. Under this assumption, we derive optimal encoding and decoding strategies which minimize the mean square decoding error, and achieve the channel capacity. We further show that under certain conditions, the proposed encoding and decoding strategies yield mean square observability. Furthermore, using this encoding scheme and standard LQG results, a mean square stability scheme is proposed for time-invariant system.

Following these constructions, it is concluded that the summation of the real parts of the unstable eigenvalues of the open loop time-invariant system is the minimum capacity under which there is an encoding scheme for mean square observability of fully observed linear time-invariant Gaussian systems driven by Brownian motion, over AWGN channels. Further, by transmitting with a rate equal to the summation of the real parts of the unstable eigenvalues of the open loop system, a mean square stability is possible. Due to similarity between the conditions found for observability and stability and the eigenvalues rate condition, it is concluded that the summation of the real parts of the unstable eigenvalues of the open loop continuous time-invariant systems is the continuous version of the eigenvalue rate condition for continuous time-invariant systems.

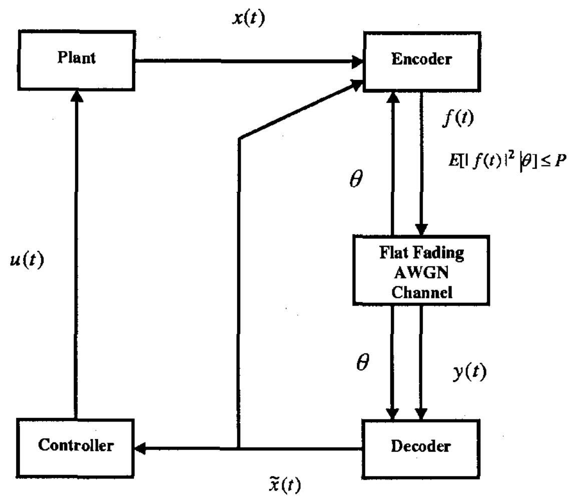

Figure 3.1: Continuous time control/communication system

This chapter is organized as follows. In Section 3.2, the problem formulation is given. In Section 3.3, a necessary condition is presented for stability of the control/communication system of Figure 3.1 when it is described by a linear continuous time-invariant noiseless plant which is controlled over a continuous time AWGN channel. In Section 3.4, the optimal encoding/decoding scheme that ensures mean square observability is given. Finally, in Section 3.5, a stabilizing control scheme is proposed. Long proofs are given in Appendix 6.1. Since tools from probability theory, information/communication theory, and systems and control theory are going to be used throughout this chapter to derive new results, in Appendix 6.2, we summarized these results.

3.2 Problem Formulation

Consider the block diagram of Figure 3.1. As in any typical communication system, the source which corresponds to the controlled plant output is communicated via a flat fading wireless AWGN channel. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ (Definition 6.2.3) and time $t \in [0, T]$, $T > 0$. Let $x \triangleq \{x(s) \in \Re^q; 0 \leq s \leq t \leq T\}$ denote the output of the controlled plant (transmitted signal), $u \triangleq \{u(s) \in \Re^o; 0 \leq s \leq t \leq T\}$ the control signal, $y \triangleq \{y(s) \in \Re; 0 \leq s \leq t \leq T\}$ the output of the communication channel, $\theta \triangleq \{\theta(s) \in \Re^g; 0 \leq s \leq t \leq T\}$ the channel state information (which is assumed to be known throughout transmission), $v \triangleq \{v(s) \in \Re; 0 \leq s \leq t \leq T\}$ the channel noise,

$w \triangleq \{w(s) \in \Re^m; 0 \leq s \leq t \leq T\}$ the plant process noise, and $\tilde{x} \triangleq \{\tilde{x}(s) \in \Re^q; 0 \leq s \leq T\}$ the decoder output. Denote the fading process by $z \triangleq \{z(s, \theta(s)) \in \Re, 0 \leq s \leq T\}$. The plant noise w, and the channel noise v are independent standard Browning motions ($Ev^2(t) = N_0 t, Cov(w(t)) = I_m.t$), which are independent of the initial state $x(0)$. Let $\{\mathcal{F}^x_{0,t}\}_{t\geq 0}$, $\{\mathcal{F}^{\tilde{x}}_{0,t}\}_{t\geq 0}$, $\{\mathcal{F}^y_{0,t}\}_{t\geq 0}$, $\{\mathcal{F}^u_{0,t}\}_{t\geq 0}$ and $\{\mathcal{F}^\theta_{0,t}\}_{t\geq 0}$ denote the complete filtration generated by $\mathcal{F}^x_{0,t} \triangleq \sigma\{x(s); 0 \leq s \leq t\}$, $\mathcal{F}^{\tilde{x}}_{0,t} \triangleq \sigma\{\tilde{x}(s); 0 \leq s \leq t\}$, $\mathcal{F}^y_{0,t} \triangleq \sigma\{y(s); 0 \leq s \leq t\}$, $\mathcal{F}^u_{0,t} \triangleq \sigma\{u(t); 0 \leq s \leq t\}$ and $\mathcal{F}^\theta_{0,t} \triangleq \sigma\{\theta(s); 0 \leq s \leq t\}$, respectively, which are sub-sigma fields of $\{\mathcal{F}_t\}_{t\geq 0}$. Here, $\mathcal{F}^x_{0,t}$, $\mathcal{F}^{\tilde{x}}_{0,t}$, $\mathcal{F}^y_{0,t}$, $\mathcal{F}^u_{0,t}$ and $\mathcal{F}^\theta_{0,t}$ are the Borel σ-algebras on the space of continuous functions $C([0,T]; \Re^q), C([0,T]; \Re^q), C([0,T]; \Re)$, $C([0,T], \Re^m)$ and $C([0,T]; \Re^g)$ respectively.

Next, the blocks of Figure 3.1 are defined as follows.

Plant. The state of the plant is described by a continuous time, controlled diffusion process given by the *Itô* equation

$$dx(t) = A(t)x(t)dt + N(t)u(t)dt + G(t)dw(t),\ x(0), \tag{3.1}$$

where $A : [0,T] \to \Re^{q\times q}$, $N : [0,T] \to \Re^{q\times o}$, and $G : [0,T] \to \Re^{q\times m}$ and $x(0)$ is Gaussian random variable $x(0) \sim N(\bar{x}_0, \bar{V}_0)$, which is independent of standard Brownian motion w. The control u is also $\{\mathcal{F}_{0,t}\}_{t\geq 0}$ adapted (Definition 6.2.4); and $A(t)$, $N(t)$ and $G(t)$ are uniformly bounded.

Encoder. The encoder f is a non-anticipative (i.e., causal) functional of the state of the plant x, the decoder output $\tilde{x}$, and the channel state information θ. Define $\mathcal{F}^{x,\tilde{x},\theta}_{0,t} \triangleq \mathcal{F}^x_{0,t} \vee \mathcal{F}^{\tilde{x}}_{0,t} \vee \mathcal{F}^\theta_{0,t}$. The set of admissible encoders is defined by $\mathcal{F}_{ad} \triangleq \Big\{f : [0,T]\times C([0,T]; \Re^q) \times C([0,T]; \Re^q) \times C([0,T]; \Re^g) \to \Re; f$ is $\{\mathcal{F}^{x,\tilde{x},\theta}_{0,t}\}_{t\geq 0}$ adapted; and $E[|f(t,x,\tilde{x},\theta)|^2 \big| \mathcal{F}^\theta_{0,t}] \leq P\Big\}$. For simplicity in notation, sometimes we may denote the encoder by $f(t)$.

Remark 3.2.1 *Please notice that the hard power constraint $E\Big[|f(t,x,\tilde{x},\theta)|^2 \Big| \mathcal{F}^\theta_{0,t}\Big] \leq P$ on the transmitted signal is the main difference between the problem considered in this chapter and the standard LQG results (Definition 6.2.46). The aim of this chapter is to relate this power constraint to the parameters of system (3.1) for observability and stability. Throughout, this relation is given in the form of necessary conditions and sufficient conditions for observability and stability, in which these conditions are new.*

Channel. The communication channel is a flat fading continuous time AWGN wireless channel whose output y is defined by the following stochastic differential equation

$$dy(t) = z(t,\theta(t))f(t,x,\tilde{x},\theta)dt + dv(t), \quad y(0)=0, \ 0 \leq t \leq T, \tag{3.2}$$

where v is the standard Brownian motion with $Ev^2(t) = N_0 t$. Throughout, we shall assume that (3.2) has a unique strong solution [51], and that $\int_0^T E[z(t,\theta(t))f(t,x,\tilde{x},\theta)]^2 < \infty$, for finite T. Further, we shall assume that $\liminf_{T\to\infty} \frac{1}{2T}\int_0^T E[z^2(t,\theta(t))]dt$ is finite. If $\lim_{t\to\infty} E[z^2(t,\theta(t))]$ exists then lim inf can be replaced by lim.

Decoder. The decoder map $\{y(s), u(s), z(s,\theta(s))\} \to \tilde{x}(t,y,\theta)$ ($= \tilde{x}(t)$ in compact notation) is adapted to $\{\mathcal{F}_{0,t}^{y,u,\theta}\}_{t\geq 0}$, where $\mathcal{F}_{0,t}^{y,u,\theta} \triangleq \mathcal{F}_{0,t}^{y} \vee \mathcal{F}_{0,t}^{u} \vee \mathcal{F}_{0,t}^{\theta}$. The set of admissible decoders is denoted by $\mathcal{D}_{admi}$. The decoder plays the role of a state estimator.

Controller. The controller u is a non-anticipative functional of the output of the decoder and the channel state information (e.g., u is $\{\mathcal{F}_{0,t}^{y,\theta}\}_{t\geq 0}$ adapted). The set of admissible controller is denoted by $\mathcal{U}_{ad}$.

The objective is to design encoder, decoder and controller for mean square observability and stability of system (3.1) over communication channel (3.2), defined as follows.

Definition 3.2.2 *(Bounded Asymptotic and Asymptotic Observability in the Mean Square Sense). Define* $\mathcal{E}(t) \triangleq E[(x(t)-\tilde{x}(t,y,u,\theta))'(x(t)-\tilde{x}(t,y,u,\theta))\big|\mathcal{F}_{0,t}^{y,\theta}]$. *System (3.1), (3.2) is bounded asymptotically (resp. asymptotically) observable, in the mean square sense, if there exists an encoder* $f \in \mathcal{F}_{ad}$, *and decoder* $\tilde{x} \in \mathcal{D}_{admin}$ *such that* $\lim_{t\to\infty} \mathcal{E}(t) < \infty$, *a.s. (resp.* $\lim_{t\to\infty} \mathcal{E}(t) = 0$, *a.s.).*

Definition 3.2.3 *(Bounded Stability and Stability in the Mean Square Sense). Define* $||x||_Q^2 \triangleq x'Qx, x \in \Re^q, Q \in \Re^{q\times q}, Q = Q' > 0$. *System (3.1), (3.2) is bounded stabilizable (resp. stabilizable) in the mean square sense, if there exists a controller, encoder and decoder, such that* $\lim_{T\to\infty} \frac{1}{T}\int_0^T E[||x(t)||_Q^2\big|\mathcal{F}_{0,t}^{\theta}]dt < \infty$, *a.s.,(resp.* $\lim_{T\to\infty} \frac{1}{T}\int_0^T E[||x(t)||_Q^2\big|\mathcal{F}_{0,t}^{\theta}]dt = 0$, *a.s).*

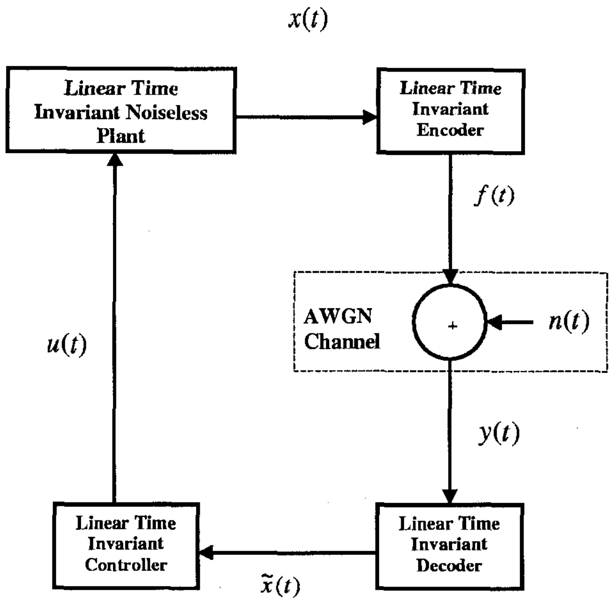

Figure 3.2: Control/communication system over AWGN

3.3 Necessary Condition for Stability

In this section, we consider the time-invariant noiseless version of system (3.1) in which, $u(t) \in \Re$. That is, the plant is given by

$$\dot{x}(t) = Ax(t) + Nu(t), \quad x(0) \in \Re^q, \quad u(t) \in \Re, \quad x(0) \sim N(\bar{x}_0, \bar{V}_0). \tag{3.3}$$

The communication channel is a continuous time AWGN channel (i.e., (3.2) when $z = 1$). That is, the output of the channel is given by

$$y(t) = f(t) + n(t), \; y(0) \in \Re, \qquad E[|f(t)|^2] \leq P \tag{3.4}$$

where n is a Gaussian white noise process with power spectral density $S_n(w) = N_0$, f is a stochastic process with power spectral density $S_f(w)$ in which it is $\mathcal{F}^x_{0,t}$ adapted (Definition 6.2.4), subject to instantaneous power constraint and it is such that $dy(t) = f(t)dt + d\int_0^t n(s)ds$ has a unique strong solution. Here, it is assumed that (f, n) are independent and encoder, decoder and controller are linear time-invariant (see Figure 3.2). Please note that the Brownian motion and the white noise are related by $v(t) = \int_0^t n(s)ds$ with $E[v(t)v(s)] = N_0 \min(t, s)$.

The mutual information between the state of plant x and the channel output y is given by

$$I_T(x;y) \triangleq E_{x,y}\Big[\log_e\big(\frac{P(dy;x)}{P(dy)}\big)\Big] \tag{3.5}$$

where $\log_e(.)$ denotes natural logarithm ,$E_{x,y}[.]$ denotes expectation with respect to sample paths x and y, $P(dy;x)$ is the stochastic kernel between sample paths y and x, $P(dy)$ is the probability distribution of sample path y, and $\frac{P(dy;x)}{P(dy)}$ is Radon-Nikodym derivative. Subsequently, the channel capacity is given by [52]

$$\mathcal{C} \triangleq \lim_{T\to\infty} \sup_{(x,f)\in\mathcal{X}\times\mathcal{F}_{ad}} \frac{1}{T} I_T(x;y) \tag{3.6}$$

where $\mathcal{X}$ is the set of all possible continuous sample paths x's.

Let $\lambda_i(A)$'s denote the eigenvalues of A. We shall show that $\mathcal{C} \geq \sum_{\{i;Re(\lambda_i(A))\geq 0\}} Re(\lambda_i(A))$ is a necessary condition for the existence of a stabilizing controller.

In this section, a controller is called stabilizable if the corresponding closed loop sensitivity transfer function $S(jw) \triangleq \frac{Y(jw)}{N(jw)}$, from n to y, is strictly stable (i.e., all of its poles have negative real parts); or alternatively $\lim_{t\to\infty} E|y(t)|^2 < \infty$ or $\lim_{t\to\infty} E||x(t)||^2 < \infty$.

The main result of this section is given in the following theorem. This theorem presents a new result.

Theorem 3.3.1 *Consider the control/communication system (3.3), (3.4) given in Figure 3.2. A necessary condition for the existence of a stabilizing controller is given by*

$$\mathcal{C} \geq \sum_{\{i:Re(\lambda_i(A))\geq 0\}} Re(\lambda_i(A)), \tag{3.7}$$

where $\mathcal{C}$ is the capacity of the AWGN channel.

Proof: The proof is original and it is given in Appendix.

Remark 3.3.2 *Condition (3.7) is important in designing the control/communication system described by (3.3), (3.4) and linear time-invariant encoder, decoder and controller. In particular, if the capacity is less than $\sum_{\{i:Re(\lambda_i(A))\geq 0\}} Re(\lambda_i(A))$, there is no stability scheme for stability of system (3.3). That is, $\sum_{\{i;Re(\lambda_i(A))\geq 0\}} Re(\lambda_i(A))$ is a fundamental limit on capacity for stability.*

3.4 Optimal Encoding/Decoding Scheme for Observability

In this section, we design an optimal encoder/decoder pair for the time-varying system defined by (3.1) and (3.2) (in this section and the next sections, we assume $N_0 = 1$) that guarantees the observability condition defined in the sense of Definition 3.2.2. The necessary and sufficient condition for the existence of such encoder/decoder pair is given in terms of the capacity of the channel and the time varying system matrix $A(t)$. First, we consider the scalar case of system (3.1) because it is easier to present the idea developed which is based on [51, 53], and then we extend the result to the vector case.

3.4.1 Optimal Encoding/Decoding Scheme for Observability: The Scalar Case

In this section, we recall the definition of the channel capacity of the flat fading continuous time AWGN channel when the channel state information θ is known to the transmitter and the receiver (since the main objective of this chapter is to find the continuous version of the eigenvalue rate condition for continuous time systems, for simplicity in analyzing, we consider the case with known channel state information which correspond to slow fading case). Further, we design an encoding scheme which minimizes the mean square decoding error, and achieves the channel capacity. In particular, the mean square error is bounded if $G(t)$ is non-zero, and tends to zero asymptotically if $G(t)$ is zero. This can be explained by the fact that if $G(t)$ is non-zero, the state equation is driven by Brownian motion which has unbounded variation. Finally, we state the necessary and sufficient conditions for bounded asymptotic and asymptotic observability in the mean square sense.

Capacity of Feedback Systems

The definition of channel capacity for a Gaussian flat fading channel, when CSI is fully known is given next.

Definition 3.4.1 *Consider the model given by (3.1) and (3.2) subject to the following instantaneous power constraint, when the fading process z or θ is completely known to the* *transmitter, and receiver*

$$E\Big[|f(t,x,\tilde{x},\theta)|^2\Big|\mathcal{F}^{\theta}_{0,t}\Big] \leq P. \tag{3.8}$$

Then, the finite-time and infinite time channel capacity are defined by

$$\mathcal{C}_f^T \triangleq \sup_{(x,f)\in\mathcal{X}\times\mathcal{F}_{ad}} \frac{1}{T} I_T(x;y|\theta), \quad \mathcal{C}_f \triangleq \liminf_{T\to\infty} \mathcal{C}_f^T, \tag{3.9}$$

where $I_T(x;y|\theta)$ is defined in Theorem 6.2.16. Here, the supremum is taken over all state processes $x \in \mathcal{X}$ which gives strong solutions to (3.1) and over all encoding functions $f \in \mathcal{F}_{ad}$ that satisfy (3.8) (see [51, 52, 54, 55]).

In Lemma 6.2.20, an upper bound on the mutual information is introduced and subsequently, it is shown that $\mathcal{C}_f = \liminf_{T\to\infty} \frac{P}{2T}\int_0^T E[z^2(t,\theta(t))]dt$.

Optimal Encoding and Decoding

In this section, we design an encoding/decoding scheme that achieves the channel capacity $\mathcal{C}_f^T$,$\mathcal{C}_f$. The proposed encoding/decoding scheme is optimal in the sense that among all admissible encoding/decoding schemes that satisfy condition (3.8), it minimizes the mean square decoding error and at the same time achieves the channel capacity. We then, employ the expression for the minimum mean square decoding error to obtain necessary and sufficient conditions for bounded asymptotic and asymptotic observability. In the subsequent development, only linear encoders are considered, because along the same lines of ([51], Section 16.4), it can be shown that linear encoders achieve the channel capacity and the minimum mean square decoding error.

Definition 3.4.2 *(Set of Linear Admissible Encoders). The set of linear admissible encoders $\mathcal{L}_{ad} \subset \mathcal{F}_{ad}$, is the set of linear non-anticipative functionals $f(t,x,\tilde{x},\theta)$ which have the following form*

$$f(t,x,\tilde{x},\theta) = f_0(t,\tilde{x},\theta) + f_1(t,\tilde{x},\theta)x(t), \tag{3.10}$$

Using linear encoders, the received signal y is given by

$$dy(t) = z(t,\theta(t))[f_0(t,\tilde{x},\theta) + f_1(t,\tilde{x},\theta)x(t)]dt + dv(t), \quad y(0) = 0. \tag{3.11}$$

Decoding. Since the desired observability criterion is the minimum mean square decoding error and because the decoded signal $\tilde{x}$ is a function of the received signal y and the channel state θ, the optimal decoder minimizing the mean square decoding error is the conditional expectation given by

$$\tilde{x}(t,y,u,\theta) = E[x(t)\Big|\mathcal{F}_{0,t}^{y,u,\theta}],\ \mathcal{F}_{0,t}^{y,u,\theta} = \mathcal{F}_{0,t}^{y}\bigvee\mathcal{F}_{0,t}^{u}\bigvee\mathcal{F}_{0,t}^{\theta}. \tag{3.12}$$

The conditional error variance for the decoder defined by (3.12) is

$$V(t,y,\theta) = E[(x(t)-\tilde{x}(t,y,u,\theta))^2\Big|\mathcal{F}_{0,t}^{y,\theta}]. \tag{3.13}$$

Moreover, following Definition 6.2.46 the decoder $\tilde{x}(t,y,u,\theta)$ and the corresponding conditional error variance $V(t,y,\theta)$ satisfy the following Generalized Kalman Filtering equations.

$$\begin{aligned} d\tilde{x}(t,y,u,\theta) &= A(t)\tilde{x}(t,y,u,\theta)dt + N(t)u(t)dt + z(t,\theta(t))V(t,y,\theta)f_1(t,\tilde{x},\theta) \\ &\quad .[dy(t) - z(t,\theta(t))(f_0(t,\tilde{x},\theta) + f_1(t,\tilde{x},\theta)\tilde{x}(t,y,u,\theta))dt], \end{aligned} \tag{3.14}$$

$$\dot{V}(t,y,\theta) = 2A(t)V(t,y,\theta) - z^2(t,\theta(t))f_1^2(t,\tilde{x},\theta)V^2(t,y,\theta) + G^2(t), \tag{3.15}$$

with initial conditions $\tilde{x}(0) = \bar{x}_0$, and $V(0) = \bar{V}_0$.

Encoding. From the point of view of the coding theorem, an encoder is efficient if it operates near the channel capacity, while ensuring a decoding error that tends to zero exponentially fast. In our case, the choice of an efficient linear encoder of the form (3.10), described by the pair (f_0, f_1) is directly related to the expression for the conditional error variance (3.15). By choosing (f_0, f_1) appropriately, the conditional error variance is minimized, and the channel capacity C_f is achieved.

The optimal encoder and decoder as well as the corresponding conditional error variance are given in Theorem 6.2.21. The methodology is similar to one found in ([51], section 16.4), except that the channel state information θ has to be taken into account here, by working with the conditional mutual information instead of the unconditional. Figure 3.3 illustrates the optimal encoder and decoder.

From (6.109) (see Appendix), it follows that by employing the proposed optimal encoding/decoding scheme of Theorem 6.2.21, the mean square estimation error $V^*(t,y,\theta)$ is independent of a control signal. Hence, under certain conditions, the decoding error can be made arbitrary small, regardless of the control signals. This suggests that the encoder

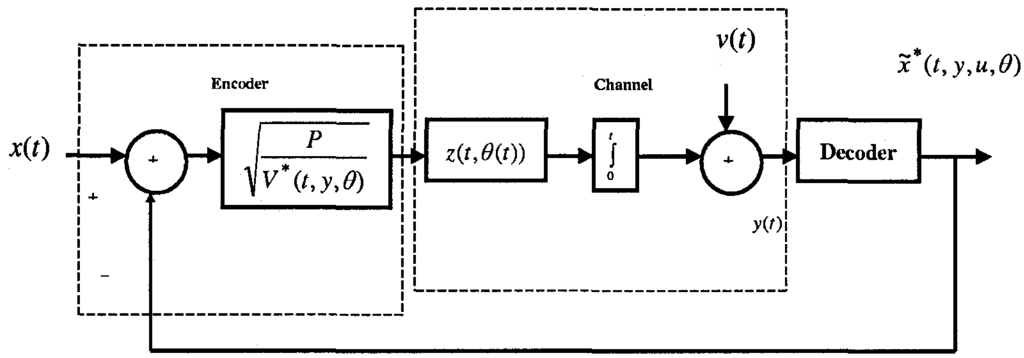

Figure 3.3: Optimal encoder and decoder

and decoder, can be designed independently of the controller. In other words, a separation principle holds between the control, and the communication part of the design.

Next, following the results of Theorem 6.2.21, we have the following necessary and sufficient conditions for mean square observability. These conditions are new.

Theorem 3.4.3 *i) When $G(t) \neq 0$, a sufficient condition for bounded asymptotic observability in the form of* $\limsup_{t\to\infty} \mathcal{E}(t) < \infty$ *is*

$$\inf_{t\in[0,\infty)} (Pz^2(t,\theta(t)) - 2[A(t)]^+) > 0, \quad a.s., \tag{3.16}$$

where $[a]^+ = a$ if $a \geq 0$ and $[a]^+ = 0$ otherwise.

ii) When $G(t) = 0$, (3.16) is a sufficient condition for asymptotic observability in the mean square sense (i.e., $\lim_{t\to\infty} \mathcal{E}(t) = 0$*).*

iii) For the time-invariant case of (3.1) when $G \neq 0$ and $z = 1$, a sufficient condition for bounded asymptotic observability (i.e., $\lim_{t\to\infty} \mathcal{E}(t) < \infty$*) is given by*

$$\mathcal{C} = \mathcal{C}_f = \frac{P}{2} > [A]^+. \tag{3.17}$$

Moreover, a necessary condition for such observability is given by

$$\mathcal{C} = \mathcal{C}_f \geq [A]^+. \tag{3.18}$$

Furthermore, (3.17) and (3.18) are also sufficient and necessary condition (if $\bar{V}_0 \neq 0$) for asymptotic observability (i.e., $\lim_{t\to\infty} \mathcal{E}(t) = 0$*), respectively, for the case of $G = 0$.*

Proof: The proof is original and it is given in Appendix.

Remark 3.4.4 *i) When* $G(t) = 0$, $\limsup_{t\to\infty} \frac{1}{t}\int_0^t A(s)ds$ *is bounded, and the channel is the continuous time AWGN channel (i.e.,* $z = 1$*), for which the channel capacity is* $\mathcal{C} = \frac{P}{2}$*, it is easily shown that another sufficient condition for asymptotic observability is*

$$\mathcal{C} = \frac{P}{2} > \limsup_{t\to\infty} \frac{1}{t}\int_0^t A(s)ds. \tag{3.19}$$

Moreover, a necessary condition for asymptotic observability is

$$\mathcal{C} \geq \liminf_{t\to\infty} \frac{1}{t}\int_0^t A(s)ds. \tag{3.20}$$

ii) Theorem 3.4.3 (or Remark 3.4.4, i) together with Theorem 6.2.21 provide conditions on the transmission power P *under which using the coding scheme of Theorem 6.2.21, the source (3.1) over the communication channel (3.2) is reliably reconstructed in the mean square sense. In particular, for time-invariant analogous plant over AWGN channels,* $P = 2[A]^+$ *is the minimum transmission power to be allocated for reliable mean square reconstruction.*

3.4.2 Optimal Encoding/Decoding Scheme for Observability: The Vector Case

In this section, we extend the previous results to the vector case. Here, we assume $A(t)$ is time-invariant (i.e., $A(t) = A$) and there exists a similarity transformation S such that $SAS^{-1} = \Lambda \triangleq diag(\lambda_1(A) \; ... \; \lambda_q(A))$ (e.g., A has linearly independent eigenvectors), where $diag(...)$ denotes diagonal matrix and $\lambda_i(A)$'s are eigenvalues of A. It is also assumed that $\lambda_i(A)$'s are real numbers (e.g., A is a symmetric matrix). We apply this similarity transformation $\gamma(t) \triangleq Sx(t)$ to transform system (3.1) into the following system

$$d\gamma(t) = \Lambda\gamma(t)dt + SN(t)u(t)dt + SG(t)dw(t), \quad \gamma(0) = S\bar{x}_0, \tag{3.21}$$

where $\Lambda = \begin{pmatrix} \Lambda_s & 0 \\ 0 & \Lambda_{us} \end{pmatrix}$, in which Λ_s block corresponds to the stable subspace and Λ_{us} block corresponds to unstable subspace. From previous part, we noticed that the stable eigenvalues do not contribute to capacity requirement for observability. Thus, without loss of generality we can restrict our attention to A matrices which contains only unstable eigenvalues (e.g., A is positive semi-definite). Similar idea, as above, for discrete time systems is used in [9]. In order to extend the previous result, the following extra assumptions are introduced regarding system (3.21).

Assumptions 3.4.5 $SG(t)G'(t)S'$ *and* $S\bar{V}_0S'$ *are diagonal matrices.*

Remark 3.4.6 *Notice that Assumptions 3.4.5 are satisfied if* $G(t)G'(t) = I_q$ *(e.g., G is orthogonal), A is a symmetric system matrix (note that for A symmetric,* $S' = S^{-1}$*), and the initial condition* $x(0)$ *has a corresponding covariance matrix of the form* $\bar{V}_0 = \alpha.I_q, \alpha \geq 0$.

Under Assumptions 3.4.5, it can be shown that the optimal mean square decoding error obtained by transmitting $\gamma(t)$ is diagonal since under these assumptions the uncontrolled version of the system (3.21) decomposed into q stochastically decoupled subsystems (please note that, as it was shown for the scalar case, the mean square decoding error is going to be independent of the control signal).

Please notice that $\gamma(t)'Q\gamma(t) = x'(t)S'QSx(t)$. Thus, stability of $\gamma(t)$ is equivalent to the stability of $x(t)$ and vice versa. Further, observability of $\gamma(t)$ is equivalent to observability of $x(t)$, particularly for A symmetric. Therefore, without loss of generality, we can consider the transformed system (3.21) instead of system (3.1) in our analysis.

By replacing $\gamma(t)$ with $x(t)$, the results obtained in Section 3.4.1 for the channel capacity do not change, and therefore the capacity of the flat fading AWGN channel is given by $C_f = \liminf_{T\to\infty} \frac{P}{2T}\int_0^T E_\theta[z^2(t,\theta(t))]dt$.

Next, by defining $f_1(t,\tilde{\gamma},\theta) \triangleq [f_{11}(t,\tilde{\gamma},\theta) \;...\; f_{qq}(t,\tilde{\gamma},\theta)]$, the set of admissible linear encoders will be similar to (3.10). Subsequently, the received signal is

$$dy(t) = z(t,\theta(t))[f_0(t,\tilde{\gamma},\theta) + f_1(t,\tilde{\gamma},\theta)\gamma(t)]dt + dv(t), \;\; y(0) = 0, \tag{3.22}$$

and the optimal mean square decoder is

$$\tilde{\gamma}(t,y,u,\theta) = E[\gamma(t)\big|\mathcal{F}_{0,t}^{y,u,\theta}], \tag{3.23}$$

with the following error covariance

$$V(t,y,\theta) = E[(\gamma(t) - \tilde{\gamma}(t,y,u,\theta))(\gamma(t) - \tilde{\gamma}(t,y,u,\theta))'\big|\mathcal{F}_{0,t}^{y,\theta}]. \tag{3.24}$$

Moreover, following Definition 6.2.46, the decoder $\tilde{\gamma}(t,y,u,\theta)$ and the corresponding error covariance $V(t,y,\theta)$ satisfy the following Generalized Kalman Filter equation.

$$\begin{aligned} d\tilde{\gamma}(t,y,u,\theta) &= \Lambda\tilde{\gamma}(t,y,u,\theta)dt + SN(t)u(t)dt + V(t,y,\theta)f_1'(t,\tilde{\gamma},\theta)z(t,\theta(t)) \\ &\quad .[dy(t) - z(t,\theta(t))(f_0(t,\tilde{\gamma},\theta) + f_1(t,\tilde{\gamma},\theta)\tilde{\gamma}(t,y,u,\theta))dt], \end{aligned} \tag{3.25}$$

$$\begin{aligned}\dot{V}(t,y,\theta) &= 2\Lambda V(t,y,\theta) - z^2(t,\theta(t))V(t,y,\theta)f_1'(t,\tilde{\gamma},\theta)f_1(t,\tilde{\gamma},\theta)V(t,y,\theta) \\ &+SG(t)G'(t)S',\end{aligned} \tag{3.26}$$

where $\tilde{\gamma}(0) = S\bar{x}_0$, and $V(0) = S\bar{V}_0S'$.
Applying Assumptions 3.4.5, $V(t,y,\theta)$ is diagonal (e.g., $V(t,y,\theta) = diag(V_{11}(t,y,\theta) \ldots V_{qq}(t,y,\theta))$, and the ith diagonal element of (3.26), is given by

$$\dot{V}_{ii}(t,y,\theta) = 2\lambda_i(A)V_{ii}(t,y,\theta) - z^2(t,\theta(t))V_{ii}^2(t,y,\theta)f_{ii}^2(t,\tilde{\gamma},\theta) + [SG(t)G'(t)S']_{ii}, \tag{3.27}$$

where $[SG(t)G'(t)S']_{ii}$ is the i-th diagonal element of $SG(t)G(t)'S'$. Consequently, following the same methodology used to prove Theorem 6.2.21, we have the following theorem.

Theorem 3.4.7 *Suppose Assumptions 3.4.5 hold, the received signal is defined by (3.22) and the source by (3.21). Then the encoder, which achieves the channel capacity* $C_f^T = \frac{P}{2T}\int_0^T E[z^2(t,\theta(t))]dt$, *the optimal decoder and the corresponding error covariance are respectively given by*

$$\begin{aligned}
&f^*(t,\gamma,\tilde{\gamma}^*,\theta) = f_0^*(t,\tilde{\gamma}^*,\theta) + \sum_{i=1}^{q} f_{ii}^*(t,\tilde{\gamma}^*,\theta)\gamma_i(t),\\
&f_0^*(t,\tilde{\gamma}^*,\theta) = -\sum_{i=0}^{q} f_{ii}^*(t,\tilde{\gamma}^*,\theta)\tilde{\gamma}_i^*(t,y,u,\theta),\\
&f_{ii}^*(t,\tilde{\gamma}^*,\theta) = \sqrt{\frac{\alpha_i P}{V_{ii}^*(t,y,\theta)}},\\
&d\tilde{\gamma}^*(t,y,u,\theta) = \Lambda\tilde{\gamma}^*(t,y,u,\theta)dt + SN(t)u(t)dt\\
&\qquad\qquad +z(t,\theta)[\sqrt{\alpha_1 PV_{11}^*(t,y,\theta)} \ \ldots \ \sqrt{\alpha_q PV_{qq}^*(t,y,\theta)}]'dy(t),
\end{aligned} \tag{3.28}$$

$$\begin{aligned}V_{ii}^*(t,y,\theta) &= [S\bar{V}_0S']_{ii}e^{(2\int_0^t \lambda_i(A)ds - \int_0^t \alpha_i z^2(s,\theta(s))Pds)} \\ &+ \int_0^t [S(s)G(s)G'(s)S'(s)]_{ii}e^{(2\int_s^t \lambda_i(A)du - \int_s^t \alpha_i z^2(u,\theta(u))Pdu)}ds,\end{aligned} \tag{3.29}$$

where $\tilde{\gamma}_i^*(t,y,u,\theta)$ *is the i-th element of* $\tilde{\gamma}^*(t,y,u,\theta)$ *and* $0 \le \alpha_i \le 1$, $\sum_{i=1}^q \alpha_i = 1$.

Next, we have the following theorem that extends Theorem 3.4.3 to the vector case.

Theorem 3.4.8 *Suppose Assumptions 3.4.5 hold.*
i) If there exists a set of $\{\alpha_i\}_{i=1}^q$ *such that* $0 \le \alpha_i \le 1$, $\sum_{i=1}^q \alpha_i = 1$ *in which*

$$\inf_{t\in[0,\infty)} (\alpha_i P z^2(t,\theta(t)) - 2[Re(\lambda_i(A))]^+) > 0, \quad \forall i = 1,2,...,q, \tag{3.30}$$

where $Re(.)$ *is an operator which returns the real parts of complex numbers and* $\lambda_i(A)$*'s are the eigenvalues, then we have bounded asymptotic observability in the form of* $\limsup_{t\to\infty} E[(\gamma(t)-\tilde{\gamma}(t))'(\gamma(t)-\tilde{\gamma}(t))|\mathcal{F}_{0,t}^{y,\theta}] < \infty$ *for the case of* $G(t) \neq 0$*; and asymptotic observability in the mean square sense (i.e.,* $\lim_{t\to\infty} E[(\gamma(t)-\tilde{\gamma}(t))'(\gamma(t)-\tilde{\gamma}(t))|\mathcal{F}_{0,t}^{y,\theta}] = 0$*) for the case of* $G(t)=0$.
ii) For the time-invariant case of (3.21) when $G \neq 0$ *and* $z = 1$*, a sufficient condition for bounded asymptotic observability (i.e.,* $\lim_{t\to\infty} E[(\gamma(t)-\tilde{\gamma}(t))'(\gamma(t)-\tilde{\gamma}(t))|\mathcal{F}_{0,t}^{y,\theta}] < \infty$*) is given by*

$$\mathcal{C} = \mathcal{C}_f = \frac{P}{2} > \sum_{\{i;Re(\lambda_i(A))\ge 0\}} Re(\lambda_i(A)). \tag{3.31}$$

Moreover, a necessary condition for such observability is given by

$$\mathcal{C} = \mathcal{C}_f = \frac{P}{2} \ge \sum_{\{i;Re(\lambda_i(A))\ge 0\}} Re(\lambda_i(A)). \tag{3.32}$$

Furthermore, the above conditions are also sufficient and necessary conditions for asymptotic observability (i.e., $\lim_{t\to\infty} E[(\gamma(t)-\tilde{\gamma}(t))'(\gamma(t)-\tilde{\gamma}(t))|\mathcal{F}_{0,t}^{y,\theta}] = 0$*), respectively, for* $G = 0$.

Proof: The proof is original and it is given in Appendix.

3.5 Optimal Controller

In this section, we propose a state feedback controller that minimizes a quadratic pay-off for stability of the time-invariant version of system (3.1) (e.g., $A(t) = A$, $N(t) = N$ and $G(t) = G$). The stability problem considered in this section has not been considered elsewhere; and thus this section presents a new stability scheme for stability of the control/communication systems. Here, similar to the previous section, we assume there exists a similarity transformation S such that $SAS^{-1} = \Lambda$ in which we can consider the time-invariant version of the transformed system (3.21) instead of system (3.1) in our analysis. Since it is assumed that the channel state information is known, the average pay-off is defined for a fixed channel sample path. That is, for a fixed channel sample path θ, the state feedback controller is chosen to minimize the following quadratic pay-off

$$J^T = \frac{1}{T} E\{\int_0^T [\gamma'(t)Q\gamma(t) + u'(t)Ru(t)]dt\}, \tag{3.33}$$

where $Q > 0$ and $R > 0$ are symmetric weighting matrices. Moreover, we also consider the infinite-horizon $\bar{J} = \limsup_{T\to\infty} J^T$. For the infinite-horizon we assume the controllability rank condition (Definition 6.2.37) $Rank(Co) = q$, $Co \triangleq [N\ AN\ \ldots\ A^{q-1}N]$. The time-invariant version of system (3.21) over the communication channel (3.22) subject to the linear encoder, can be viewed as the following partially observed controlled system at the communication end.

$$\begin{cases} d\gamma(t) = \Lambda\gamma(t)dt + SNu(t)dt + SGdw(t) & \gamma(0) = S\bar{x}_0 \\ dy(t) - z(t,\theta(t))f_0(t,\tilde{\gamma},\theta)dt = z(t,\theta(t))f_1(t,\tilde{\gamma},\theta)\gamma(t)dt + dv(t), & y(0) = 0, \end{cases} \tag{3.34}$$

Next, according to the classical separation theorem of estimation and control (Definition 6.2.46), the optimal controller that minimizes (3.33) subject to a flat fading AWGN communication channel and the linear encoder $f(t,\gamma,\tilde{\gamma},\theta) = f_0(t,\tilde{\gamma},\theta) + f_1(t,\tilde{\gamma},\theta)\gamma(t)$, is separated into a state estimator and a certainty equivalent controller given by

$$u^*(t) = -K(t)\tilde{\gamma}(t,y,u,\theta), \quad K(t) = R^{-1}N'P(t), \tag{3.35}$$

where $\tilde{\gamma}(t,y,u,\theta)$ is the solution of (3.25) with the corresponding observer Riccati equation (3.26) and $P(t)$ is the solution of the following regulator Riccati equation

$$-\dot{P}(t) = Q - P(t)SNR^{-1}N'S'P(t) + 2\Lambda P(t), \quad P(T) = 0. \tag{3.36}$$

For a fixed sample path of the channel θ, it follows that if the observer and regulator Riccati equations have steady state solution $\bar{V} \triangleq \lim_{t\to\infty} V(t,y,\theta)$ ($V(t,y,\theta)$ is the solution of the observer Riccati equation (3.26))and $\bar{P} \triangleq \lim_{t\to\infty} P(t)$, respectively, the averaged criterion

$$\bar{J} = \limsup_{T\to\infty} \frac{1}{T}E\{\int_0^T [\gamma'(t)Q\gamma(t) + u^{*'}(t)Ru^*(t)]dt\} \tag{3.37}$$

can be expressed in the alternative form

$$\bar{J} = \lim_{T\to\infty} \frac{1}{T}E\{\int_0^T [\gamma'(t)Q\gamma(t) + u^{*'}(t)Ru^*(t)]dt\} = trac[\bar{P}SGG'S' + \bar{V}\bar{K}'R\bar{K}], \tag{3.38}$$

where $\bar{K} = R^{-1}N'\bar{P}$. From Lemma 6.2.47 follows that the regulator Riccati equation (3.36) has the steady state solution $\bar{P}$ as $t \to \infty$, if the rank condition holds. Moreover, under Assumptions 3.4.5, from Theorem 3.4.8 follows that the observer Riccati equation (3.26) has the steady-state solution, $\bar{V} = \lim_{t\to\infty} V(t,y,\theta)$ a.s., if the optimal encoding/decoding scheme of Theorem 3.4.7 is used, condition (3.30) holds and $G = 0$; or $G \neq 0$, but $z = 1$.

Under these assumptions when $G = 0$, then $\bar{V} = 0$, a.s.; and when $G \neq 0$ but $z = 1$, then $\bar{V} = diag(\frac{[SGG'S']_{11}}{\alpha_1 P - 2\lambda_1(A)} \cdots \frac{[SGG'S']_{qq}}{\alpha_q P - 2\lambda_q(A)})$. Next, for a time-invariant version of system (3.21), we have the following proposition for stability defined in the sense of Definition 3.2.3. This proposition contains new results.

Proposition 3.5.1 *Consider the time-invariant version of system (3.21) and assume the rank condition holds.*
Then, for a fixed sample path of the channel, we have the followings (a.s.).
i) Assuming $G \neq 0$ *and* $V(t, y, \theta) \rightarrow \bar{V}$ *as* $t \rightarrow \infty$, *by using the optimal policy (3.35),* $\lim_{T\rightarrow\infty} \frac{1}{T} \int_0^T E||\gamma(t)||_Q^2 dt < \infty$.
ii) Assuming $G = 0$ *and* $V(t, y, \theta) \rightarrow 0$ *as* $t \rightarrow \infty$, *by using the optimal policy (3.35),* $\lim_{T\rightarrow\infty} \frac{1}{T} \int_0^T E||\gamma(t)||_Q^2 dt = 0$.

Proof: The proof is original and it is given in Appendix.

Next, from Theorem 3.4.8 and Proposition 3.5.1, we have the following theorem for bounded stability and stability in the mean square sense. This theorem contains new results.

Theorem 3.5.2 *Consider the time-invariant version of the system (3.21). Assume the rank condition and Assumptions 3.4.5 hold and the sample path of the channel is completely known.*
Then
i) For the case of $G \neq 0$ *and* $z = 1$, *a sufficient condition for bounded stability in the mean square sense is given by*

$$\mathcal{C} = \frac{P}{2} > \sum_{\{i; Re(\lambda_i(A)) \geq 0\}} Re(\lambda_i(A)), \tag{3.39}$$

where $\mathcal{C}$ *is the capacity,* $Re(.)$ *is an operator which returns the real parts of complex numbers and* $\lambda_i(A)$*'s are the eigenvalues of the system matrix* A. *ii) For the case of* $G = 0$ *and* $z = 1$, *(3.39) is also a sufficient condition for stability in the mean square sense.*
iii) For the case of $G = 0$ *and in the presence of fading (i.e.,* $z \neq 1$*), (3.30) is a sufficient condition for stability in the mean square sense.*

Proof: i) Suppose the condition (3.39) is satisfied. Then, by using the optimal encoding/decoding scheme of Theorem 3.4.7, $V(t, y, \theta) \rightarrow \bar{V}$. Consequently, from Proposition

3.5.1, it follows that by using the optimal control policy (3.35), and the optimal encoding/decoding scheme of Theorem 3.4.7, the stability condition is guaranteed.
ii) It is shown along the same lines of i.
iii) It is shown along the same lines of i.

Remark 3.5.3 *i) First, we note that the above conditions are new.*
ii) Theorem 3.5.2 provides condition on the transmission power under which using the encoding scheme of Theorem 3.4.7 and certainly equivalent controller (3.35), the plant is mean square stabilizable, while the reliable mean square reconstruction is also guaranteed. In particular for time-invariant analogous plant over AWGN channels, $P = 2\sum_{\{i;Re(\lambda_i(A))\geq 0\}} Re(\lambda_i(A))$ *is the minimum transmission power to be allocated to have reliable mean square reconstruction, in which the mean square stability is also achieved using the encoding scheme of Theorem 3.4.7 and certainly equivalent controller (3.35).*
iii) The case of $P = \infty$ *corresponds to the case of unlimited capacity. In this case from (3.29) follows that* $\gamma(t) = \tilde{\gamma}(t)$ *(i.e.,* $x(t) = \tilde{x}(t)$*) and subsequently, encoder, channel and decoder in Figure 3.1 can be replaced by a straight line; and thus the proposed stability scheme is reduced to the standard results (Definition 6.2.46). That is, our proposed stability and encoding schemes extend the standard results to the cases where the observed process is transmitted over an AWGN channel subject to limited transmission power and subsequently limited capacity.*

3.6 Conclusion

In this chapter, we first found a necessary condition for bounded stability in the mean square sense for a fully observed linear continuous time-invariant noiseless plant over continuous time AWGN communication channels when the subsystems are linear and time-invariant. Then, for the fully observed linear continuous time-varying systems driven by Brownian motion over flat fading continuous time AWGN channels, the optimal encoding scheme was constructed for bounded asymptotic and asymptotic observability in the mean square sense. From this construction, necessary and sufficient conditions for bounded asymptotic and asymptotic observability in the mean square sense were derived. Following this construction, we also proposed a mean square stabilizing control scheme which stabilizes the fully observed linear continuous time-invariant systems subject to Brownian motion. From the obtained conditions, it is concluded that a lower bound on the capacity in terms of the summation of the real parts of the unstable eigenvalues of the open loop time-invariant system, is the continuous version of the eigenvalue rate condition.

Chapter 4

Uniform Observability and Robust Control

4.1 Introduction

This chapter is concerned with uniform reliable data reconstruction (i.e., uniform observability) and robust stability of discrete time uncertain dynamical systems which are controlled over discrete time channels subject to limited capacity constraint (see Figure 4.1). The uncertainty in the dynamical system is described by a relative entropy constraint. As it will be shown, such uncertainty description is a natural generalization of the sum quadratic uncertainty description [57], [58]. The methodology used to address uniform reliable data reconstruction and robust stability is information theoretic and it invokes a robust version of the Shannon lower bound.

The problem of uniform observability and/or robust stability of fully observed uncertain dynamical systems subject to bounded uncertainty have been considered in [4], [9], [10] and [16]. In particular, the cited references are concerned with the following dynamical system $\begin{cases} X_{t+1} = AX_t + NU_t + Z_t, \\ Y_t = X_t, \end{cases}$ where $X_t, Z_t, Y_t \in \Re^q$, $U_t \in \Re^o$ and Z_t is the disturbance such that $||Z_t|| \leq d$ almost surely. This chapter complements the already existing results in the literature by considering the uncertainty in the dynamical system described by a relative entropy constraint in which this description is a natural generalization of the sum quadratic uncertainty description. Sum quadratic uncertainty description is more suitable description for modeling uncertain dynamical systems than bounded uncertainty description considered in [4], [9], [10], [16]. We study the observability and stability problem of the following

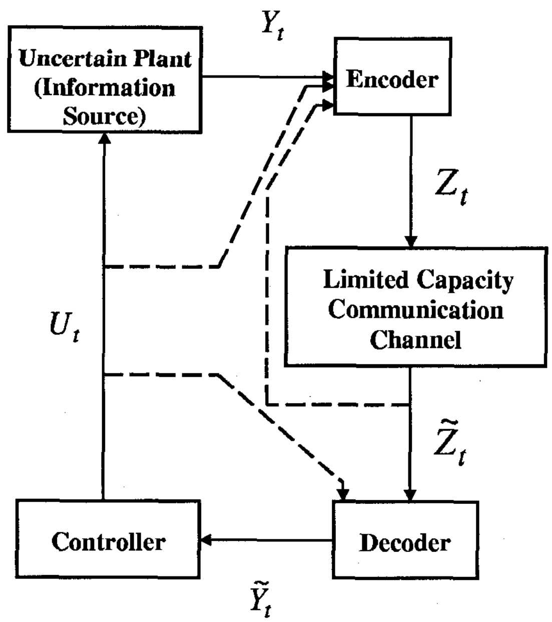

Figure 4.1: Control/communication system described by an uncertain plant

uncertain dynamical system

$$\begin{cases} X_{t+1} = AX_t + NU_t + BW_t + B\bar{W}_t, \\ Y_t = X_t, \end{cases} \tag{4.1}$$

where $X_t, Y_t \in \Re^q$, $U_t \in \Re^o$, $W_t, \bar{W}_t \in \Re^m$, $A \in \Re^{q\times q}$, $N \in \Re^{q\times o}$, $B \in \Re^{q\times m}$, W_t orthogonal $\sim N(0, \Sigma_W)$, $\Sigma_W > 0$, and $\bar{W}_t$ is the perturbed noise process such that it satisfies the following sum quadratic constraint $\frac{1}{2T}\sum_{t=0}^{T-2} E[\bar{W}_t'\Sigma_W^{-1}\bar{W}_t] \leq R_c + \frac{1}{2T}\sum_{t=0}^{T-1} E[Y_t' M Y_t]$, where R_c is a non-negative scalar, $M = M' \in \Re^{q\times q} \geq 0$ and T can be arbitrary large.

The problem of reliable data reconstruction and stability of uncertain system (4.1) over limited capacity communication channels has not been considered elsewhere; and thus this chapter contains original results.

In this chapter first the robust entropy defined as a maximization of the Shannon entropy over the class of sources is considered. When the class of sources is described by a relative entropy constraint, the explicit solution to the maximization problem is found and the connection between this solution and Rényi and Tsallis entropies is established. Then, by developing and invoking a robust version of the information transmission theorem and the Shannon lower bound, necessary conditions for uniform observability and robust stability of the control/communication system of Figure 4.1 are derived. These conditions are given in the form of a lower bound on the capacity in terms of the robust entropy rate.

Throughout by calculating the robust entropy rate, the obtained necessary conditions are applied to specific uncertain plants described by a relative entropy constraint. Furthermore for the uncertain system (4.1), a relation between the robust entropy rate with the solution of the Algebraic Riccati equation appearing in the H^∞ estimation and control problem, is established. In addition, a description of a class of sources using power spectral density functions is briefly discussed and the robust entropy rate is calculated for a class of sources with the corresponding Gaussian nominal distribution. Subsequently, necessary conditions for uniform observability and robust stability of such system are also given.

In addition, the robust stability of the fully observed control uncertain Gauss Markov system (4.1) over AWGN channels, is addressed; and subsequently, it is concluded that the necessary condition for uniform observability is also a sufficient condition.

This chapter is organized as follows. In Section 4.2, the problem formulation is given. Then, in Section 4.3, the definition of robust entropy is recalled; and the explicit solution to the robust entropy when the class of sources is described by a relative entropy uncertainty constraint, is found. Then, specific examples are worked out. In Section 4.4, the robust information transmission theorem and the robust Shannon lower bound are developed. Subsequently, the necessary conditions for uniform observability and robust stability of the control/communication system of Figure 4.1 are presented. Furthermore, stability problem of system (4.1) over AWGN channels is addressed. Long proofs are given in Appendix 6.1. Similar to previous chapters, in Appendix 6.2 we summarized the known concepts, measures, and results from different fields that we are going to use them throughout this chapter, to derive new results.

4.2 Problem Formulation

In this chapter, we are concerned with the control/communication system of Figure 4.1. The description of signals and spaces associated with Figure 4.1 are similar to Figure 2.1 (see Section 2.2). Throughout, $\log(.)$ and $\log_e(.)$ denote logarithm of base 2 and natural logarithm, respectively, and $\mathcal{M}_1(\Re^{Td})$ is the set of all finitely additive probability measures on Borel measurable space $(\Re^{Td}, \mathcal{B}(\Re^{Td}))$, where the sequence of R.V.'s $Y^{T-1} \triangleq (Y_0, Y_1, ..., Y_{T-1})$ is defined on this space.

The different blocks of Figure 4.1 are described below.

Information Source. The information source is the output of the controlled dynamical system with input U_t and output Y_t. Throughout this chapter it is assumed that the controlled dynamical system is subject to certain unknown terms (known as perturbed terms) which are unknown but they belong to certain known classes. Different values for the perturbed terms correspond to different controlled dynamical systems. Throughout this chapter we are particularly interested to those uncertain controlled dynamical systems which are described by the following relative entropy constraint.

$$f_{Y^{T-1}} \in \mathcal{D}_{SU}(g_{Y^{T-1}}) \triangleq \Big\{ f_{Y^{T-1}}; \frac{1}{T} H(f_{Y^{T-1}} || g_{Y^{T-1}}) \leq R_c + E_{f_{Y^{T-1}}}[\frac{1}{2T}\sum_{t=0}^{T-1} Y_t' M Y_t] \Big\} \quad (4.2)$$

where $f_{Y^{T-1}}$ is the joint density function of the observations process Y^{T-1} associated with the controlled dynamical system subject to the perturbed terms, $g_{Y^{T-1}}$ is the joint density function of the observations process of the controlled dynamical system in the absence of perturbed terms (i.e., nominal density), $H(.||.)$ is the relative entropy (Definition 6.2.27), R_c is a non-negative scalar and $M = M' \in \Re^{d\times d}$ is a positive semi-definite matrix, and $E_{f_{Y^{T-1}}}[\cdot]$ is the expectation with respect to the joint density function $f_{Y^{T-1}}$.

Thus, the relative entropy $H(f_{Y^{T-1}}||g_{Y^{T-1}})$ can be thought of as a measure of the difference between the controlled dynamical system subject to the perturbed terms; and the controlled dynamical system in the absence of the perturbed terms (i.e., nominal system). Therefore, this relative entropy constraint sets a restriction on the class where the perturbed terms belong to. Particularly, this relative entropy constraint does not set any constraints on the control sequence and subsequently encoder, decoder, and controller. Please note that the most restrictive case on the relative entropy constraint (4.2) occurs when $R_c = 0$ and $M = 0$. But, this case implies $f_{Y^{T-1}} = g_{Y^{T-1}}$ which corresponds to the case where the perturbed terms belong to the empty set. That is, more restriction on the relative entropy constraint puts more restriction only on the class where the perturbed terms belong to.

One example of controlled dynamical systems subject to the perturbed terms (i.e., uncertain dynamical systems) that can be described by the relative entropy constraint (4.2) is the following family of fully observed Gauss Markov systems.

$$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t\geq 0}) : \begin{cases} X_{t+1} = AX_t + NU_t + BW_t + B\bar{W}_t, & X_0 = X, \\ Y_t = H_t, & H_t = X_t \end{cases} \quad (4.3)$$

where $X_t \in \Re^q$, $U_t \in \Re^o$, $W_t \in \Re^m$, $\bar{W}_t \in \Re^m$, $X_0 \sim N(\bar{x}_0, \bar{V}_0)$, $Y_t, H_t \in \Re^q$, W_t is i.i.d. $\sim N(0, \Sigma_W)$, $\Sigma_W > 0$, $\bar{W}_t$ is the perturbed noise random process which is $\{\sigma\{W_l\}; l =$

$1, 2, ..., t-1\}$ adapted (Definition 6.2.4) ($\sigma\{.\}$ denotes the σ-algebra) and H_t is the signal to be controlled.

The nominal system associated with the above uncertain system, is the following fully observed system.

$$(\Omega, \mathcal{F}(\Omega), \Pi; \{\mathcal{F}_t\}_{t\geq 0}) : \left\{ \begin{array}{ll} X_{t+1} = AX_t + NU_t + BW_t, & X_0 = X, \\ Y_t = H_t, & H_t = X_t \end{array} \right. \tag{4.4}$$

In [59], it is shown that for the uncertain system (4.3) with the nominal system (4.4), we have $H(f_{Y^{T-1}}||g_{Y^{T-1}}) = \frac{1}{2}E_P\left[\sum_{t=0}^{T-2} \bar{W}_t'\Sigma_{\bar{W}}^{-1}\bar{W}_t\right]$, where $P(dY^{T-1}) = f_{Y^{T-1}}dY^{T-1}$ and $\Pi(dY^{T-1}) = g_{Y^{T-1}}dY^{T-1}$. That is, the relative entropy constraint (4.2) holds for the uncertain system (4.3) with the nominal system (4.4), provided the following sum quadratic constraint holds.

$$\left\{\{\bar{W}_t\}_{t=0}^{T-2}; E_P[\frac{1}{2T}\sum_{t=0}^{T-2}(\bar{W}_t'\Sigma_{\bar{W}}^{-1}\bar{W}_t)] \leq R_c + E_P[\frac{1}{2T}\sum_{t=0}^{T-1}(Y_t'MY_t)]\right\}, \tag{4.5}$$

where R_c is a non-negative scalar and $M = M' \in \Re^{q\times q} \geq 0$.

Remark 4.2.1 *As it is clear from above examples, our motivations for introducing the relative entropy uncertainty description (4.2) is that it provides a natural stochastic generalization of the sum quadratic constraint uncertainty description. Such uncertainty description has been considered in [57], [58], [59]; and for continuous time systems (in the form of integral quadratic constraint uncertainty description) has been considered in [60]-[62]. Sum quadratic uncertainty description is an appropriate description for modeling uncertain dynamical systems.*

Communication Channel: The communication channel at time $t \in \mathbf{N}_+$ is modeled by a feedback channel with memory via a sequence of stochastic kernels $\{P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1}); t \in \mathbf{N}_+\}$, $t \in \mathbf{N}_+ \triangleq \{0, 1, 2, ...\}$, where $Z^t = z^t$ is the specific realization of the channel input, and $\tilde{Z}^{t-1} = \tilde{z}^{t-1}$, is the specific realization of the previous channel outputs.

Encoder: We define and discuss the following types of encoders.

Class A) The encoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(dZ_t; y^t, u^{t-1}, \tilde{z}^{t-1})$.

Class B) The encoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(dZ_t; y^t, u^{t-1})$.

Class C) The encoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(dZ_t; y^t, \tilde{z}^{t-1})$.

Decoder: We define and discuss the following types of decoders.

Class A) The decoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(d\tilde{Y}_t; \tilde{z}^t)$.

Class B) The decoder at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(d\tilde{Y}_t; \tilde{z}^t, u^{t-1})$.

Controller: The controller at any time $t \in \mathbf{N}_+$ is modeled by a stochastic kernel $P(dU_t; \tilde{z}^t, u^{t-1})$.

Please note that the above descriptions for channel, encoder, decoder and controller are similar to the descriptions presented in [9] and [11].

In this chapter, we are concerned with the following observability and stability criterion.

Definition 4.2.2 *(Uniform Observability in Probability and r- Mean). Consider the general control/communication system of Figure 4.1 described by an uncertain plant.*

i) For a given $\delta \geq 0$ and $D_v \in [0,1)$, the uncertain plant is called (δ, D_v)- uniform observable in probability if there exist an encoder and decoder such that

$$\lim_{T\to\infty} \sup_{P(dY^{T-1})\in\mathcal{M}_{SU}^T} \frac{1}{T}\sum_{t=0}^{T-1} \Pr(||Y_t - \tilde{Y}_t|| > \delta) \leq D_v, \tag{4.6}$$

where $\mathcal{M}_{SU}^T$ is the set of all admissible probability measure associated with the observations process Y^{T-1} of the uncertain plant.

ii) For a given $r > 0$ and a finite $D_v \geq 0$, the uncertain plant is called (r, D_v)-uniform observable if there exist an encoder and a decoder such that $\lim_{T\to\infty} \sup_{P(dY^{T-1})\in\mathcal{M}_{SU}^T} \frac{1}{T}\sum_{t=0}^{T-1} E_P$ $||Y_t - \tilde{Y}_t||^r \leq D_v$.

Definition 4.2.3 *(Robust Stability in Probability and r- Mean). Consider the general control/communication system of Figure 4.1 described by an uncertain plant, in which $Y_t = H_t + \Gamma_t$ where H_t is the signal to be controlled and Γ_t represents the effects of measurement noise and the perturbed terms.*

i) For a given $\delta \geq 0$ and $D_v \in [0,1)$, the uncertain plant is called (δ, D_v)- robust stabilizable in probability if there exist a controller, an encoder, and decoder such that

$$\lim_{T\to\infty} \sup_{P(dY^{T-1})\in\mathcal{M}_{SU}^T} \frac{1}{T}\sum_{t=0}^{T-1} \Pr(||H_t|| > \delta) \leq D_v. \tag{4.7}$$

ii) For a given $r > 0$ and finite $D_v \geq 0$, the uncertain plant is called (r, D_v)- robust stabilizable if there exist a controller, encoder and a decoder such that $\lim_{T\to\infty} \sup_{P(dY^{T-1})\in\mathcal{M}_{SU}^T} \frac{1}{T}\sum_{t=0}^{T-1}$ $E_P||H_t||^r \leq D_v$.

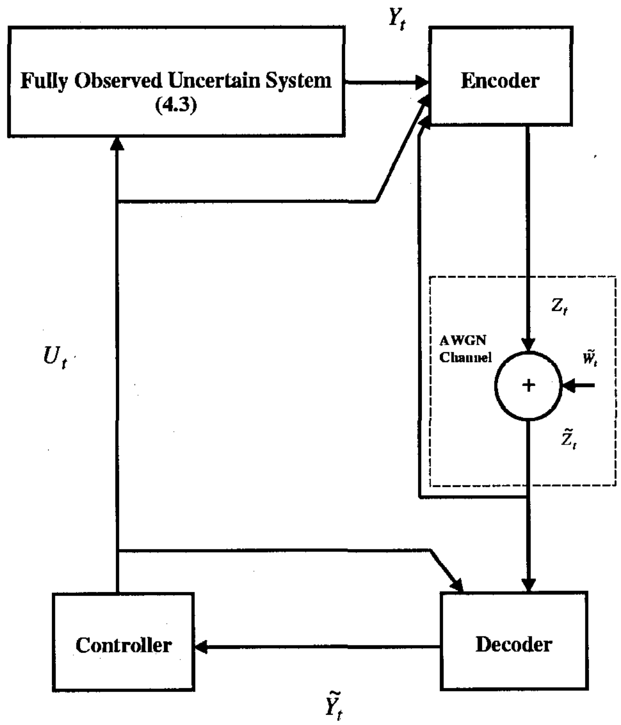

Figure 4.2: Control/communication system described by the fully observed uncertainty plant

The above definitions for observability and stability which uses sequences, are variants of the almost sure observability and stability criterion introduced in [9].

Here, it is assumed that the uncertain controlled dynamical system, communication channel, the types of encoder, decoder and controller are given. The objective is to design encoder, decoder and controller for uniform observability and robust stability in probability and r-mean when the capacity is as minimum as possible.

Throughout this chapter we are particularly interested to the control/communication system of Figure 4.2 (which is a special case of the control/communication system of Figure 4.1) described by the uncertain controlled dynamical system (4.3) subject to the sum quadratic uncertainty description (4.5), over the following AWGN channel when the encoder is of type A and the decoder is of type B.

$$\tilde{Z}_t = Z_t + \tilde{W}_t, \ \ \tilde{W}_t \text{ orthogonal} \sim N(0, W_c), \quad E_P[Z_t' Z_t] \leq P_t, \tag{4.8}$$

where Z_t is the channel input, $\tilde{Z}_t$ is the channel output, $\tilde{W}_t$ is the channel noise and P_t is the power constraint.

The observations process of system (4.3) is written as follows.

$$Y_t = A^t X_0 + \sum_{i=0}^{t-1} A^{t-1-i} B W_i + \sum_{i=0}^{t-1} A^{t-1-i} B \bar{W}_i + \sum_{i=0}^{t-1} A^{t-1-i} N U_i, \quad 0 \le t \le T-1, \tag{4.9}$$

where $\bar{W}_t$ belongs to the following class

$$\Big\{ \{\bar{W}_t\}_{t=0}^{T-2}; E_P[\frac{1}{2T} \sum_{t=0}^{T-2} (\bar{W}_t' \Sigma_W^{-1} \bar{W}_t)] \le R_c + E_P[\frac{1}{2T} \sum_{t=0}^{T-1} (Y_t' M Y_t)] \Big\}. \tag{4.10}$$

But, by the knowledge of the control sequence U^{t-1} at the encoder and decoder, the last term in (4.9) (i.e., $\sum_{i=0}^{t-1} A^{t-1-i} N U_i$) is reconstructed perfectly at the encoder and at the communication end. Thus, the problem of reliable data reconstruction of Y_t is equivalent to the problem of reliable data reconstruction of $\bar{Y}_t$, given as follow.

$$\bar{Y}_t = A^t X_0 + \sum_{i=0}^{t-1} A^{t-1-i} B W_i + \sum_{i=0}^{t-1} A^{t-1-i} B \bar{W}_i. \tag{4.11}$$

But, $\bar{Y}_t$ is the outputs of the uncontrolled analogous system, that is, the following system.

$$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t \ge 0}) : \begin{cases} X_{t+1} = A X_t + B W_t + B \bar{W}_t, & X_0 = X, \\ \bar{Y}_t = H_t, & H_t = X_t \end{cases} \tag{4.12}$$

where for this system, the perturbed term $\bar{W}_t$ belongs to the following class.

$$\Big\{ \{\bar{W}_t\}_{t=0}^{T-2}; E_P[\frac{1}{2T} \sum_{t=0}^{T-2} (\bar{W}_t' \Sigma_W^{-1} \bar{W}_t)] \le R_c + E_P[\frac{1}{2T} \sum_{t=0}^{T-1} (\bar{Y}_t' M \bar{Y}_t)] \Big\}. \tag{4.13}$$

But, this class is independent of the control sequence. Thus, without loss of generality in addressing the uniform observability of the control/communication system of Figure 4.2, we can consider the uncontrolled uncertain system (4.12). Furthermore, following similar discussion that we had in (Chapter 2, Section 2.2), in addressing the necessary conditions for robust stability of the control/communication system of Figure 4.2, we can also consider the uncontrolled uncertain system (4.12). Subsequently, when we are concerned with the control/communication system of Figure 4.2, the definition of uniform observability 4.2.2 involves the existence of a control sequence, an encoder and a decoder. Furthermore, the distribution of the control sequence does not contribute to the expectation and the probability criterion (4.6).

4.3 Robust Entropy Rate

The uniform observability and robust stability of the control/communication system of Figure 4.1 are addressed by finding the entropy rate of the observations process. Since in this chapter the information source belongs to a class, the robust entropy rate must be calculated. The robust entropy is defined as the maximum of the Shannon entropy over the class of sources. When the class of sources is described by a relative entropy constraint, the explicit solution to the robust entropy is derived. Then, specific examples are presented to illustrate the theory. The results of this section is used in Section 4.4 to address the question of uniform observability and robust stability of the uncertain plants which are described via the relative entropy constraint. An example of such systems is the uncertain system (4.3) subject to the sum quadratic constraint (4.5).

4.3.1 Robust Entropy Rate-Definition and Solution

In this section, first we define the robust entropy and the robust entropy rate. Then, we provide the explicit solution to the robust entropy when the class of sources is described by a relative entropy constraint. The main results of this section which are concerned with the calculation of the robust entropy when the maximization is over the relative entropy constraint, are given in Lemma 4.3.2, Theorem 4.3.3, Lemma 4.3.4, Example 4.3.5 and Remark 4.3.6. This lemma, theorem, example and remark are new contributions of this chapter.

Definition 4.3.1 *(Robust Entropy and Robust Entropy Rate)*
i) Probabilistic Model. Let Y be a R.V. corresponding to a source with probability measure $P(dY) = f_Y dY$, where the density function $f_Y \in \mathcal{D}_{SU} \subset \mathcal{D}$ is unknown, $\mathcal{D}_{SU}$ is set of admissible source density functions and $\mathcal{D}$ is the set of all possible density functions associated with a R.V., Y. The robust entropy associated with the class $\mathcal{D}_{SU}$ is defined by

$$H_r(f_Y^*) \triangleq \sup_{f_Y \in \mathcal{D}_{SU}} H_S(f_Y) \tag{4.14}$$

where $H_S(f_Y)$ is the Shannon entropy (Definition 6.2.9).
Moreover, for the sequence Y^{T-1} with the unknown probability measure $P(dY^{T-1}) = f_{Y^{T-1}}$ dY^{T-1} where the joint density function $f_{Y^{T-1}} \in \mathcal{D}_{SU}^T \subset \mathcal{D}^T$ is unknown ($\mathcal{D}_{SU}^T$ is the set of all admissible joint density functions and $\mathcal{D}^T$ is the set of all possible joint density functions), the robust entropy rate associated with the family $\mathcal{D}_{SU}^T$ is defined by

$$\begin{aligned} \mathcal{H}_r(\mathcal{Y}) &\triangleq \lim_{T\to\infty} \frac{1}{T} H_r(f_{Y^{T-1}}^*), \\ H_r(f_{Y^{T-1}}^*) &= \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^T} H_S(f_{Y^{T-1}}), \end{aligned} \tag{4.15}$$

provided the limit exists.

ii) Frequency Domain Model. Consider a class of Random Processes (R.P's) $\{Y_t \in \Re^d; t \in \mathbf{N}_+\}$ *of Wide Sense Stationary (WSS) Gaussian processes, which generates a class of Power Spectral Density's (PSD's),* $S_Y(e^{jw}) \in \mathcal{P}_{SU}$.

The robust entropy rate associated with the class $\mathcal{P}_{SU}$ *is defined by*

$$\mathcal{H}_r(\mathcal{Y}) \triangleq \sup_{S_Y \in \mathcal{P}_{SU}} \mathcal{H}_S(S_Y), \tag{4.16}$$

where $\mathcal{H}_S(S_Y)$ *is the Shannon entropy rate of the WSS Gaussian random process* $\{Y_t \in \Re^d; t \in \mathbf{N}_+\}$ *in which from lemma 2.3.1 is given by*

$$\mathcal{H}_S(S_Y) = \frac{d}{2}\log_e(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \log_e \det S_Y(e^{jw})dw. \tag{4.17}$$

.

The above extension of entropy corresponds to the maximum amount of information generated by a class of sources.

Throughout this section, we are concerned with the relative entropy constraint set $\mathcal{D}_{SU}(g_Y)$ defined by

$$\mathcal{D}_{SU}(g_Y) \triangleq \{f_Y \in \mathcal{D}; H(f_Y||g_Y) \leq R_c + E_{f_Y}[L(Y)]\} \subset \mathcal{D}, \tag{4.18}$$

where $R_c \in [0, \infty)$ is fixed, g_Y is the nominal (known) density function and $L(Y) \geq 0$.

Next, consider the robust entropy definition (4.14) over the class of sources (4.18). That is,

$$\sup_{f_Y \in \mathcal{D}_{SU}(g_Y)} H_S(f_Y). \tag{4.19}$$

The constraint problem (4.19) can be converted into an unconstraint problem by introducing the Lagrangian function and its dual, as follows.

For $s \in [0, \infty)$, define the Lagrangian associated with (4.19) by

$$L(s, f_Y) \triangleq H_S(f_Y) - s\Big(H(f_Y||g_Y) - R_c - E_{f_Y}[L(Y)]\Big) \tag{4.20}$$

and its associated dual function by

$$L(s, f_Y^{*,s}) \triangleq \sup_{f_Y \in \mathcal{D}} L(s, f_Y). \tag{4.21}$$

In addition, define the quantity

$$L(s^*, f_Y^{*,s^*}) \triangleq \min_{s \geq 0} L(s, f_Y^{*,s}). \tag{4.22}$$

Next, by invoking the duality theorem (Lemma 6.2.33), we show that the constraint problem (4.19) is equivalent to the unconstraint problem (4.22).

Lemma 4.3.2 *The constraint problem (4.19) is equivalent to the unconstraint problem (4.22).*

Proof: It follows from duality theorem. See Appendix.

Next, by invoking Remark 6.2.28, i, the solution to the problem (4.19) is given by solving the unconstraint problem (4.22).

Theorem 4.3.3 *Suppose for some* $s \in [0, \infty)$, $(e^{L(y)} g_Y(y))^{\frac{s}{1+s}} \in L_1(\Re^d, \Re^+)$, *where* $L_1(\Re^d, \Re^+)$ *denotes the set of all integrable functions defined on* $\Re^d$ *which take values in* $\Re^+$ *(the set of non-negative real numbers).*

Then,

i)

$$L(s, f_Y^{*,s}) = sR_c + \log_e \int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy \tag{4.23}$$

and subsequently,

$$L(s^*, f_Y^{*,s^*}) = \min_{s \geq 0} \{sR_c + (1+s) \log_e \int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy\} \tag{4.24}$$

Further, if for some $s \geq 0$, $(\frac{s}{1+s} L(y) - \frac{1}{1+s} \log_e g_Y(y)) e^{\frac{s}{1+s} L(y)} g_Y(y)^{-\frac{1}{1+s}} \in L_1(\Re^d, \Re)$, *then the suprimizing* $f_Y^{*,s}$ *is given by*

$$f_Y^{*,s} = \frac{(e^{L(y)} g_Y(y))^{\frac{s}{1+s}}}{\int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy} \tag{4.25}$$

ii) If for some $s \in [0, \infty)$, $(e^{L(y)} g_Y(y))^{\frac{s}{1+s}} (L(y) + \log_e g_Y(y)) \in L_1(\Re^d, \Re)$ *and* $(e^{L(y)} g_Y(y))^{\frac{s}{1+s}} (L(y) + \log_e g_Y(y))^2 \in L_1(\Re^d, \Re^+)$. *Then,* $L(s, f_Y^{*,s^*})$ *is a convex function of* $s \geq 0$.

iii) Under the assumption that ii holds, the minimizing $s^* \geq 0$, *in which it minimizes* $L(s, f_Y^{*,s})$ *is the solution of the following equation*

$$T(s) \triangleq H(f_Y^{*,s} || g_Y) - E_{f_Y^{*,s}}[L(Y)] \Big|_{s=s^*} = R_c \tag{4.26}$$

That is, the solution, f_Y^{*,s^*}, *occurs on the boundary of the relative entropy uncertainty constraint (4.18).*

Moreover, $T(s)$ *is a non-increasing function of* $s \geq 0$, *that is*

$$0 \leq T(s)\Big|_{s=s_2} \leq T(s)\Big|_{s=s_1} \leq T(s)\Big|_{s=s^*} = R_c, \quad 0 \leq s^* \leq s_1 \leq s_2. \tag{4.27}$$

Proof: The proof is original and it is given in Appendix.

Above solution is related to the Rényi entropy $H_R(f_Y)$ (see (6.92) for definition) and subsequently to the Tsallis entropy $H_T(f_Y)$ (see (6.93)) as follows (notice that $H_R(f_Y) = \frac{1}{1-\alpha}\log_e(1+(1-\alpha)H_T(f_Y))$). Assume $s^* > 0$ and let $\alpha = \frac{s}{1+s}$, $s > 0$. Then,

$$\min_{\alpha\in(0,1)} H_R(e^{L(y)}g_Y) \leq H_r(f_Y^{*,s^*}) \leq \frac{\alpha}{1-\alpha}R_c + H_R(e^{L(y)}g_Y). \tag{4.28}$$

Clearly, the results of Theorem 4.3.3 are also applicable to the sequences of R.V.'s. We summarize it in the following Lemma.

Lemma 4.3.4 *Consider a sequence of R.V.'s, Y^{T-1} with the joint density function $f_{Y^{T-1}} \in \mathcal{D}_{SU}(g_{Y^{T-1}}) \subset \mathcal{D}^T$, where*

$$\mathcal{D}_{SU}(g_{Y^{T-1}}) \triangleq \{f_{Y^{T-1}} \in \mathcal{D}^T; \frac{1}{T}H(f_{Y^{T-1}}||g_{Y^{T-1}}) \leq R_c + E_{f_{Y^{T-1}}}[\frac{1}{T}L(Y^{T-1})]\}. \tag{4.29}$$

Then, the statement of Theorem 4.3.3 holds for the sequences of R.V.'s, Y^{T-1} by replacing $R_c \rightarrow TR_c$.

Next, in the following example, using the results of the Lemma 4.3.4, we find the robust entropy rate when the nominal source is Gaussian distributed.

Example 4.3.5 *Let the nominal source be the Gaussian distributed. That is, $g_{Y^{T-1}} \sim N(0,\Gamma_{Y^{T-1}})$, $\Gamma_{Y^{T-1}} \in \Re^{Td\times Td}$; and let $L(Y^{T-1}) = \frac{1}{2}\sum_{t=0}^{T-1} Y_t' M Y_t$; $M \in \Re^{d\times d}$, $M = M' \geq 0$. Further, assume $\bar{\sigma}(\Gamma_{Y^{T-1}}\bar{M}) < 1$ where $\bar{\sigma}(.)$ denotes the maximum singular value and $\bar{M} \triangleq MI_{Td} \in \Re^{Td\times Td}$ where $I_{Td} \in \Re^{Td\times Td}$ denotes the identity matrix.*
Then,

$$\begin{aligned} f_{Y^{T-1}}^{*,s^*} &\sim N\Big(0, \frac{1+s^*}{s^*}\Gamma_{Y^{T-1}}(I_{Td} - \Gamma_{Y^{T-1}}\bar{M})^{-1}\Big) \\ H_r(f_{Y^{T-1}}^{*,s^*}) &= \frac{Td}{2}\log_e(\frac{1+s^*}{s^*}) + H_S(g_{Y^{T-1}}) - \frac{1}{2}\log_e\det(I_{Td} - \Gamma_{Y^{T-1}}\bar{M}) \end{aligned} \tag{4.30}$$

where I_{Td} is the identity matrix with dimension Td and $s^* > 0$ is the solution of the following equation

$$\begin{aligned} -\frac{d}{2}\log_e\frac{1+s^*}{s^*} - \frac{d}{2} + \frac{1}{2T}\log_e\det(I_{Td} - \Gamma_{Y^{T-1}}\bar{M}) \\ + \frac{1}{2T}\frac{1+s^*}{s^*} trac[(I_{Td} - \bar{M}\Gamma_{Y^{T-1}})^{-1}(I_{Td} - \Gamma_{Y^{T-1}}\bar{M})] = R_c. \end{aligned} \tag{4.31}$$

Proof: It follows from Lemma 4.3.4 and by direct substitution.

Remark 4.3.6 *For the special case of* $M = 0$, $\forall t \in \mathbf{N}_+$, *Example 4.3.5 is reduced to the following results.*

$$H_r(f^{*,s^*}_{Y^{T-1}}) = \frac{Td}{2}\log_e \frac{1+s^*}{s^*} + H_S(g_{Y^{T-1}}) \tag{4.32}$$

in which, $s^* > 0$ *is the unique solution of the following equation*

$$-\frac{d}{2}\log_e \frac{1+s^*}{s^*} + \frac{d}{2s^*} = R_c. \tag{4.33}$$

Subsequently, if the Shannon entropy rate $\mathcal{H}_S(\mathcal{Y}) = \lim_{T\to\infty}\frac{1}{T}H_S(g_{Y^{T-1}})$ *exists, then the robust entropy rate is given by the following.*

$$\mathcal{H}_r(\mathcal{Y}) = \frac{d}{2}\log_e \frac{1+s^*}{s^*} + \mathcal{H}_S(\mathcal{Y}). \tag{4.34}$$

Proof: It follows from Example 4.3.5. See Appendix.

4.3.2 Example

In this section, the robust entropy rate is calculated for the uncontrolled version of the uncertain system (4.3) when it is subject to the sum quadratic constraint.
Furthermore, the robust entropy rate of a class of fully observed nonlinear system and the robust entropy rate of a class of sources which is described by the power spectral density, are calculated. These calculations are used in the next section to address the question of uniform observability and robust stability of these uncertain systems when they are controlled over limited capacity communication channels. The robust entropy rate which are calculated in this section, are new contributions of this chapter.

Uncertain Fully Observed Gauss Markov System

In this section, we are concerned with the following uncertain system

$$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t\geq 0}) : \left\{ \begin{array}{ll} X_{t+1} = AX_t + BW_t + B\bar{W}_t, & X_0 = X, \\ Y_t = H_t, & H_t = X_t \end{array} \right. \tag{4.35}$$

where $X_t \in \Re^q$, $W_t \in \Re^m$, $\bar{W}_t \in \Re^m$, $Y_t, H_t \in \Re^q$, $X_0 \sim N(\bar{x}_0, \bar{V}_0)$, W_t is i.i.d. $\sim N(0, \Sigma_W)$, $\Sigma_W > 0$, $\bar{W}_t$ is the perturbed (unknown) noise process which is $\{\sigma\{W_l\}; l \leq$

$t-1\}$ adapted ($\sigma\{\}$ denotes the σ-algebra); $\{X_0, W_t, \bar{W}_t\}$ are mutually independent and $\lim_{T\to\infty}\frac{1}{T}E_P[\sum_{t=0}^{T-2}\bar{W}_t'\Sigma_{\bar{W}}^{-1}\bar{W}_t] < \infty$.

The nominal system corresponding to the uncertain system (4.35) is the following fully observed system

$$(\Omega, \mathcal{F}(\Omega), \Pi; \{\mathcal{F}_t\}_{t\geq 0}) : \begin{cases} X_{t+1} = AX_t + BW_t, & X_0 = X, \\ Y_t = H_t, & H_t = X_t \end{cases} \tag{4.36}$$

In [59], it is shown that for the sequence Y^{T-1}, $H(P||\Pi) = \frac{1}{2}E_P[\sum_{t=0}^{T-2}\bar{W}_t'\Sigma_{\bar{W}}^{-1}\bar{W}_t]$.

For $P(dY^{T-1}) = f_{Y^{T-1}}dY^{T-1}$ and $\Pi(dY^{T-1}) = g_{Y^{T-1}}dY^{T-1}$ associated with the uncertain and the nominal systems (4.35) and (4.36), respectively, consider the following robust entropy rate problem.

$$\frac{1}{T}H_r(f^*_{Y^{T-1}}) = \sup_{f_{Y^{T-1}}\in\mathcal{D}_{SU}(g_{Y^{T-1}})}\frac{1}{T}H_S(f_{Y^{T-1}}) \tag{4.37}$$

where

$$\mathcal{D}_{SU}(g_{Y^{T-1}}) = \{f_{Y^{T-1}}; \frac{1}{T}H(f_{Y^{T-1}}||g_{Y^{T-1}}) \leq R_c + E_{f_{Y^{T-1}}}[\frac{1}{2T}\sum_{t=0}^{T-1}Y_t'MY_t]\} \tag{4.38}$$

Since

$$\begin{aligned} H_S(f_{Y^{T-1}}) &= -H(f_{Y^{T-1}}||g_{Y^{T-1}}) - E_P[\log_e g_{Y^{T-1}}], \\ H(f_{Y^{T-1}}||g_{Y^{T-1}}) &= \frac{1}{2}E_P[\sum_{t=0}^{T-2}\bar{W}_t'\Sigma_{\bar{W}}^{-1}\bar{W}_t], \end{aligned} \tag{4.39}$$

the problem (4.37) is equivalent to the following problem

$$\begin{aligned} \frac{1}{T}H_r(f^*_{Y^{T-1}}) &= \sup_{f_{Y^{T-1}}\in\mathcal{D}_{SU}(g_{Y^{T-1}})}\frac{1}{T}H_S(f_{Y^{T-1}}) \\ &= \sup_{\left\{\{\bar{W}_t\}_{t=0}^{T-2};\frac{1}{2T}E_P[\sum_{t=0}^{T-2}\bar{W}_t'\Sigma_{\bar{W}}^{-1}\bar{W}_t]\leq R_c+E_P[\frac{1}{2T}\sum_{t=0}^{T-1}Y_t'MY_t]\right\}} -f(\bar{W}^{T-2}) \end{aligned} \tag{4.40}$$

where

$$f(\bar{W}^{T-2}) = \frac{1}{2T}E_P[\sum_{t=0}^{T-2}\bar{W}_t'\Sigma_{\bar{W}}^{-1}\bar{W}_t] + \frac{1}{T}E_P[\log_e g_{Y^{T-1}}]. \tag{4.41}$$

Next, under the assumption that $B'(B\Sigma_W B')^{-1}B \leq \Sigma_{\bar{W}}^{-1}$, the conditions for applying Lagrange duality theorem (Lemma 6.2.33) are satisfied.

Subsequently, by applying Lemma 6.2.33, we have

$$\begin{aligned} \frac{1}{T}H_r(f^{*,s^*}_{Y^{T-1}}) &= \min_{s\geq 0}\sup_{\{\bar{W}_t\}_{t=0}^{T-2}}\Big\{-\frac{1+s}{2T}E_P[\sum_{t=0}^{T-2}\bar{W}_t'\Sigma_{\bar{W}}^{-1}\bar{W}_t] \\ &\quad -\frac{1}{T}E_P[\log_e g_{Y^{T-1}}] + sR_c + \frac{s}{2T}E_P[\sum_{t=0}^{T-1}Y_t'MY_t]\Big\}. \end{aligned} \tag{4.42}$$

Thus, the robust entropy rate is given by

$$\mathcal{H}_r(\mathcal{Y}) = \lim_{T\to\infty} \frac{1}{T} H_r(f^{*,s^*}_{Y^{T-1}}). \tag{4.43}$$

Next, the solution to the robust entropy problem (4.43) is given in the following theorem.

Theorem 4.3.7 *Consider the robust entropy problem (4.43) and (4.42). Let* $B'(B\Sigma_W B')^{-1}B < (1+s)\Sigma_W^{-1}$ *for some* $s \geq 0$. *Then,*

i)

$$\bar{W}_t^* = -[B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1} + B'\Xi_{t+1}B]^{-1}B'\Xi_{t+1}AX_t \tag{4.44}$$

where Ξ_t *is a real symmetric solution of the following Riccati equation.*

$$\begin{aligned} \Xi_t &= A'\Xi_{t+1}A - A'\Xi_{t+1}B[B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1} + B'\Xi_{t+1}B]^{-1}B'\Xi_{t+1}A + sM \\ \Xi_{T-1} &= sM \end{aligned} \tag{4.45}$$

and $s \geq 0$ *is the minimizing solution of the following equation*

$$Z(s^*) = \min_{s\geq 0}\{sR_c + \frac{1}{2T}trac(\Xi_0\bar{V}_0) + \frac{1}{2T}\sum_{t=1}^{T-1} trac(B'\Xi_t B\Sigma_W)\} \tag{4.46}$$

ii) If (A,B) *is controllable (Definition 6.2.37),* A *and* $B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1}$ *are invertible, and* $\beta(\eta) > 0$ *for some* η; $|\eta| = 1$ *where* $\beta(\eta)$ *is the rational matrix function given by the following.*

$$\beta(\eta) = B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1} + B'(\eta^{-1}I_q - A')sM(\eta I_q - A)^{-1}B, \quad s \geq 0. \tag{4.47}$$

Then

$$\mathcal{H}_r(\mathcal{Y}) = \frac{q}{2}\log_e(2\pi e) + \frac{1}{2}\log_e\det(B\Sigma_W B') + \min_{s\geq 0}\{sR_c + \frac{1}{2}trac(B'\Xi_\infty B\Sigma_W)\} \tag{4.48}$$

where Ξ_∞ *is the solution of the following Algebraic Riccati equation appearing in the* H^∞ *estimation and control problem*

$$\Xi_\infty = A'\Xi_\infty A - A'\Xi_\infty B[B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1} + B'\Xi_\infty B]^{-1}B'\Xi_\infty A + sM. \tag{4.49}$$

Proof: The proof is original and it is given in Appendix.

Remark 4.3.8 *Since the Algebraic Riccati equation (4.49) is quadratic, it has two symmetric solutions. Nevertheless, in (4.48) the solution under which* $\{sR_c + \frac{1}{2}trac(B'\Xi_\infty B\Sigma_W)\}$ *is bigger, must be used. Please note that conditions of Theorem 4.3.7, ii guarantee that the Algebraic Riccati equation (4.49) has real solutions. These conditions just need to be valid for* $s \geq 0$ *which minimizes* $\{sR_c + \frac{1}{2}trac(B'\Xi_\infty B\Sigma_W)\}$. *Furthermore, the condition* $B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1} < 0$ *is critical for the validity of the results. Similarly, this condition must be hold just for* $s \geq 0$ *which minimizes* $\{sR_c + \frac{1}{2}trac(B'\Xi_\infty B\Sigma_W)\}$.

Class of Fully Observed Nonlinear System

Consider the following fully observed nonlinear nominal system

$$(\Omega, \mathcal{F}(\Omega), \Pi; \{\mathcal{F}_t\}_{t\geq 0}) : \begin{cases} X_{t+1} = F_t(X^t) + BW_t, & X_0 = X, \\ Y_t = X_t, & t \in \mathbf{N}_+, \end{cases} \tag{4.50}$$

where $X_t \in \Re^q$ is the state process, $Y_t \in \Re^q$ is the observation process, $W_t \in \Re^m$ is i.i.d with distribution $W_t \sim N(0, \Sigma_W) = g_W$, $\Sigma_W > 0$, $X_0 \in \Re^q$ is the initial condition with distribution $X_0 \sim N(\bar{x}_0, \bar{V}_0) = g_{X_0}$ independent of W_t, and $F_t(.)$ defines a Borel measurable mapping $\Re^{q(t+1)} \longrightarrow \Re^q$.
Denote by $g_{Y^{T-1}} \in \mathcal{D}^T$ the joint density function of the sequence Y^{T-1} produced by the nominal system (4.50), and consider the relative entropy constraint (4.2) with $M = 0$. Since $\Pr(X_t \leq x_t | X^{t-1} = x^{t-1}) = \int_{X_t \leq x_t} g_W(X_t - F_t(x^{t-1}))dX_t$, then $g_{Y^{T-1}} = \prod_{t=1}^{T-1} g_W(x_t - F_t(x^{t-1}))g_{X_0}(x_0)$. Subsequently, the maximizing density is given by (4.25) with g_Y replaced by $g_{Y^{T-1}}$, while the entropy for this class is computed from Theorem 4.3.3 when R_c is replaced by TR_c.
Note that a specific uncertain system that can be described by the relative entropy constraint (4.2) with the nominal system (4.50), is the following system [59]

$$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t\geq 0}) : \begin{cases} X_{t+1} = F_t(X^t) + BW_t + B\bar{W}_t, & X_0 = X, \\ Y_t = X_t, & t \in \mathbf{N}_+, \end{cases} \tag{4.51}$$

where $X_t, Y_t \in \Re^q$, $F_t(.) : \Re^{q(t+1)} \longrightarrow \Re^q$ defines a Borel measurable function, $X_0 \sim N(\bar{x}_0, \bar{V}_0)$, $W_t \in \Re^m$ i.i.d. $\sim N(0, \Sigma_W)$, $\Sigma_W > 0$, (X_0, W_t) are independent, and $\bar{W}_t$ is a mean square summable perturbed noise process adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ which is subject to the following sum quadratic constraint.

$$H(f_{Y^{T-1}} || g_{Y^{T-1}}) = \frac{1}{2} \sum_{t=0}^{T-2} E_P[\bar{W}_t' \Sigma_W^{-1} \bar{W}_t] \leq TR_c. \tag{4.52}$$

In (4.52) $E_P[.]$ denotes expectation with respect to the probability measure P, $P(dY^{T-1}) = f_{Y^{T-1}}dY^{T-1}$ and $\Pi(dY^{T-1}) = g_{Y^{T-1}}dY^{T-1}$.
From Theorem 4.3.3 follows that for the class of systems described by the relative entropy constraint (4.2), we have the following result for the entropy rate of the class of systems.

Proposition 4.3.9 *For the class of systems described by the relative entropy constraint (4.2) with $M = 0$ and the nominal system (4.50), the robust entropy rate is given by*

$$\mathcal{H}_r(\mathcal{Y}) = \frac{q}{2}\log(\frac{1+s^*}{s^*}) + \mathcal{H}_S(\mathcal{Y}), \tag{4.53}$$

where $\mathcal{H}_S(\mathcal{Y}) = \frac{1}{2}\log_e\left((2\pi e)^q \det(B\Sigma_W B')\right)$ is the Shannon entropy rate of the nominal system; and for a given $R_c \in [0,\infty)$, $s^* > 0$ is the unique solution of the following nonlinear equation.

$$R_c = -\frac{q}{2}\log_e(\frac{1+s^*}{s^*}) + \frac{q}{2s^*}. \tag{4.54}$$

Proof: It follows from Theorem 4.3.3. See Appendix.

Remark 4.3.10 *For the special case of $F_t(X^t) = AX_t$, the robust entropy rate of (4.53) also represents the robust entropy rate of the uncontrolled version of uncertain system (4.3) subject to the sum quadratic constraint (with $M = 0$). This robust entropy rate calculation for such uncertain system is a valuable result since when $M = 0$, the conditions for having the robust entropy rate (4.48) do not hold.*

Class of Systems Described by the Power Spectral Density

Let $\beta(1) \triangleq \{z; |z| \geq 1\}$ where z is a complex number, and let H^∞ be the space of scalar, rational proper transfer functions of z (e.g., the space of $\frac{N(z)}{D(z)}$, where the degree of the polynomial $D(z)$ is greater or equal to the degree of the polynomial $N(z)$), in which these rational functions are analytic functions of $z \in \beta(1)$. When this space is endowed with the norm $||H||_\infty \triangleq \sup_{-\pi\leq w\leq\pi}|H(e^{jw})|$, $H(z) \in H^\infty$, $(z = e^{jw})$, then $(H^\infty, ||.||_\infty)$ is a Banach space.
The class of systems is obtained by passing a stationary Gaussian random process $\{X_t; t \in \mathbf{N}_+\}$, $X_t \in \Re$ with known PSD $S_X(e^{jw})$ through an unknown linear filter $\tilde{H}(z) \in H^\infty$, in which $\tilde{H}(z)$ belongs to the following class

$$\tilde{H}(z) \in \mathcal{D}_{ad} \triangleq \Big\{\tilde{H}(z) \in H^\infty; \tilde{H}(z) = H(z) + \Delta(z)W(z); \tilde{H}(z), H(z), \Delta(z), W(z) \in H^\infty,$$
$$H(z), W(z) \text{ are fixed (known)}, \Delta(z) \text{ is unknown with } ||\Delta||_\infty \leq 1\Big\}, \tag{4.55}$$

where $H(z) \in H^\infty$ is the nominal (known) system transfer function based on previous experience or belief, and $\Delta(z)W(z)$ represents the unknown (uncertain) part of the system. Clearly, this model implies $|\tilde{H}(e^{jw}) - H(e^{jw})| \leq |W(e^{jw})|$, $\forall w \in [-\pi, \pi]$ and thus the class (4.55) is described by a ball in the H^∞ space, centered at $H(e^{jw})$, with radius $W(e^{jw})$ which is a function of frequency.

Since $\{X_t; t \in \mathbf{N}_+\}$, $X_t \in \Re$ is a Gaussian random process and $\tilde{H}(z) \in H^\infty$, $\{Y_t; t \in \mathbf{N}_+\}$ is an asymptotic stationary Gaussian random process. Consequently, from Lemma 2.3.1 follows that the entropy rate of the class of systems obtained by passing $\{X_t; t \in \mathbf{N}_+\}$, $X_t \in \Re$ through the linear filter $\tilde{H}(z) \in H^\infty$, in nats per time step, is given by

$$\mathcal{H}_S(S_Y) = \frac{1}{2}\log_e(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \log_e S_Y(e^{jw})dw, \quad S_Y(e^{jw}) = |\tilde{H}(e^{jw})|^2 S_X(e^{jw}) \in \mathcal{P}_{SU}, \tag{4.56}$$

where $\mathcal{P}_{SU} \triangleq \{S_Y(e^{jw}); S_Y(e^{jw}) = |\tilde{H}(e^{jw})|^2 S_X(e^{jw}),\ \tilde{H}(z) \in \mathcal{D}_{ad}\}$. Subsequently, the robust entropy rate associated with the uncertain system is defined by $\mathcal{H}_r(\mathcal{Y}) = \sup_{S_Y \in \mathcal{P}_{SU}} \mathcal{H}_S(S_Y)$. Next, in the following proposition, the robust entropy rate associated with the family $\mathcal{P}_{SU}$ is calculated.

Proposition 4.3.11 *Let the observed process $\{Y_t; t \in \mathbf{N}_+\}$ is obtained by passing a stationary Gaussian random process $\{X_t \in \Re; t \in \mathbf{N}_+\}$ with the known PSD, $S_X(e^{jw})$ through an unknown linear filter $\tilde{H}(z) \in H^\infty$, described by (4.55). Then, the robust entropy rate of this system, in nats per time step, is given by*

$$\mathcal{H}_r(\mathcal{Y}) = \frac{1}{2}\log_e(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \log_e\Big((|H(e^{jw})| + |W(e^{jw})|)^2 S_X(e^{jw})\Big)dw, \tag{4.57}$$

where $|\Delta^*(e^{jw})| = 1$, $\arg(\Delta^*(e^{jw})) = \arg(H(e^{jw})) - \arg(W(e^{jw}))$.

Proof: The proof is original and it is given in Appendix.

One may consider different uncertain models as well as generalization of (4.56) to the vector case. However, this is beyond the scoop of this expository section.

4.4 Necessary Conditions for Observability and Stability

In this section, we develop a robust version of information transmission theorem and a robust version of the Shannon lower bound. Subsequently, by invoking these results, we find necessary conditions for uniform observability and robust stability in the form of a lower bound on the capacity in terms of the robust Shannon lower bound. In addition, the robust stability problem of the uncertain fully observed controlled Gauss Markov system (4.3) subject to the sum quadratic constraint (4.5) over AWGN channels, is also considered.

4.4.1 Robust Information Transmission Theorem and Robust Shannon Lower Bound

In this section, we develop a robust version of information transmission theory and the Shannon lower bound. These results are used in the next section to relate the channel capacity to the robust Shannon lower bound for uniform observability and robust stability. In the following, we present a robust version of the information transmission theorem. Here, we assume the channel is DMC's or AWGN channel. The following theorem is obtained following the same methodology used in [63] by implementing a slight modification.

Theorem 4.4.1 *(Robust Information Transmission Theorem) Consider the block diagram of Figure 4.1 subject to the conditional independence assumption (as discussed in Section 2.4.1). Let the source belong to the following class* $P(dY^{T-1}) \in \mathcal{M}_{SU}^{T} \subset \mathcal{M}_1(\Re^{Td})$. *Then, a necessary condition on n channel uses capacity, i.e.,* $\mathcal{C}_n$ *(Definition 6.2.18) for reproducing a sequence of the source messages* Y^{T-1} *by* $\tilde{Y}^{T-1}$ *up to the distortion value* D_v *at the communication end (i.e.,* $E\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v$, $\forall P(dY^{T-1}) \in \mathcal{M}_{SU}^{T} \subset \mathcal{M}_1(\Re^{Td})$*) using a sequence of the channel inputs and channel outputs with length n* ($T \leq n$) *is*

$$\mathcal{C}_n \geq R_{T,r}(D_v), \tag{4.58}$$

where $R_{T,r}(D_v)$ *is the robust (Sakrison) rate distortion function (Definition 6.2.35).*

Proof: See Appendix.

Finding an explicit solution for the robust rate distortion is very difficult. Thus, it is important to approximate it by a lower bound that can be computed easily. In the following lemma, we present a lower bound for the robust rate distortion in terms of the robust entropy of the source messages. The following lemma is obtained following a slight modification on the Shannon lower bound.

Lemma 4.4.2 *(Robust Shannon Lower Bound) Let* Y^{T-1}, $Y_t \in \Re^d$, $0 \leq t \leq T-1$ *be a sequence with length T of the source messages and* $P(dY^{T-1}) = f_{Y^{T-1}} dY^{T-1}$, *where the joint*

density function $f_{Y^{T-1}}$ belong to the set $f_{Y^{T-1}} \in \mathcal{D}_{SU}^T \subset \mathcal{D}^T$ where $\mathcal{D}_{SU}^T$ is the set of admissible joint density functions and $\mathcal{D}^T$ is the set of all possible joint density functions. 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 the difference distortion measure $\rho(y_t, \tilde{y}_t) = \rho(y_t - \tilde{y}_t) : \Re^d \to [0, \infty)$ is continuous.
Then, a lower bound for $\frac{1}{T}R_{T,r}(D_v)$ is given by

$$\frac{1}{T}R_{T,r}(D_v) \geq \sup_{f_{Y^{T-1}} \in \mathcal{D}_{SU}^T} \frac{1}{T}H_S(f_{Y^{T-1}}) - \max_{h \in G_D} H_S(h), \tag{4.59}$$

where $h(\xi)$ is a density function which belongs to the set G_D defined as follow.

$$G_D \triangleq \{h : \Re^d \to [0, \infty); \int_{\Re^d} h(\xi)d\xi = 1, \int_{\Re^d} \rho(\xi)h(\xi)d\xi \leq D_v, \xi \in \Re^d\}. \tag{4.60}$$

Moreover, when $\int_{\Re^d} e^{s\rho(\xi)}d\xi < \infty$ for all $s < 0$, then $h^*(\xi) \in G_D$ that maximizes $H_S(h)$ is given by

$$h^*(\xi) = \frac{e^{s\rho(\xi)}}{\int_{\Re^d} e^{s\rho(\xi)}d\xi}, \quad \int_{\Re^d} \rho(\xi)h^*(\xi)d\xi = D_v. \tag{4.61}$$

Subsequently, when the robust rate distortion $R_r(D_v)$ (Definition 6.2.35) and $\mathcal{H}_r(\mathcal{Y})$ exist, the robust Shannon lower bound is given by

$$R_r(D_v) \geq \mathcal{H}_r(\mathcal{Y}) - \max_{h \in G_D} H_S(h) \triangleq R_{S,r}(D_v), \tag{4.62}$$

where $\mathcal{H}_r(\mathcal{Y})$ is the robust entropy rate and $R_{S,r}(D_v)$ is the robust Shannon lower bound.

Proof: See Appendix.

When we have the equality (6.125) (see Appendix) for the robust rate distortion, then under certain conditions the lower bound described above is equivalent to the robust rare distortion function when distortion value D_v is small. Subsequently, this lower bound can be refereed as the robust version of the Shannon lower bound (which is a tight lower bound for the rate distortion function).

4.4.2 Necessary Conditions for Uniform Observability and Robust Stability

The main theorem which connects capacity, uniform observability and robust stability is given next. This is applied to the control/communication system of Figure 4.1 described by the stochastic uncertain control systems of Section 4.2. The following theorem is a direct result of Theorem 4.4.1 and Lemma 4.4.2 and it is obtained following similar methodology used in [9] and [11].

Theorem 4.4.3 *Consider the control/communication system of Figure 4.1 described by a class of sources ($Y_t \in \Re^d$) and under conditional independence assumption. 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

$$\mathcal{C} \geq \mathcal{H}_r(\mathcal{Y}) - \frac{1}{2}\log_e[(2\pi e)^d \det \Gamma_g] \triangleq R_{S,r}(D_v), \tag{4.63}$$

where $\mathcal{C}$ is the capacity, $\mathcal{H}_r(\mathcal{Y})$ is the robust entropy rate of the observations process, $R_{S,r}(D_v)$ is the robust Shannon lower bound, and Γ_g is the covariance matrix of the Gaussian distribution $h^*(\xi) \sim N(0,\Gamma_g)$, ($\xi \in \Re^d$) which satisfies

$$\int_{||\xi||>\delta} h^*(\xi)d\xi = D_v. \tag{4.64}$$

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

$$\mathcal{C} \geq \mathcal{H}_r(\mathcal{Y}) - \frac{d}{r} + \log_e\left(\frac{r}{dV_d\Gamma(\frac{d}{r})}\left(\frac{d}{rD_v}\right)^{\frac{d}{r}}\right) \triangleq R_{S,r}(D_v), \tag{4.65}$$

where $\Gamma(.)$ is the gamma function and V_d is the volume of the unit sphere (e.g., $V_d = Vol(S_d); S_d \triangleq \{\xi \in \Re^d; ||\xi|| < 1\}$).
Further, when $Y_t = H_t + \Gamma_t$ for the class of sources, (4.63) and (4.65) are also necessary conditions for robust stability in probability and r-mean, respectively.

Proof: See Appendix.

Remark 4.4.4 *We have the following remarks regarding the results of Theorem 4.4.3.*
i) First, we notice that the conditions (4.63) and (4.65) are new. These conditions are important in designing control/communication systems subject to the perturbations in the dynamical system. If the capacity is less than the bounds (4.63) and (4.65), there are no encoding scheme and/or stability scheme for uniform observability and/or robust stability. That is, conditions (4.63) and (4.65) provide fundamental limits on the capacity for uniform observability and/or robust stability of sequences.
ii) The lower bounds (4.63) and (4.65) given in Theorem 4.4.3 hold for any observed process and they are independent of the encoder and decoder information patterns.

Next, in the following corollary, we apply the results of Theorem 4.4.3 to the specific uncertain dynamical systems. The necessary conditions presented in the following corollary for uniform reliable data reconstruction and robust stability, are new.

Corollary 4.4.5 *i) For a class of sources which is described by the relative entropy constraint (4.2), a necessary condition for uniform observability is given by conditions (4.63) and (4.65) when these conditions are described by the robust entropy of Lemma 4.3.4. In particular, when the nominal source is Gaussian distributed, the robust entropy solution of Example 4.3.5 must be used in (4.63) and (4.65).*

ii) Consider the control/communication system of Figure 4.2 described by the uncertain system (4.3) subject to the sum quadratic constraint. Let $\{X_0, W_t, \bar{W}_t\}$ are mutually independent. Then, conditions (4.63) and (4.65) when they are described by the robust entropy rate of (Theorem 4.3.7, (4.48)) are necessary conditions for uniform observability and robust stability of the uncertain dynamical system (4.3) subject to the sum quadratic constraint.

iii) Consider a class of controlled systems described by relative entropy (4.2) with $M = 0$, where the nominal system is a controlled version of the nominal system (4.50) described by the following state space model.

$$(\Omega, \mathcal{F}(\Omega), \Pi; \{\mathcal{F}_t\}_{t\geq 0}) : \begin{cases} X_{t+1} = F_t(X^t) + G_t(U^t) + BW_t, & X_0 = X, \\ Y_t = X_t, & t \in \mathbf{N}_+, \end{cases} \tag{4.66}$$

Here, $U_t \in \Re^o$ and $G_t(.)$ defines a Borel measurable mapping. Then, following the same methodology used in Corollary 2.3.4, the robust entropy rate of this class of controlled systems is bounded below by the robust entropy rate of the uncontrolled analogous systems which is given by (4.53).

Subsequently, no matter what the information patterns of the encoder and decoder are (either A, B, or C), the lower bound (4.53) must be used in conditions (4.63) and (4.65) to address the necessary condition for uniform observability and/or robust stability of such uncertain systems.

iv) Let the class of systems is described by $Y(z) = \tilde{H}(z)X(z) + H_U(z)U(z)$, where $\tilde{H}(z)$ is an uncertain filter described by (4.55), $\{X_t \in \Re; t \in \mathbf{N}_+\}$ is a stationary Gaussian R.P. with known PSD, $S_X(e^{jw})$, $H_U(z) \in H^\infty$ is known filter and $\{U_t \in \Re; t \in \mathbf{N}_+\}$ (the control signal) is a Gaussian R.P. independent of the process $\{X_t; t \in \mathbf{N}_+\}$ with PSD, $S_U(e^{jw})$. Then, it can be easily shown that

$$\mathcal{H}_r(\mathcal{Y}) = \sup_{S_Y \in \mathcal{P}_{SU}} \mathcal{H}_S(\mathcal{Y})$$

$$\geq \frac{1}{2}\log_e(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi}\log_e(|H(e^{jw})| + |W(e^{jw})|)^2 S_X(e^{jw})dw \tag{4.67}$$

where

$$\begin{aligned} S_Y(e^{jw}) &= |\tilde{H}(e^{jw})|^2 S_X(e^{jw}) + |H_U(e^{jw})|^2 S_U(e^{jw}) \\ \mathcal{H}_S(\mathcal{Y}) &= \frac{1}{2}\log_e(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi}\log_e S_Y(e^{jw})dw \\ \mathcal{P}_{SU} &= \{S_Y(e^{jw}); \tilde{H}(z) \in \mathcal{D}_{ad}\}. \end{aligned} \tag{4.68}$$

Subsequently, no matter what the information patterns of the encoder and decoder are (either A, B, or C), the lower bound (4.67) must be used in conditions (4.63) and (4.65) to address the necessary condition for uniform observability and/or robust stability of such uncertain system. Furthermore, these results also hold for an arbitrary control signal (i.e., not necessarily Gaussian and independent of process X_t) if the encoder is of Class A or B and decoder is of Class B such that the reliability of the observations process associated with the controlled system is reduced to the equivalent reliability of the observations process of the uncontrolled analogous system.

4.4.3 Controlled Uncertain System

In this section we are concerned with the robust stability of the uncertain system (4.3) subject to the sum quadratic constraint (4.5), over AWGN channels. Throughout, we consider the case of $Y_t \in \Re$. The vector case is treated similarly. The problem considered in this section has not been considered elsewhere. Therefore, in this section a new encoding/decoding scheme and a stability scheme are proposed for robust stability of the control/communication systems.

Consider the control/communication system of Figure 4.3 described by the following fully observed uncertain system which is the system of (4.3) with $Y_t, X_t \in \Re$.

$$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t\geq 0}) : \begin{cases} X_{t+1} = AX_t + NU_t + BW_t + B\bar{W}_t, & X_0 = X, \\ Y_t = H_t, & H_t = X_t \end{cases} \tag{4.69}$$

where $X_t \in \Re$, $U_t \in \Re^o$, $W_t \in \Re^m$, $\bar{W}_t \in \Re^m$, $X_0 \sim N(\bar{x}_0, \bar{V}_0)$, $Y_t, H_t \in \Re$, W_t is i.i.d. $\sim N(0, \Sigma_W)$, $\Sigma_W > 0$, and $\bar{W}_t$ is the perturbed noise random process which is $\{\sigma\{W_l\}; l = 1, 2, ..., t-1\}$ adapted; and it is subject to the following sum quadratic constraint.

$$\Big\{\{\bar{W}_t\}_{t=0}^{T-2}; E_P[\frac{1}{2T}\sum_{t=0}^{T-2}(\bar{W}_t'\Sigma_W^{-1}\bar{W}_t)] \leq R_c + E_P[\frac{1}{2T}\sum_{t=0}^{T-1}(MY_t^2)]\Big\}, \tag{4.70}$$

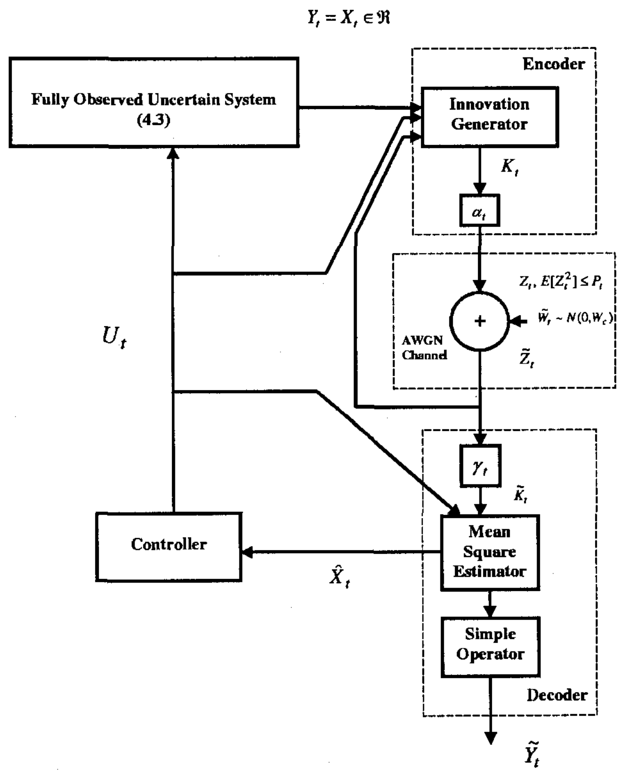

Figure 4.3: Encoder, decoder and controller for robust stability of the uncertain system

where R_c and M are non-negative scalars. The control/communication system of Figure 4.3 is described by the following AWGN channel.

$$\tilde{Z}_t = Z_t + \tilde{W}_t, \quad \tilde{W}_t \text{ orthogonal} \sim N(0, W_c), \quad E[Z_t^2] \le P_t < \infty, \tag{4.71}$$

where $Z_t \in \Re$ is the channel input, $\tilde{Z}_t \in \Re$ is the channel output and $\tilde{W}_t \in \Re$ is the channel noise.

Please note that in the control/communication system of Figure 4.3, α_t and γ_t are non-negative scalars.

Encoder and Decoder. The encoder is of type A and the decoder is of type B. Here, we use the same methodology used in Section 2.4.4 to design encoder and decoder. The encoder consists of a pre-encoding producing the innovations process $K_t = X_t - \hat{X}_t$, where $\hat{X}_t$ is the mean square state estimation in the presence of uncertainty in the plant, in which this estimation is obtained by the knowledge of U^{t-1} and $\tilde{Z}^{t-1}$ at the encoder (a recursive equation for $\hat{X}_t$ will be presented shortly). The decoder scales the channel outputs by γ_t and produces $\tilde{K}_t = \gamma_t \tilde{Z}_t$ ($\tilde{K}_t$ is the reconstruction of K_t); and subsequently it produces $\tilde{Y}_t = \tilde{K}_t + \hat{X}_t$ ($\tilde{Y}_t$ is the reconstruction of the observations process Y_t). The parameters α_t and γ_t must be chosen such that the innovations process is matched to the AWGN channel.

Robust Stabilizing Controller.

At the communication end, decoder receives the process $\tilde{K}_t$ which is produced by the following dynamic

$$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t\geq 0}) : \begin{cases} X_{t+1} = AX_t + NU_t + BW_t + B\bar{W}_t, \quad X_0 = X, \\ \tilde{K}_t = \gamma_t\alpha_t(X_t - \hat{X}_t) + \gamma_t\tilde{W}_t, \end{cases} \tag{4.72}$$

where $X_t, Y_t \in \Re$, $U_t \in \Re^o$, $W_t, \bar{W}_t \in \Re^m$, $X_0 \sim N(\bar{x}_0, \bar{V}_0)$, W_t i.i.d. $\sim N(0, \Sigma_W)$, $\Sigma_W > 0$. Let $U_t \in \mathcal{U}_t \triangleq \{U_t : \Re^{t(1+o)} \to \Re^o; U_t \in \mathcal{G}^U_{t-1}\}$, $\mathcal{G}^U_t \triangleq \sigma\{\tilde{K}_0, ..., \tilde{K}_t, U_0, ..., U_t\}$. Then, following the same methodology used in [59], by implementing the Legendre-Fenchel transformation (Remark 6.2.28, i) and by using the partial information risk sensitive optimal control results, a control U_t which stabilizes the uncertain system (4.72) (and subsequently the system (4.69)) in the sense that $\frac{1}{T}\sum_{t=0}^{T-1} E_P||X_t||^2 \leq D_v^c$, when it is subject to the sum quadratic constraint (4.70), is given by the followings.

$$U_t = -H^{-1}N'\Big[\Pi_{t+1} - \Pi_{t+1}[(NH^{-1}N' - \frac{B\Sigma_W B'}{\tau})^{-1} + \Pi_{t+1}]^{-1}\Pi_{t+1}\Big]A(1 - \frac{\Sigma_t\Pi_t}{\tau}) .\hat{X}_t. \tag{4.73}$$

$$\Pi_t = 1 + \tau M + A\Pi_{t+1}A - A\Pi_{t+1}[(NH^{-1}N' - \frac{B\Sigma_W B'}{\tau})^{-1} + \Pi_{t+1}]^{-1}\Pi_{t+1}A. \tag{4.74}$$

$$\begin{aligned} &t = 0, 1, ..., T-1 \\ &\Pi_T = 0 \\ &\Pi_{t+1}^{-1} - \frac{B\Sigma_W B'}{\tau} > 0 \\ &\Pi_t^{-1} - \frac{\Sigma_t}{\tau} > 0 \end{aligned} \tag{4.75}$$

$$\begin{aligned} \hat{X}_{t+1} &= A\hat{X}_t + NU_t + G_t\tilde{K}_t + A\Big[\Sigma_t - \Sigma_t[(\frac{\alpha_t^2}{W_c} - \frac{1}{\tau} - M)^{-1} + \Sigma_t]^{-1}\Sigma_t\Big](\frac{1}{\tau} + M)\hat{X}_t \\ \hat{X}_0 &= \bar{x}_0. \end{aligned} \tag{4.76}$$

$$G_t = A\Big[\Sigma_t - \Sigma_t[(\frac{\alpha_t^2}{W_c} - \frac{1}{\tau} - M)^{-1} + \Sigma_t]^{-1}\Sigma_t\Big]\frac{\alpha_t}{\gamma_t W_c}. \tag{4.77}$$

$$\begin{aligned} \Sigma_{t+1} &= B\Sigma_W B' + A\Sigma_t A - A\Sigma_t[(\frac{\alpha_t^2}{W_c} - \frac{1}{\tau} - M)^{-1} + \Sigma_t]^{-1}\Sigma_t A. \\ \Sigma_0 &= \bar{V}_0. \end{aligned} \tag{4.78}$$

$$\begin{aligned} &\Sigma_t^{-1} + \frac{\alpha_t^2}{W_c} - \frac{1}{\tau} - M > 0 \\ &\Sigma_t > 0 \end{aligned} \tag{4.79}$$

where $H \in \Re^{o \times o}$ is a positive definite matrix and $\tau > 0$ is a scalar which is defined shortly.

Next, in the limit as $T \to \infty$, the controller Riccati equation is reduced to the following Algebraic (indefinite) Riccati equation.

$$\Pi_\infty = 1 + \tau M + A\Pi_\infty A - A\Pi_\infty[(NH^{-1}N' - \frac{B\Sigma_W B'}{\tau})^{-1} + \Pi_\infty]^{-1}\Pi_\infty A. \quad (4.80)$$

Similarly, the observer Riccati equation (4.78) is reduced to the following Algebraic (indefinite) Riccati equation

$$\Sigma_\infty = B\Sigma_W B' + A\Sigma_\infty A - A\Sigma_\infty[(\frac{\alpha_\infty^2}{W_c} - \frac{1}{\tau} - M)^{-1} + \Sigma_\infty]^{-1}\Sigma_\infty A, \quad (4.81)$$

in which the solution to these Riccati equations are required to satisfy the followings. $\Pi_\infty^{-1} - \frac{B\Sigma_\infty B'}{\tau} > 0$, $\Pi_\infty^{-1} - \frac{\Sigma_\infty}{\tau} > 0$, $\Sigma_\infty > 0$, and $\Sigma_\infty^{-1} + \frac{\alpha_\infty^2}{W_c} - \frac{1}{\tau} - M > 0$.
For the case of $T \to \infty$, let $\tau > 0$ be the one that minimizes $\tau(\lim_{T\to\infty} \frac{\tilde{V}_\tau}{T} + R_c)$ where

$$\lim_{T\to\infty} \frac{\tilde{V}_\tau}{T} = -\frac{1}{2}\log_e[(1 - (\frac{1}{\tau} + M)\Sigma_\infty)] - \frac{1}{2}\log_e \Theta_\infty \quad (4.82)$$

in which, $\Theta_\infty = (1 - \frac{1}{\tau}G_\infty[\gamma_\infty^2 W_c + \gamma_\infty^2\alpha_\infty^2(\Sigma_\infty^{-1} - \frac{1}{\tau} - M)^{-1}]G_\infty(\Pi_\infty^{-1} - \frac{\Sigma_\infty}{\tau})^{-1})$, and $G_\infty = A\Big[\Sigma_\infty - \Sigma_\infty[(\frac{\alpha_\infty^2}{W_c} - \frac{1}{\tau} - M)^{-1} + \Sigma_\infty]^{-1}\Sigma_\infty\Big]\frac{\alpha_\infty}{W_c\gamma_\infty}$. Then, the controller (4.73) with Π_{t+1}, Σ_t and Π_t replaced by Π_∞, Σ_∞ and Π_∞, along with the estimator (4.76) stabilizes the uncertain system (4.72) (and subsequently the system (4.69)) in the sense that $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} E_P||X_t||^2 \leq D_v^c$ (where $D_v^c = 2\tau(\lim_{T\to\infty} \frac{\tilde{V}_\tau}{T} + R_c)$), when it is subject to the sum quadratic constraint (4.70). Please note that when $R_c = 0$ and $M = 0$ (i.e., the case without uncertainty) corresponds to the case where $\tau \to \infty$. For this case, the equations (4.73)-(4.79) are reduced to the standard Linear Quadratic Gaussian (LQG) results with the following cost functional $J = \frac{1}{2T}\sum_{t=0}^{T-1} E_P[X_t'X_t + U_t'HU_t]$.

Selecting α_t and γ_t:

In this section we find α_t and γ_t such that the innovations process K_t to be matched to the AWGN channel (as discussed in Section 2.4.4). The innovation process K_t is an orthogonal process; and the process K_t associated with the maximum entropy (robust entropy) and subsequently associated with the maximum rate distortion (maximization is over the class where the innovations process belong to), is the one with the following distribution $K_t \sim N(0, \Sigma_t)$, where Σ_t is the solution of the recursive equation (4.78). Subsequently, the minimizing kernel associated with the maximum rate distortion is given by

$$P^*(d\tilde{K}^{T-1}; k^{T-1}) = \Big(\prod_{t=0}^{T-1} q^*_{\tilde{K}_t|K_t}\Big)d\tilde{K}^{T-1}, \quad q^*_{\tilde{K}_t|K_t} \sim N(\eta_t k_t, \eta_t D_v), \eta_t = 1 - \frac{D_v}{\Sigma_t}, \quad (4.83)$$

where Σ_t is the solution of the recursive equation (4.78) and the distortion value D_v satisfies $D_v < \min_{t\in\mathbf{N}_+} \Sigma_t$.

Consequently, following the solution (4.83), if α_t and γ_t are chosen as follows, then the innovations process to reconstructions (i.e., K_t to $\tilde{K}_t$) behalves like the minimizing kernel (4.83).

$$\alpha_t = \sqrt{\frac{\eta_t W_c}{D_v}}, \quad \gamma_t = \sqrt{\frac{D_v \eta_t}{W_c}}, \quad \eta_t = 1 - \frac{D_v}{\Sigma_t}. \tag{4.84}$$

The power constraint of this encoding scheme is $E_P[Z_t^2] \leq \frac{\eta_t W_c}{D_v}\Sigma_t$. It is easily shown that when α_t and γ_t is given by (4.84), then $E_P(K_t - \tilde{K}_t)^2 \leq D_v$ over all values for the perturbed term (and subsequently $E_P(Y_t - \tilde{Y}_t)^2 \leq D_v$ where $\tilde{Y}_t = \tilde{K}_t + \hat{X}_t$ is the reconstruction of Y_t at the communication end). Furthermore, under assumption that the solution to the Riccati equation (4.78) converges to the solution of the Algebraic Riccati equation (4.81), then the capacity $\mathcal{C}$ is given as follow. $\mathcal{C} = \mathcal{H}_r(\mathcal{K}) - \frac{1}{2}\log_e(2\pi e D_v)$ where $\mathcal{H}_r(\mathcal{K}) = \frac{1}{2}\log_e(2\pi e) + \frac{1}{2}\log_e \Sigma_\infty$ is the robust entropy rate of the innovations process.

Remark 4.4.6 *i) From (4.65) and following above analysis follows that for a given distortion value D_v, $\mathcal{C} = R_{S,r}(D_v) = \mathcal{H}_r(\mathcal{K}) - \frac{1}{2}\log_e(2\pi e D_v)$ is the minimum capacity for uniform observability of the innovations process, where $R_{S,r}(D_v)$ is the robust Shannon lower bound of the innovations process.*

ii) In the real life application the desired stability criterion (i.e., $\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1} E_P||X_t||^2 \leq D_v^c$) is given. The desired stability criterion determines the admissible distortion value D_v via the relation $D_v^c = 2\tau(\lim_{T\to\infty}\frac{V_T}{T} + R_c)$ (where H is chosen to be small). Subsequently, by implementing the proposed encoding scheme and stability scheme we have uniform observability (of the observations as well as innovations process) up to the admissible distortion value D_v, while the capacity is as minimum as possible (for uniform observability of the innovations process); and the robust stability is being guaranteed. Please note that the necessary condition presented in Corollary 4.4.5,ii will be very useful to determine if our design is optimal for a given uncertain plant and an admissible distortion value D_v. In particular, if the capacity $\mathcal{C} = \mathcal{H}_r(\mathcal{K}) - \frac{1}{2}\log_e(2\pi e D_v)$ is close to the necessary condition presented in Corollary 4.4.5,ii, then we know that the stability is guaranteed when the capacity is also minimum for uniform observability of the observations process. This will happen when the distortion value D_v is sufficiently small.

4.5 Conclusion

In this chapter we addressed uniform observability and robust stability of discrete time uncertain systems. The uncertainty in the dynamical system is described by a relative entropy constraint which is a natural generalization of the sum quadratic uncertainty description. By developing and invoking a robust version of the information transmission theorem and the Shannon lower bound, we derived necessary conditions for uniform observability and robust stability. These conditions are given in terms of the robust Shannon lower bound. Robust Shannon lower bound is given in terms of the robust entropy rate of the inputs to the encoder and an extra term which is related to the observability and stability criteria. We have shown that under certain conditions these necessary conditions are also sufficient. From the results of this chapter, it is concluded that the (robust) Shannon lower bound is also an adequate measure for describing conditions for uniform observability and robust stability of sequences. It can be easily calculated (Section 4.3.2) and it is applicable to time domain (Section 4.3.2) and frequency domain (Proposition 4.3.11). Furthermore, it can be applied to nonlinear systems (Proposition 4.3.9).

Chapter 5

Conclusion

5.1 Synopsis

In the present thesis we addressed the problem of reliable data reconstruction and stability of dynamical systems controlled over limited capacity communication channels. Throughout, it has been shown that the Shannon lower bound is an adequate measure for describing conditions for reliable data reconstruction and stability of sequences. This conclusion was obtained throughout the thesis by considering the discrete time case in Chapter 2 and discrete time case subject to uncertainty in Chapter 4.

In Chapter 2 using the information theoretic tools and by invoking the information transmission theorem and the Shannon lower bound, we derived necessary conditions for reliable data reconstruction and stability of general discrete time systems in terms of Shannon lower bound. These conditions are independent of the information patterns of encoder, decoder, and controller; and they are applicable to both linear and nonlinear systems and time and frequency domains. Using these necessary conditions, some of necessary conditions found in the literature (e.g., [41]) were obtained as a special case. It has been also shown that under certain conditions these necessary conditions are also sufficient. The obtained conditions are important because the already existing conditions in the literature have been developed for specific dynamical systems and observability and stability criteria (e.g., [4]-[13]). Furthermore, from these conditions we can conclude that, unlike [4], the Shannon capacity is an adequate measure for describing the conditions for moment observability and stability. We then applied the results to linear stochastic partially observed control systems subject to measurement noise over AWGN channels. Here, we derived an encoder, decoder, and controller for mean square stability and reconstruction. These results have extended the results of [11] and [42] which are concerned with stability of fully observed systems over AWGN channels. Shannon lower bound appeared to be quite nice to use because it was related to the existing necessary and sufficient conditions derived in the literature; and it was applicable to time domain and frequency domain methods.

In Chapter 3 we considered the mean square reliable data reconstruction and stability of a fully observed linear continuous time system driven by Brownian motion over flat fading AWGN channels. The problem considered in this chapter was the continuous version of the problem considered in [11] and [42]. In this chapter, it has been shown that the summation of the real parts of the unstable eigenvalues of the open loop time-invariant systems is the continuous version of the eigenvalue rate condition [4]-[13] for continuous time-invariant systems controlled over analog channels.

In Chapter 4 we considered uniform reliable data reconstruction and robust stability of uncertain dynamical systems controlled over limited capacity communication channels. By developing and invoking a robust version of the information transmission theory and the Shannon lower bound, we derived necessary conditions for uniform reliable data reconstruction and robust stability of sequences. The uncertainty in the dynamical system was described by a relative entropy constraint. As it was shown such uncertainty description is a natural generalization of the sum quadratic uncertainty description. These necessary conditions are given in the form of a lower bound on the Shannon capacity in terms of the robust Shannon lower bound. Throughout, we applied these conditions to several uncertain dynamical systems by calculating the corresponding robust entropy rate. Furthermore, a relation between the robust entropy rate of an uncertain system with the Algebraic Riccati equation of the type appearing in the H^∞ estimation and control problem was established. Under certain conditions, these necessary conditions were also sufficient. Subsequently, it has been concluded that the robust Shannon lower bound is also an adequate measure for describing conditions for uniform reliable data reconstruction and robust stability of sequences. This chapter particularly complemented the already existing results presented in [4], [9], [10] and [16] by considering the sum quadratic uncertainty description.

5.2 Direction for the Future Research

In the present thesis we addressed some basic problems in reliable data reconstruction and stability of dynamical systems controlled over limited capacity communication channels. In particular, we addressed reliable data reconstruction and stability problem of linear systems over AWGN channels. Nevertheless, AWGN channel is a basic model for wireless communication channels. Wireless communication channels are mostly subject to fading, interference and sometimes transmission delay. Although in Chapter 3 we considered slow flat fading AWGN channels when the criteria for reliable reconstruction was mean square criterion, throughout this chapter it was assumed that the transmitter and receiver knew the fading process. Nevertheless, for fast fading or frequency nonselective multiple fading channels, in which the received carrier amplitude is modeled by the fading amplitude, it is hard for transmitter and receiver to measure the fading process. They may only know the distribution of this process. For such channels the expected value of the mean square criterion with respect to the fading process (i.e., $E_\theta[(x(t)-\tilde{x}(t))^2|\mathcal{F}_{0,t}^{y,\theta}])$ which is independent of the fading process, can be used to design decoding scheme. It is clear that when the fading process is slow, this design scheme is optimal. Nevertheless, the lack of knowledge of the value of the fading process in the transmitter results in smaller capacity. In addition, the capacity of fading channels for non-i.i.d. fading process when just the statistic of the fading process is known, is an open problem for almost all fading distributions. Furthermore, how we can extend our results for the encoder (presented in Chapter 3) to achieve this capacity must be investigated.

On the other hand; most of the controlled systems are subject to the nonlinear terms. In the present thesis the obtained conditions for reliable data reconstruction and stability were given in terms of the information theoretic measures such as Shannon lower bound. Subsequently, the obtained conditions (particularly, necessary conditions) are applicable to nonlinear dynamical systems (see Section 4.3.2). By designing an encoding scheme and a stability scheme that guarantee reliability/stability of a given nonlinear system, by transmitting with a rate close to the bounds found as necessary conditions, we can also address the reliability/stability of the nonlinear systems controlled over limited capacity communication channels.

Furthermore, throughout the thesis we considered the cases where there was high capacity communication link from the controller to the dynamical system. Nevertheless, in many practical applications this assumption may not be feasible and the effects of limited capacity communication constraint in both feedforward (i.e., from sensors to controller) and feedback (i.e., from controller to the dynamical system) must be considered.

Therefore, for creating a mathematical framework which can be used to address real life problems, we suggest the following extension on the results of the present thesis for the future direction.

i) Reliable data reconstruction and stability subject to fading, interference and transmission delay.

ii) Reliable data reconstruction and stability of nonlinear dynamical systems.

iii) Reliable data reconstruction and stability subject to communication constraints in both feedforward and feedback direction.

Publications. Some of the contributions of the current thesis have been already published in the conference proceedings; or they are accepted for publication in the refereed journals. Below is the list of these publications.

Articles accepted in refereed journals

1- C. D. Charalambous, Alireza Farhadi, and S. Z. Denic, Control of continuous-time linear Gaussian systems over additive Gaussian wireless fading channels: a separation principle, accepted in the IEEE Transactions on Automatic Control. It was scheduled to be published in the March edition of transactions in 7 pages.

2- C. D. Charalambous and Alireza Farhadi, LQG optimality and separation principle for general discrete time partially observed systems over finite capacity communication channels, accepted in Automatica.

3- Alireza Farhadi and C. D. Charalambous, Robust coding for a class of sources: applications in control and reliable communication over limited capacity channels, accepted (subject to a minor revision) in Systems and Control Letters.

Other contributions (e.g., communications, papers in refereed conference proceedings, etc.)

1- Alireza Farhadi and C. D. Charalambous, Robust stabilizing scheme for uncertain systems

controlled over limited capacity Additive White Gaussian Noise channels, in proceedings of the 2008 American Control Conference, June 11-13, 2008.

2- Alireza Farhadi and C. D. Charalambous, Robust control of feedback systems subject to limited capacity constraints, in proceedings of 46th IEEE Conference on Decision and Control, December 2007.

3- C. D. Charalambous and Alireza Farhadi, Control of feedback systems subject to the finite rate constraints via Shannon lower bound, in 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Cyprus, April 16-20, 2007.

4- Alireza Farhadi and C. D. Charalambous, Control of tele-operation systems subject to capacity limited channels and uncertainty, in proceedings of CCECE/CCGEI, Ottawa, May 7-10, 2006, pp. 492-497.

5- C. D. Charalambous and Alireza Farhadi, A mathematical framework for robust control over uncertain communication channels, in proceedings of 44th IEEE Conference on Decision and Control, Seville, Spain, December 12-15, 2005, pp. 2530-2535.

6- C. D. Charalambous, S. Denic and Alireza Farhadi, Control over wireless communication channel for continuous-time systems, in proceedings of 44th IEEE Conference on Decision and Control, Seville, Spain, December 12-15, 2005, pp. 3225-3230.

7- C. D. Charalambous and Alireza Farhadi, Robust entropy rate for uncertain sources and its applications in controlling systems subject to capacity constraints, in proceedings of 43rd Annual Allerton Conference on Communication, Control, and Computation, Allerton House, Chicago, September 28-30, 2005.

8- C. D. Charalambous, Alireza Farhadi, S. Denic and F. Rezaei , Robust control over uncertain communication channels, in proceedings of 13th Mediterranean Conference on Control and Automation, Cyprus, June 27-29, 2005, pp. 737-742.

9- C. D. Charalambous and Alireza Farhadi, Robust entropy rate for uncertain sources: applications to communication and control systems, in proceedings of 9th Canadian Workshop on Information Theory, Montreal, June 5-8, 2005, pp. 307-310.

10- C. D. Charalambous, F. Rezaei, S. Denic, A. Kyprianou, and Alireza Farhadi, Robust information transmission and control subject to uncertainty and power constraints, workshop H-2 (half a day) presented by above authors in 43rd IEEE Conference on Decision and Control, Bahamas, December 2004.

Chapter 6

Appendix

6.1 Proofs

6.1.1 Chapter 2

Proof of Corollary 2.3.4.

Let Y^{T-1} be a sequence with length T of the observation process corresponding to (2.1). Then, from chain rule of the Shannon entropy (Remark 6.2.12, ii) and Remark 6.2.12, i, it follows that

$$H_S(Y^{T-1}) = \sum_{i=0}^{T-1} H_S(Y_i|Y^{i-1}) \geq \sum_{i=0}^{T-1} H_S(Y_i|Y^{i-1}, U^{i-1}). \tag{6.1}$$

The conditional probability density function $f_{Y_i|Y^{i-1},U^{i-1}}(y_i)$ corresponding to the R.V. Y_i conditioned on (Y^{i-1}, U^{i-1}) is Gaussian distributed. Consequently,

$$\begin{aligned}
&H_S(Y_i|Y^{i-1}, U^{i-1}) \\
&\triangleq \int_{\Re^{id}\times\Re^{io}} \Big(- \int_{\Re^d} \log[f_{Y_i|Y^{i-1},U^{i-1}}(y_i)] f_{Y_i|Y^{i-1},U^{i-1}}(y_i) dy_i \Big) f_{Y^{i-1},U^{i-1}} dy^{i-1} du^{i-1} \\
&= \int_{\Re^{id}\times\Re^{io}} \Big(\frac{d}{2}\log(2\pi e) + \frac{1}{2}\log\det Cov(Y_i|Y^{i-1}, U^{i-1}) \Big) f_{Y^{i-1},U^{i-1}} dy^{i-1} du^{i-1}, \qquad (6.2)
\end{aligned}$$

where $Cov(Y_i|Y^{i-1}, U^{i-1}) \triangleq E\Big[(Y_i - E(Y_i|Y^{i-1}, U^{i-1}))(Y_i - E(Y_i|Y^{i-1}, U^{i-1}))'\Big|Y^{i-1}, U^{i-1}\Big]$. Consequently, by the total probability theorem

$$\begin{aligned}
H_S(Y_i|Y^{i-1}, U^{i-1}) &= \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log\det E\Big[(Y_i - E(Y_i|Y^{i-1}, U^{i-1})) \\
&\qquad .(Y_i - E(Y_i|Y^{i-1}, U^{i-1}))'\Big] \\
&= \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log\det[CV_iC' + DD'], \qquad (6.3)
\end{aligned}$$

where V_i is defined in Lemma 2.3.2.

Subsequently, from (6.1), it follows that

$$H_S(Y^{T-1}) \geq \frac{Td}{2}\log(2\pi e) + \frac{1}{2}\sum_{i=0}^{T-1}\log\det[CV_iC' + DD']. \tag{6.4}$$

Next, using Cesaro average formula (Lemma 6.2.32), $\lim_{T\to\infty}\frac{1}{T}\sum_{i=0}^{T-1}\log\det[CV_iC' + DD']$ $= \log\det\Lambda_\infty$. Subsequently, dividing both sides of (6.4) by T and taking $T\to\infty$, yields the result.

Proof of Theorem 2.4.2.

Under the assumption that there exists an encoding scheme that provides a reliable data reconstruction up to the distortion value D_v, from data processing inequality, it follows that

$$I(Z^{n-1};\tilde{Z}^{n-1}) \geq I(Y^{T-1};\tilde{Y}^{T-1}). \tag{6.5}$$

subsequently,

$$\begin{aligned}\mathcal{C}_n = \sup_{P(dZ^{n-1})\in\mathcal{M}_{CI}} I(Z^{n-1};\tilde{Z}^{n-1}) &\geq I(Y^{T-1};\tilde{Y}^{T-1})\\ &\geq \inf_{P(d\tilde{Y}^{T-1};y^{T-1})\in\mathcal{M}_{DC}} I(Y^{T-1};\tilde{Y}^{T-1}) = R_T(D_v).\end{aligned} \tag{6.6}$$

That is, (6.6) is a necessary condition for existence of such encoding scheme.

Proof of Theorem 2.4.3. (Observability). Assume there exists an encoder/decoder pair such that the observability in probability is obtained. This implies that 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 the distortion measure $\rho_t(Y^{t-1};\tilde{Y}^{t-1})\triangleq\frac{1}{t}\sum_{k=0}^{t-1}\rho(Y_k,\tilde{Y}_k)$, where $\rho(.,.)$ is defined as follow $\rho(Y,\tilde{Y}) = \begin{cases}1, & ||Y-\tilde{Y}||>\delta\\ 0, & ||Y-\tilde{Y}||\leq\delta\end{cases}$ Then, for $t\geq T(\delta,D_v)$,

$$E\rho_t(Y^{t-1},\tilde{Y}^{t-1}) = \frac{1}{t}\sum_{k=0}^{t-1}E\rho(Y_k,\tilde{Y}_k) = \frac{1}{t}\sum_{k=0}^{t-1}\Pr(||Y_k-\tilde{Y}_k||>\delta)\leq D_v. \tag{6.7}$$

That is, a rate distortion with distortion value D_v is obtained for $t\geq T(\delta,D_v)$. Then by Theorem 2.4.2 the channel capacity and rate distortion must for all $t\geq T(\delta,D_v)$ satisfy

$$\begin{aligned}\frac{1}{t}\mathcal{C}_t\geq\frac{1}{t}R_t(D_v) &\Leftrightarrow \lim_{t\to\infty}\frac{1}{t}\mathcal{C}_t\geq\lim_{t\to\infty}\frac{1}{t}H_S(Y^{t-1})-\max_{h\in G_D}H_S(h)\\ &\Leftrightarrow \mathcal{C}\geq\mathcal{H}_S(\mathcal{Y})-\max_{h\in G_D}H_S(h).\end{aligned} \tag{6.8}$$

Since among all distributions with the same covariance, the Gaussian distribution has the biggest entropy (Remark 6.2.12, v), $h^*(\xi) \in G_D$ that maximizes $H_S(h)$ is a Gaussian distributed which occurs on the boundary of G_D. That is, $h^*(\xi) \sim N(0,\Gamma_g)$ in which Γ_g satisfies (2.15). Consequently, letting $H_S(h^*) = \frac{1}{2}\log[(2\pi e)^d \det \Gamma_g]$ (Remark 6.2.12, iv), the result follows.

Necessary condition for observability in r-mean is obtained along the same lines of the above proof. The only difference is that from Remark 6.2.25, iii, it follows that for this case, $\max_{h\in G_D} H_S(h) = \log e^{\frac{d}{r}} - \log(\frac{r}{dV_d\Gamma(\frac{d}{r})}(\frac{d}{rD_v})^{\frac{d}{r}})$ bits per time step.

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

Proof of Corollary 2.4.5. Assume the system of Figure 2.3 is r-mean stable via the linear time-invariant stable controller $-K_c(z)$.

From Figure 2.3, it follows that

$$Y(z) = S(z)F(z)W(z) + S(z)DG(z) - S(z)L(z)\tilde{W}(z), \tag{6.9}$$

where $S(z) = \frac{1}{1+L(z)}$ is the closed loop sensitivity transfer function and $W(z)$ and $G(z)$ are the z-transformation of the Gaussian processes $\{W_t; t \in \mathbb{N}_+\}$ and $\{G_t; t \in \mathbb{N}_+\}$, respectively. Since we have assumed the r-mean stability, $S(z)$ is stable transfer function (e.g., $S(z)$ has poles inside the unit circle). r-mean stability implies that the jointly Gaussian process $\{Y_t; t \in \mathbb{N}_+\}$ is a stationary Gaussian process. Consequently, from Lemma 2.3.1, it follows that

$$\mathcal{H}_S(\mathcal{Y}) = \frac{1}{2}\log(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \log S_Y(e^{jw})dw, \tag{6.10}$$

where from (6.9), $S_Y(e^{jw}) = |S(e^{jw})|^2\Big(F(e^{jw})F'(e^{-jw}) + DD' + L(e^{jw})L'(e^{-jw})W_c\Big)$. Then, an application of the Bode integral formula (Theorem 6.2.41) results that

$$\begin{aligned}\frac{1}{4\pi}\int_{-\pi}^{\pi} \log S_Y(e^{jw})dw \;&=\; \sum_{\{i;|\lambda_i(A)|\geq 1\}} \log|\lambda_i(A)| \\ &+\frac{1}{4\pi}\int_{-\pi}^{\pi} \log(F(e^{jw})F'(e^{-jw}) + DD' + L(e^{jw})L'(e^{-jw})W_c)dw.\end{aligned} \tag{6.11}$$

Consequently, substituting the resulting $\mathcal{H}_S(\mathcal{Y})$ in (2.16), we obtain the desired result.

6.1.2 Chapter 3

Proof of Theorem 3.3.1.

Assume there exist a stabilizing controller, and an encoder/decoder pair such that the system (3.3) over the communication channel (3.4) is stable. Define $l(t) \triangleq \int_0^t y(s)ds$. Then, $l(t) = \int_0^t f(s)ds + \int_0^t n(s)ds$, a.s. Since $v(t) = \int_0^t n(s)ds$ is Brownian motion, then we can apply Lemma 6.2.15 to get

$$I_T(l;x) = \frac{1}{2N_0}E[\int_0^T |f(t)-\hat{f}(t)|^2dt] = \frac{1}{2N_0}\int_0^T E[|f(t)-\hat{f}(t)|^2]dt = \frac{1}{2N_0}\int_0^T \Sigma_t dt, \quad (6.12)$$

where $\hat{f}(t) \triangleq E[f(t)\big|\mathcal{F}^y_{0,t}]$, and $\Sigma_t \triangleq E[|f(t)-\hat{f}(t)|^2]$. From data processing inequality (Remark 6.2.14, ii), it follows that

$$I_T(y;x) \geq I_T(l;x). \quad (6.13)$$

Thus,

$$I_T(y;x) \geq \frac{1}{2N_0}\int_0^t \Sigma_t dt. \quad (6.14)$$

On the other hand, the mean square error is related to the power spectral densities via Lemma 6.2.15 as follow

$$\lim_{T\to\infty}\Sigma_T = \frac{N_0}{2\pi}\int_{-\infty}^{+\infty}\log_e(1+\frac{S_f(w)}{N_0})dw, \quad (6.15)$$

where $S_f(w)$ is the PSD of f. Moreover, from Lemma 6.1.1

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \Sigma_t dt = \lim_{T\to\infty}\Sigma_T = \frac{N_0}{2\pi}\int_{-\infty}^{+\infty}\log_e(1+\frac{S_f(w)}{N_0})dw. \quad (6.16)$$

Next, an application of Bode integral formula (Theorem 6.2.42) results that

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{2N_0}\frac{1}{T}\int_0^T \Sigma_t dt &= \frac{1}{4\pi}\int_{-\infty}^{+\infty}\log_e\left(1+\frac{S_f(w)}{N_0}\right)dw \\ &= \frac{1}{4\pi}\int_{-\infty}^{+\infty}\log_e\left(\frac{N_0+S_f(w)}{N_0}\right)dw = \frac{1}{4\pi}\int_{-\infty}^{+\infty}\log_e\frac{S_y(w)}{S_n(w)}dw \\ &= \frac{1}{4\pi}\int_{-\infty}^{+\infty}\log_e|S(w)|^2dw = \sum_{\{i:Re(\lambda_i(A))\geq 0\}} Re(\lambda_i(A)). \quad (6.17)\end{aligned}$$

Please notice that in above derivation, the Bode integral formula is employed under mild assumption that the stabilizing controller is chosen such that the corresponding open loop transfer function (i.e., multiplications of channel, encoder, plant, controller, and decoder transfer functions) is strictly proper with degree at least two. Next, from (6.14) and (6.17), it follows that

$$\begin{aligned}\sup_{(x,f)\in\mathcal{X}\times\mathcal{F}_{ad}}\frac{1}{T}I_T(y;x) &\geq \frac{1}{T}I_T(y;x)\geq\frac{1}{2N_0}\frac{1}{T}\int_0^T\Sigma_t dt\\ \mathcal{C}\triangleq\lim_{T\to\infty}\sup_{(x,f)\in\mathcal{X}\times\mathcal{F}_{ad}}\frac{1}{T}I_T(y;x) &\geq \lim_{T\to\infty}\frac{1}{2N_0}\frac{1}{T}\int_0^T\Sigma_t dt=\sum_{\{i:Re(\lambda_i(A))\geq 0\}}Re(\lambda_i(A)).\end{aligned}\tag{6.18}$$

That is, $\mathcal{C}\geq\sum_{\{i:Re(\lambda_i(A))\geq 0\}}Re(\lambda_i(A))$ is a necessary condition for existence of a stabilizing controller.

Lemma 6.1.1 *If* $\lim_{t\to\infty}a_t=a$ *and* $b_t=\frac{1}{t}\int_0^t a_s ds$*, then* $\lim_{t\to\infty}b_t=a$.

Proof: As $\lim_{t\to\infty}a_t=a$, $\forall\epsilon>0,\exists\, T(\epsilon)>0$; such that$|a_t-a|\leq\epsilon$, $\forall t\geq T(\epsilon)$. Hence

$$\begin{aligned}|b_t-a|=|\frac{1}{t}\int_0^t(a_s-a)ds| &\leq \frac{1}{t}\int_0^t|a_s-a|ds\\ &\leq \frac{1}{t}\int_0^{T(\epsilon)}|a_s-a|ds+\frac{1}{t}\int_{T(\epsilon)}^t|a_s-a|ds\\ &\leq \frac{1}{t}\int_0^{T(\epsilon)}|a_s-a|ds+\frac{t-T(\epsilon)}{t}\epsilon,\quad\forall t\geq T(\epsilon).\end{aligned}\tag{6.19}$$

Since the first right side term in (6.19) goes to 0 as $t\to\infty$, we can make $|b_t-a|\leq\epsilon$ by taking t large enough. Hence, $\lim_{t\to\infty}b_t=a$ as $t\to\infty$.

Proof of Theorem 3.4.3.

(**Sufficient part**) i) Suppose (3.16) holds. Then, from (6.109), it follows that by using the optimal encoder and decoder proposed in Theorem 6.2.21, $\limsup_{t\to\infty}V^*(t,y,\theta)<\infty$ a.s (see Lemma 6.1.2, i,ii).

ii) This follows from Lemma 6.1.2, i, along the same lines of i.

iii) Follows easily from Lemma 6.1.2, iii.

(**Necessary part**) iii) Consider the case of $G\neq 0$. If condition (3.18) is not satisfied, then $A>\frac{P}{2}$ and consequently, $V^*(t,y,\theta)\to\infty$, as $t\to\infty$, a.s. Therefore, since among all admissible encoding/decoding schemes (including nonlinear ones), the proposed encoding/decoding scheme is optimal (*i.e.*, $\mathcal{E}(t)\geq V^*(t,y,\theta)$, a.s.), the mean square estimation error $\mathcal{E}(t)$ associated with all other admissible encoding scheme is going to be unbounded asymptotically, a.s. This implies that condition (3.18) is a necessary condition for bounded asymptotic observability. The result for the case of $G = 0$ follows similarly.

Lemma 6.1.2 *Let* $\epsilon = \inf_{t\in[0,\infty)}(Pz^2(t,\theta(t)) - 2[A(t)]^+) > 0$, *a.s. Then,*
i)

$$\lim_{t\to\infty} e^{(2\int_0^t A(s)ds - \int_0^t z^2(s,\theta(s))Pds)} = 0. \tag{6.20}$$

ii) For uniformly bounded $G(t)$, *we have*

$$\limsup_{t\to\infty} \int_0^t G^2(s)e^{(2\int_s^t A(u)du - \int_s^t z^2(u,\theta(u))Pdu)}ds < \infty. \tag{6.21}$$

iii) If $\frac{P}{2} > [A]^+$ *we have the following*

$$\lim_{t\to\infty} \int_0^t G^2 e^{(2\int_s^t Adu - \int_s^t Pdu)}ds = -\frac{G^2}{2A-P}. \tag{6.22}$$

Proof: i) Under the assumptions of the Lemma,

$$2A(t) - z^2(t,\theta(t))P \leq -\epsilon \quad a.e. - t \geq 0,\ a.s. \tag{6.23}$$

$$\begin{aligned}
2\int_0^t A(s)ds - \int_0^t z^2(s,\theta(s))Pds &\leq -\epsilon t,\ a.s. \\
e^{(2\int_0^t A(s)ds - \int_0^t z^2(s,\theta(s))Pds)} &\leq e^{-\epsilon t},\ a.s. \\
\limsup_{t\to\infty} e^{(2\int_0^t A(s)ds - \int_0^t z^2(s,\theta(s))Pds)} &\leq \limsup_{t\to\infty} e^{-\epsilon t} = \lim_{t\to\infty} e^{-\epsilon t} = 0,\ a.s.
\end{aligned} \tag{6.24}$$

On the other hand,

$$\begin{aligned}
e^{(2\int_0^t A(s)ds - \int_0^t z^2(s,\theta(s))P)ds} &\geq 0 \\
\liminf_{t\to\infty} e^{(2\int_0^t A(s)ds - \int_0^t z^2(s,\theta(s))P)ds} &\geq 0.
\end{aligned} \tag{6.25}$$

This completes the proof of (6.20).
ii) From (6.23), it also follows that, *a.s.*,

$$\int_0^t G^2(s)e^{(2\int_s^t A(u)du - \int_s^t z^2(u,\theta(u))Pdu)}ds \leq \int_0^t G^2(s)e^{-\epsilon(t-s)}ds. \tag{6.26}$$

Consequently, as $G(t)$ is uniformly bounded (i.e., there exist $G = \sup_{t\in[0,\infty)}|G(t)| < \infty$ such that $G(t) \leq G,\ \forall t \geq 0$)

$$\begin{aligned}
\limsup_{t\to\infty} \int_0^t G^2(s)e^{-\epsilon(t-s)}ds &\leq \liminf_{t\to\infty} \frac{G^2}{\epsilon}(1 - \exp(-\epsilon t)) \\
&= \lim_{t\to\infty} \frac{G^2}{\epsilon}(1 - e^{-\epsilon t}) = \frac{G^2}{\epsilon} < \infty,
\end{aligned} \tag{6.27}$$

and hence (6.21) follows.

iii) It is easily obtained by direct calculation.

Proof of Theorem 3.4.8.

(Sufficient part) i) Follows from Lemma 6.1.2 and (3.29).

ii) If condition (3.31) holds there exists a set of $\{\alpha_i\}_{i=1}^q$ such that $0 \leq \alpha_i \leq 1$ and $\sum_{i=1}^q \alpha_i = 1$ in which

$$\frac{\alpha_i P}{2} > [Re(\lambda_i(A))]^+. \tag{6.28}$$

Subsequently, from Lemma 6.1.2 and (3.29) the result is obtained.

(Necessary part) ii) Consider the case of $G \neq 0$. If condition (3.32) is not satisfied, then for each set $\{\alpha_i\}_{i=1}^q$ such that $0 \leq \alpha_i \leq 1$, and $\sum_{i=1}^q \alpha_i = 1$, there exists one element $\alpha_j \in \{\alpha_i\}_{i=1}^q$ such that $\frac{\alpha_j P}{2} < [Re(\lambda_j(A))]^+$. This implies that $V_{jj}^*(t, y, \theta) \to \infty$ as $t \to \infty$. Subsequently, there is no other encoding scheme with asymptotic bounded mean square estimation error. The result for the case of $G = 0$ follows similarly.

Proof of Proposition 3.5.1.

i) Under the assumption that the rank condition holds, from Lemma 6.2.47 follows that $\bar{P}(t) = \bar{P} \geq 0$ exists, which is the solution of the following Algebraic Riccati equation

$$0 = Q - \bar{P}SBR^{-1}B'S'\bar{P} + 2\Lambda\bar{P}. \tag{6.29}$$

Therefore, since $V(t, y, \theta)$ tends to $\bar{V}$, as $t \to \infty$, the optimal cost functional (3.37) is equal to the right side of (3.38) which is bounded. Consequently, from (3.38) it follows that for a fixed sample path θ, $\lim_{T\to\infty} \frac{1}{T} \int_0^T E||\gamma(t)||_Q^2 dt < \infty$.

ii) Again under the assumption that the rank condition holds, $\bar{P}(t) = \bar{P} \geq 0$ exists. Moreover, as $V(t, y, \theta) \to \bar{V} = 0$, for a fixed sample path of the channel, the right side of (3.38) is zero. Consequently, from (3.38), it follows that $\bar{J} = 0$. This implies that for a fixed sample path of the channel $\lim_{T\to\infty} \frac{1}{T} \int_0^T E||\gamma(t)||_Q^2 dt = 0$.

6.1.3 Chapter 4

Proof of the Lemma 4.3.2.

From Remark 6.2.12, vi follows that $H_S(f_Y)$ is concave so $-H_S(f_Y)$ is a convex function of $f_Y \in \mathcal{D}$ in which $\mathcal{D}$ is a convex subset of vector space $L_1(\Re^d, \Re^+)$ with the ground field $\Re^+$, where $L_1(\Re^d, \Re^+)$ denotes the set of all integrable functions which are defined on the space $\Re^d$ and taking values in $\Re^+$ (the set of non-negative scalars). Let $\Pi(dY) = g_Y dY$ be the probability measure associated with the nominal density $g_Y \in \mathcal{D}$ and $P(dY) = f_Y dY$ be a positive measure associated with $f_Y \in L_1(\Re^d, \Re^+)$. Then, from Remark 6.2.28, iv follows that $H(P||\Pi) = H(f_Y||g_Y)$ is a real valued convex function of $f_Y \in L_1(\Re^d, \Re^+)$. Moreover, as $E_{P(dY)}[L(Y)] = E_{f_Y}[L(Y)]$ is a linear function of $f_Y \in L_1(\Re^d, \Re^+)$, $G(f_Y) \triangleq H(f_Y||g_Y) - R_c - E_{f_Y}[L(Y)]$ is a convex mapping from vector space $L_1(\Re^d, \Re^+)$ to the normed space $\Re$. Furthermore, for $f_Y = g_Y$, $G(f_Y) < 0$. Subsequently, the conditions for applying the Lagrange Duality Theorem (Lemma 6.2.33, with $f = -H_S(f_Y)$, $G = G(f_Y)$, $\Omega = \mathcal{D}$, $X = L_1(\Re^d, \Re^+)$ and $Z = \Re$) are satisfied and the constraint problem (4.19) is equivalent to the following unconstraint problem.

$$\begin{aligned} \sup_{f_Y \in \mathcal{D}_{SU}(g_Y)} H_S(f_Y) &= - \inf_{f_Y \in \mathcal{D}_{SU}(g_Y)} -H_S(f_Y) \\ &= - \max_{s \geq 0} \inf_{f_Y \in \mathcal{D}} \Big(- H_S(f_Y) + s(H(f_Y||g_Y) - R_c - E_{f_Y}[L(Y)]) \Big) \\ &= \min_{s \geq 0} \sup_{f_Y \in \mathcal{D}} L(s, f_Y) \\ &= L(s^*, f_Y^{*,s^*}). \end{aligned} \tag{6.30}$$

Proof of Theorem 4.3.3.

i)

$$\begin{aligned} L(s, f_Y) &= H_S(f_Y) - sH(f_Y||g_Y) + sR_c + sE_{f_Y}[L(Y)] \\ &= -\int f_Y(y) \log_e f_Y(y) dy - sH(f_Y||g_Y) + sR_c + sE_{f_Y}[L(Y)] \\ &= -\int log_e(\frac{f_Y(y)}{g_Y(y)} . g_Y(y)) f_Y(y) dy - sH(f_Y||g_Y) + sR_c + sE_{f_Y}[L(Y)] \\ &= -H(f_Y||g_Y) - E_{f_Y}[\log_e g_Y] - sH(f_Y||g_Y) + sR_c + sE_{f_Y}[L(Y)] \\ &= sR_c + (1+s)\Big(E_{f_Y}[\frac{s}{1+s} L(Y) - \frac{1}{1+s} \log_e g_Y] - H(f_Y||g_Y)\Big). \end{aligned} \tag{6.31}$$

Subsequently, from Remark 6.2.28, i, it follows that

$$L(s, f_Y^{*,s}) = \sup_{f_Y \in \mathcal{D}} L(s, f_Y)$$

$$= sR_c + (1+s)\log_e \int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy \tag{6.32}$$

where

$$f_Y^{*,s} = \frac{(e^{L(y)} g_Y(y))^{\frac{s}{1+s}}}{\int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy} \tag{6.33}$$

provided $(\frac{s}{1+s}L(y) - \frac{1}{1+s}\log_e g_Y(y))(e^{L(y)} g_Y(y))^{\frac{s}{1+s}} \in L_1(\Re^d, \Re)$ for some $s \geq 0$.

ii) In the following, we will show that $\frac{d^2}{ds^2}L(s, f_Y^{*,s}) \geq 0$, $\forall s \geq 0$. This means that $L(s, f_Y^{*,s})$ is a convex function of $s > 0$.

$$\begin{aligned}
\frac{d}{ds}L(s, f_Y^{*,s}) &= \frac{d}{ds}\Big[sR_c + (1+s)\log_e \int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy\Big] \\
&= R_c + \log_e \int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy \\
&\quad + \frac{1}{1+s}\frac{\int (L(y) + \log_e g_Y(y))(e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy}{\int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy}
\end{aligned} \tag{6.34}$$

Subsequently,

$$\begin{aligned}
\frac{d^2 L(s, f_Y^{*,s})}{ds^2} &= \frac{1}{(1+s)^3}\Big\{\int (L(y) + \log_e g_Y(y))^2 f_Y^{*,s}(y) dy \\
&\quad -\Big(\int (L(y) + \log_e g_Y(y)) f_Y^{*,s}(y) dy\Big)^2\Big\} \\
&= \frac{1}{(1+s)^3} Var_{f_Y^{*,s}}[L(Y) + \log_e g_Y(y)] \geq 0, \quad \forall s \geq 0
\end{aligned} \tag{6.35}$$

where for $f(x)$ positive function, $Var_{f(x)}[X] \triangleq \int (x - \int x f(x)dx)^2 f(x) dx$. In the above analysis, under the assumptions of part ii (see Theorem 4.3.3), we can exchange the differentiation with the integral since for any real valued function $k(s, y) \in L_1(\Re \times \Re^d, \Re)$ in which it is differentiable with respect to s, we have

$$\frac{d}{ds}\int k(s,y)dy \triangleq \lim_{s_n \to s} \int \frac{k(s_n, y) - k(s, y)}{s_n - s} dy. \tag{6.36}$$

On the other hand, by the mean value theorem, there exists a c_n between s_n and s such that

$$\frac{k(s_n, y) - k(s, y)}{s_n - s} = \frac{d}{ds}k(s, y)\Big|_{s_n = c_n} \triangleq \dot{k}(c_n, y) \tag{6.37}$$

and $\lim_{n\to\infty} c_n = s$. Subsequently, if $\dot{k}(s, y) \in L_1(\Re \times \Re^d, \Re)$, from Lebesgue's dominated convergence theorem (Lemma 6.2.34) follows that

$$\begin{aligned}
\frac{d}{ds}\int k(s,y)dy &\triangleq \lim_{s_n \to s} \int \frac{k(s_n, y) - k(s, y)}{s_n - s} dy \\
&= \int \lim_{s_n \to s} \frac{k(s_n, y) - k(s, y)}{s_n - s} dy \\
&= \int \dot{k}(s, y) dy.
\end{aligned} \tag{6.38}$$

iii) Since $L(s, f_Y^{*,s})$ is a convex function of $s \geq 0$, there exists $s^* \geq 0$ such that $L(s, f_Y^{*,s})$ attains its minimum at this point. This minimizing point, $s^* \geq 0$, is the solution of the following equation

$$\frac{d}{ds}L(s, f_Y^{*,s})\Big|_{s=s^*} = 0$$

$$\Leftrightarrow$$

$$R_c + \log_e \int (e^{L(y)} g_Y(y))^{\frac{s^*}{1+s^*}} dy + \frac{1}{1+s^*} \int (L(y) + \log_e g_Y(y)) f_Y^{*,s^*}(y) = 0 \tag{6.39}$$

On the other hand,

$$\begin{aligned} H(f_Y^{*,s} \| g_Y) &= \int L(y) f_Y^{*,s}(y) dy - \frac{1}{1+s} \int L(y) f_Y^{*,s}(y) dy - \frac{1}{1+s} \int \log_e(g_Y(y)) f_Y^{*,s}(y) dy \\ &\quad - \log_e \int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy \\ &= E_{f_Y^{*,s}}[L(Y)] - \frac{1}{1+s} \int (L(y) + \log_e g_Y(y)) f_Y^{*,s}(y) dy \\ &\quad - \log_e \int (e^{L(y)} g_Y(y))^{\frac{s}{1+s}} dy. \end{aligned} \tag{6.40}$$

Subsequently, $s^* \geq 0$ is the solution of the following equation

$$\begin{aligned} &R_c + E_{f_Y^{*,s^*}}[L(Y)] - H(f_Y^{*,s^*} \| g_Y) = 0 \\ &T(s^*) \triangleq H(f_Y^{*,s^*} \| g_Y) - E_{f_Y^{*,s^*}}[L(Y)] = R_c. \end{aligned} \tag{6.41}$$

Moreover, from (6.41), (6.40) and (6.39), it follows that

$$\begin{aligned} \frac{d}{ds}L(s, f_Y^{*,s}) &= R_c - T(s) \\ T(s) &= R_c - \frac{d}{ds}L(s, f_Y^{*,s}) \\ \frac{dT(s)}{ds} &= -\frac{d^2}{ds^2}L(s, f_Y^{*,s}) \end{aligned} \tag{6.42}$$

and since $\frac{d^2}{ds^2}L(s, f_Y^{*,s}) \geq 0$, $\forall s \geq 0$ (part ii),

$$\frac{dT(s)}{ds} \leq 0, \ \forall s \geq 0. \tag{6.43}$$

Thus, $T(s)$ is a non-increasing function of $s \geq 0$.

Proof of Remark 4.3.6.

(4.32) and (4.33) are the direct result of (4.30) and (4.31). Thus, we just need to show that for a given $R_c \in [0,\infty)$, there exists a unique solution (i.e., unique $s^* > 0$) for equations (4.33).

(4.33) can be written into the following form

$$e^{-2R_c} = e^{-\frac{d}{s^*}}(\frac{1+s^*}{s^*})^d \triangleq M(s^*). \tag{6.44}$$

Since $\frac{dM(s^*)}{ds^*} > 0$, $M(s^*)$ is a strictly increasing function of $s^* > 0$. Further, since $M(s^*)$ is a continuous function of $s^* > 0$, the minimum and the maximum of $M(s^*)$ are obtained at $s^* \to 0$ and $s^* \to \infty$, respectively. That is, the minimum is $\lim_{s^*\to 0} M(s^*) = 0$ and the maximum is $\lim_{s^*\to\infty} M(s^*) = 1$. Subsequently, $M(s^*) \in (0,1)$. On the other hand, $\forall R_c \geq 0$, $e^{-2R_c} \in (0,1]$. Consequently, since $M(s^*)$ is a continuous and strictly increasing function of $s^* > 0$ which covers the range of e^{-2R_c}, for each $R_c \geq 0$, there exists a unique $s^* > 0$ such that

$$e^{-2R_c} = M(s^*), \tag{6.45}$$

or equivalently, there exists a unique $s^* > 0$ such that

$$R_c = -\frac{d}{2}\log_e(\frac{1+s^*}{s^*}) + \frac{d}{2s^*}, \quad R_c \in [0,\infty). \tag{6.46}$$

Proof of Theorem 4.3.7.

From (4.42), we have

$$\begin{aligned}\frac{1}{T}H_r(f^{*,s^*}_{Y^{T-1}}) &= \min_{s\geq 0}\sup_{\{\bar{W}_t\}_{t=0}^{T-2}}\Big\{-\frac{1+s}{2T}E_P[\sum_{t=0}^{T-2}\bar{W}_t'\Sigma_W^{-1}\bar{W}_t] \\ &\quad -\frac{1}{T}E_P[\log_e g_{Y^{T-1}}] + sR_c + \frac{s}{2T}E_P[\sum_{t=0}^{T-1}Y_t'MY_t]\Big\}.\end{aligned} \tag{6.47}$$

On the other hand,

$$\begin{aligned}\log_e[g_{Y^{T-1}}] &= -\frac{Tq}{2}\log_e(2\pi) - \frac{1}{2}\log_e\det\bar{V}_0 - \frac{T-1}{2}\log_e\det(B\Sigma_W B') - \frac{1}{2}Y_0'\bar{V}_0^{-1}Y_0 \\ &\quad -\frac{1}{2}\sum_{t=0}^{T-2}(Y_{t+1}-AY_t)'(B\Sigma_W B')^{-1}(Y_{t+1}-AY_t)\end{aligned} \tag{6.48}$$

Subsequently, (6.47) is equal to the following.

$$\frac{1}{T}H_r(f^{*,s^*}_{Y^{T-1}}) = \min_{s\geq 0}\sup_{\{\bar{W}_t\}_{t=0}^{T-2}}\Big\{-\frac{1+s}{2T}E_P[\sum_{t=0}^{T-2}\bar{W}_t'\Sigma_W^{-1}\bar{W}_t] - \frac{1}{T}\Big\{-\frac{Tq}{2}\log_e(2\pi)$$

$$-\frac{1}{2}\log_e \det \bar{V}_0 - \frac{T-1}{2}\log_e \det(B\Sigma_W B') - \frac{1}{2}E_P[Y_0\bar{V}_0^{-1}Y_0]$$
$$-\frac{1}{2}E_P[\sum_{t=0}^{T-2}(Y_{t+1}-AY_t)'(B\Sigma_W B')^{-1}(Y_{t+1}-AY_t)] + sR_c$$
$$+\frac{s}{2T}E_P[\sum_{t=0}^{T-1}Y_t'MY_t]\}\}. \tag{6.49}$$

Next, since

$$\begin{aligned}\frac{1}{2}E_P[Y_0'\bar{V}_0^{-1}Y_0] &= \frac{1}{2}E_P[trac(Y_0'\bar{V}_0^{-1}Y_0)]\\ &= \frac{1}{2}E_P[trac(\bar{V}_0^{-1}Y_0Y_0')]\\ &= \frac{1}{2}trac(\bar{V}_0^{-1}E_{P(dY^{T-1})}[Y_0Y_0'])\\ &= \frac{1}{2}trac(\bar{V}_0^{-1}\bar{V}_0)\\ &= \frac{q}{2}\end{aligned} \tag{6.50}$$

and

$$E_P[\sum_{t=0}^{T-2}(Y_{t+1}-AY_t)'(B\Sigma_W B')^{-1}(Y_{t+1}-AY_t)]$$
$$= E_P[\sum_{t=0}^{T-2}(W_t+\bar{W}_t)'B'(B\Sigma_W B')^{-1}B(W_t+\bar{W}_t)]$$
$$= E_P[\sum_{t=0}^{T-2}W_t'B'(B\Sigma_W B')^{-1}BW_t] + E_P[\sum_{t=0}^{T-2}\bar{W}_t'B'(B\Sigma_W B')^{-1}B\bar{W}_t]$$
$$= \sum_{t=0}^{T-2}trac(B'(B\Sigma_W B')^{-1}BE_P[W_tW_t']) + E_P[\sum_{t=0}^{T-2}\bar{W}_t'B'(B\Sigma_W B')^{-1}B\bar{W}_t]$$
$$= \sum_{t=0}^{T-2}trac((B\Sigma_W B')^{-1}B\Sigma_W B') + E_P[\sum_{t=0}^{T-2}\bar{W}_t'B'(B\Sigma_W B')^{-1}B\bar{W}_t]$$
$$= (T-1)q + E_P[\sum_{t=0}^{T-2}\bar{W}_t'B'(B\Sigma_W B')^{-1}B\bar{W}_t] \tag{6.51}$$

Then, (6.49) is equal to the following

$$\begin{aligned}\frac{1}{T}H_r(f^{*,s^*}_{Y^{T-1}}) &= \min_{s\geq 0}\sup_{\{\bar{W}_t\}_{t=0}^{T-2}}\Big\{sR_c + \frac{q}{2}\log_e(2\pi) + \frac{1}{2T}\log_e\det\bar{V}_0 + \frac{T-1}{2T}\log_e\det(B\Sigma_W B')\\ &+\frac{q}{2T} + \frac{T-1}{2T}q + \frac{1}{2T}E_P\Big[\sum_{t=0}^{T-2}[\bar{W}_t'(B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1})\bar{W}_t\\ &+Y_t'sMY_t] + \frac{1}{2T}Y_{T-1}'sMY_{T-1}\Big]\Big\}\end{aligned} \tag{6.52}$$

and subsequently,

$$\begin{aligned}\frac{1}{T}H_r(f^{*,s^*}_{Y^{T-1}}) &= \frac{q}{2}\log_e(2\pi) + \frac{T-1}{2T}q + \frac{1}{2T}\log_e\det\bar{V}_0 + \frac{T-1}{2T}\log_e\det(B\Sigma_W B') + \frac{q}{2T}\\ &+ \min_{s\geq 0}\Big\{sR_c + \frac{1}{2T}\sup_{\{\bar{W}_t\}_{t=0}^{T-2}} E_P[\sum_{t=0}^{T-2}[\bar{W}_t'(B'(B\Sigma_W B')^{-1}B\\ &-(1+s)\Sigma_W^{-1})\bar{W}_t + Y_t' sMY_t] + \frac{1}{2T}Y_{T-1}'sMY_{T-1}]\Big\}.\end{aligned} \quad (6.53)$$

Under the assumption that $B'(B\Sigma_W B')^{-1}B < (1+s)\Sigma_W^{-1}$ for some $s \geq 0$,

$$\sup_{\{\bar{W}_t\}_{t=0}^{T-2}}\Big\{E_P[\sum_{t=0}^{T-2}[\bar{W}_t'(B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1})\bar{W}_t + Y_t' sMY_t] + Y_{T-1}'sMY_{T-1}]\Big\} \quad (6.54)$$

has a unique solution which can be obtained by invoking the stochastic dynamic programming [23] as follows.

Define the value function

$$\begin{aligned}J(k,Y) &\triangleq \sup_{\{\bar{W}_t\}_{t=k}^{T-2}}\Big\{E_P[\sum_{t=k}^{T-2}[\bar{W}_t'(B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1})\bar{W}_t + Y_t' sMY_t]\\ &+Y_{T-1}'sMY_{T-1}\Big|Y_k]\Big\}\end{aligned} \quad (6.55)$$

Then, this value function satisfies the following stochastic dynamic programming [23]

$$J(k,Y) = \sup_{\bar{W}_k} E_P[\bar{W}_k'(B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1})\bar{W}_k + Y_k'sMY_k + J(k+1,Y)\Big|Y_k] \quad (6.56)$$

where

$$\begin{aligned}&\sup_{\{\bar{W}_t\}_{t=0}^{T-2}}\Big\{E_P[\sum_{t=0}^{T-2}[\bar{W}_t'(B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1})\bar{W}_t + Y_t' sMY_t] + Y_{T-1}'sMY_{T-1}\Big]\Big\}\\ &= E_P J(0,Y).\end{aligned} \quad (6.57)$$

Next, we pick up the following candidate as the solution of the dynamic programming (6.56).

$$J(k,Y) = Y_k'\Xi_k Y_k + \Theta_k, \qquad \Xi_k \text{ is symmetric.} \quad (6.58)$$

Subsequently,

$$\begin{aligned}Y_k'\Xi_k Y_k + \Theta_k &= \sup_{\bar{W}_k} E_P[\bar{W}_k'(B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1})\bar{W}_k\\ &+Y_k'sMY_k + Y_{k+1}'\Xi_{k+1}Y_{k+1} + \Theta_{k+1}\Big|Y_k].\end{aligned} \quad (6.59)$$

On the other hand

$$\begin{aligned} E_P[Y_{k+1}'\Xi_{k+1}Y_{k+1}\Big|Y_k] &= E_P[(AY_k + BW_k + B\bar{W}_k)'\Xi_{k+1}(AY_k + BW_k \\ &\quad +B\bar{W}_k)\Big|Y_k] \\ &= E_P[Y_k'A'\Xi_{k+1}AY_k\Big|Y_k] \\ &\quad +E_P[W_k'B'\Xi_{k+1}BW_k\Big|Y_k] \\ &\quad +E_P[\bar{W}_k'B'\Xi_{k+1}B\bar{W}_k|Y_k] \\ &\quad +2E_P[\bar{W}_k'B'\Xi_{k+1}AY_k|Y_k] \end{aligned} \tag{6.60}$$

where

$$\begin{aligned} E_P[W_k'B'\Xi_{k+1}BW_k\Big|Y_k] &= trac(B'\Xi_{k+1}BE_P[W_kW_k']) \\ &= trac(B'\Xi_{k+1}B\Sigma_W). \end{aligned} \tag{6.61}$$

Subsequently,

$$\begin{aligned} Y_k'\Xi_kY_k + \Theta_k &= \sup_{\bar{W}_k} E_P[\bar{W}_k'(B'(B\Sigma_WB')^{-1}B - (1+s)\Sigma_{\bar{W}}^{-1} + B'\Xi_{k+1}B)\bar{W}_k \\ &\quad +Y_k'sMY_k + trac(B'\Xi_{k+1}B\Sigma_W) + 2\bar{W}_k'B'\Xi_{k+1}AY_k + \Theta_{k+1}\Big|Y_k]. \end{aligned} \tag{6.62}$$

where $\bar{W}_k^*$ that suprimizes the right side of (6.62) is given by

$$\bar{W}_k^* = -[B'(B\Sigma_WB')^{-1}B - (1+s)\Sigma_{\bar{W}}^{-1} + B'\Xi_{k+1}B]^{-1}B'\Xi_{k+1}AY_k \tag{6.63}$$

where

$$\begin{aligned} \Xi_k &= A'\Xi_{k+1}A - A'\Xi_{k+1}B[B'(B\Sigma_WB')^{-1}B - (1+s)\Sigma_{\bar{W}}^{-1} + B'\Xi_{k+1}B]^{-1}B'\Xi_{k+1}A + sM \\ \Theta_k &= \Theta_{k+1} + trac(B'\Xi_{k+1}B\Sigma_W) \end{aligned} \tag{6.64}$$

in which Ξ_{T-1} and Θ_{T-1} are defined from the following equations.

$$Y_{T-1}'\Xi_{T-1}Y_{T-1} + \Theta_{T-1} = Y_{T-1}'sMY_{T-1}. \tag{6.65}$$

That is,

$$\begin{aligned} \Theta_{T-1} &= 0 \\ \Xi_{T-1} &= sM. \end{aligned} \tag{6.66}$$

Subsequently,

$$\begin{aligned} E_P J(0,Y) &= E_P[Y_0'\Xi_0 Y_0] + \sum_{t=1}^{T-1} trac(B'\Xi_t B\Sigma_W) \\ &= trac(\Xi_0 \bar{V}_0) + \sum_{t=1}^{T-1} trac(B'\Xi_t B\Sigma_W) \end{aligned} \quad (6.67)$$

Thus,

$$\begin{aligned} \frac{1}{T}H_r(f^{*,s^*}_{Y^{T-1}}) &= \frac{q}{2}\log_e(2\pi) + \frac{T-1}{2T}q + \frac{1}{2T}\log_e \det \bar{V}_0 + \frac{T-1}{2T}\log_e \det(B\Sigma_W B') + \frac{q}{2T} \\ &+ \min_{s\geq 0}\{sR_c + \frac{1}{2T}trac(\Xi_0\bar{V}_0) + \frac{1}{2T}\sum_{t=1}^{T-1} trac(B'\Xi_t B\Sigma_W)\} \end{aligned} \quad (6.68)$$

where Ξ_t is obtained from the following Riccati equation which appears in the H^∞ control problem

$$\begin{aligned} \Xi_t &= A'\Xi_{t+1}A - A'\Xi_{t+1}B[B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1} + B'\Xi_{t+1}B]^{-1}B'\Xi_{t+1}A + sM \\ \Xi_{T-1} &= sM. \end{aligned} \quad (6.69)$$

Next, under the assumption that (A,B) is controllable, A and $B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1}$ are invertible, and $\beta(\eta) > 0$ (see (4.47)) for some η; $|\eta| = 1$, from Lemma 6.2.48 follows that the Riccati equation (6.69) has a real symmetric steady state solution Ξ_∞ which is the solution of the Algebraic Riccati equation associated with the Riccati equation (6.69). Subsequently,

$$\lim_{T\to\infty}\frac{1}{2T}\sum_{t=1}^{T-1} trac(B'\Xi_t B\Sigma_W) = \frac{1}{2}trac(B'\Xi_\infty B\Sigma_W) \quad (6.70)$$

Thus,

$$\mathcal{H}_r(\mathcal{Y}) = \frac{q}{2}\log_e(2\pi e) + \frac{1}{2}\log_e \det(B\Sigma_W B') + \min_{s\geq 0}\{sR_c + \frac{1}{2}trac(B'\Xi_\infty B\Sigma_W)\} \quad (6.71)$$

where Ξ_∞ is the solution of the following equation

$$\Xi_\infty = A'\Xi_\infty A - A'\Xi_\infty B[B'(B\Sigma_W B')^{-1}B - (1+s)\Sigma_W^{-1} + B'\Xi_\infty B]^{-1}B'\Xi_\infty A + sM. \quad (6.72)$$

Proof of Proposition 4.3.9.

The entropy for this class of systems is computed from Theorem 4.3.3 when R_c is replaced by TR_c and g_Y by $g_{Y^{T-1}}$. In particular,

$$\int_{\Re^{Tq}} (g_{Y^{T-1}}(y^{T-1}))^{\frac{s}{1+s}} dy^{T-1} = (2\pi)^{\frac{Tq}{2}\frac{1}{1+s}}\Big(\det(B\Sigma_W B')\Big)^{\frac{T-1}{2}\frac{1}{1+s}}(\frac{1+s}{s})^{\frac{Tq}{2}}(\det \bar{V}_0)^{\frac{1}{2}\frac{1}{1+s}}. \quad (6.73)$$

Furthermore, the maximizing density is given by (4.25) with g_Y replaced by $g_{Y^{T-1}}$. Subsequently, from Theorem 4.3.3, iii follows that the minimizing $s^* > 0$ is the solution of $H(f^{*,s}_{Y^{T-1}}||g_{Y^{T-1}})\Big|_{s=s^*} = -\frac{Tq}{2}\log_e(\frac{1+s^*}{s^*}) + \frac{Tq}{2s^*} = TR_c$. Consequently, the robust entropy is computed from (4.23) with minimizing $s \geq 0$ replaced by $s^* > 0$ which is the solution of the following nonlinear equation.

$$-\frac{q}{2}\log(\frac{1+s^*}{s^*}) + \frac{q}{2s^*} = R_c. \tag{6.74}$$

Proof of Proposition 4.3.11.

Since $||\Delta||_\infty \leq 1$, it follows that $|H(e^{jw}) + \Delta(e^{jw})W(e^{jw})| \leq |H(e^{jw})| + |W(e^{jw})|$. Consequently,

$$\begin{aligned} \log_e S_Y(e^{jw}) &\leq \log_e\Big((|H(e^{jw})| + |W(e^{jw})|)^2 S_X(e^{jw})\Big), \ \ \forall S_Y \in \mathcal{P}_{SU} \\ \int_{-\pi}^{\pi} \log_e S_Y(e^{jw})dw &\leq \int_{-\pi}^{\pi} \log_e\Big((|H(e^{jw})| + |W(e^{jw})|)^2 S_X(e^{jw})\Big)dw, \ \ \forall S_Y \in \mathcal{P}_{SU} \\ \sup_{S_Y \in \mathcal{P}_{SU}} \int_{-\pi}^{\pi} \log_e S_Y(e^{jw})dw &\leq \int_{-\pi}^{\pi} \log_e\Big((|H(e^{jw})| + |W(e^{jw})|)^2 S_X(e^{jw})\Big)dw. \end{aligned} \tag{6.75}$$

On the other hand, it is easily shown that there exists a $\Delta^*(e^{jw})$, with $|\Delta^*(e^{jw})| = 1$, and $\arg(\Delta^*(e^{jw})) = \arg(H(e^{jw})) - \arg(W(e^{jw}))$ so that $|H(e^{jw}) + \Delta^*(e^{jw})W(e^{jw})|^2 = \Big(|H(e^{jw})| + |W(e^{jw})|\Big)^2$. Consequently, for $\Delta(e^{jw}) = \Delta^*(e^{jw})$,

$$\int_{-\pi}^{\pi} \log_e\Big(|H(e^{jw})| + |W(e^{jw})|\Big)^2 S_X(e^{jw})dw = \int_{-\pi}^{\pi} \log_e S^*_Y(e^{jw})dw, \tag{6.76}$$

where $S^*_Y(e^{jw}) = |H(e^{jw}) + \Delta^*(e^{jw})W(e^{jw})|^2 S_X(e^{jw}) \in \mathcal{P}_{SU}$. It is evident that $\int_{-\pi}^{\pi}\log_e S^*_Y(e^{jw})dw \leq \sup_{S_Y\in\mathcal{P}_{SU}} \int_{-\pi}^{\pi} \log_e S_Y(e^{jw})dw$. Subsequently, from (6.76), it follows that

$$\int_{-\pi}^{\pi} \log_e\Big((|H(e^{jw}) + W(e^{jw})|)^2 S_X(e^{jw})\Big)dw \leq \sup_{S_Y \in \mathcal{P}_{SU}} \int_{-\pi}^{\pi} \log_e S_Y(e^{jw})dw. \tag{6.77}$$

Thus, from (6.75) and (6.77), the result is obtained.

Proof of Theorem 4.4.1.

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 (Remark 6.2.14, ii) follows that

$$\begin{aligned} &I(Z^{n-1}; \tilde{Z}^{n-1}) \geq I(Y^{T-1}; \tilde{Y}^{T-1}), \ \ \forall P(dY^{T-1}) \in \mathcal{M}^T_{SU} \subset \mathcal{M}_1(\Re^{Td}), \\ &\sup_{\{P(dZ^{n-1}); P(dY^{T-1}) \in \mathcal{M}^T_{SU}\}} I(Z^{n-1}; \tilde{Z}^{n-1}) \geq \sup_{P(dY^{T-1}) \in \mathcal{M}^T_{SU}} I(Y^{T-1}; \tilde{Y}^{T-1}), \end{aligned}$$

$$
\begin{aligned}
\sup_{\{P(dZ^{n-1})\in\mathcal{M}_{CU}\}} I(Z^{n-1};\tilde{Z}^{n-1}) &\geq \sup_{\{P(dZ^{n-1});P(dY^{T-1})\in\mathcal{M}_{SU}^{T}\}} I(Z^{n-1};\tilde{Z}^{n-1}) \\
&\geq \sup_{P(dY^{T-1})\in\mathcal{M}_{SU}^{T}} I(Y^{T-1};\tilde{Y}^{T-1}), \\
\mathcal{C}_n \triangleq \sup_{\{P(dZ^{n-1})\in\mathcal{M}_{CU}\}} I(Z^{n-1};\tilde{Z}^{n-1}) &\geq \inf_{P(d\tilde{Y}^{T-1};y^{T-1})\in\mathcal{M}_{DC}} \sup_{P(dY^{T-1})\in\mathcal{M}_{SU}^{T}} I(Y^{T-1};\tilde{Y}^{T-1}) \\
&\triangleq R_{T,r}(D_v)
\end{aligned}
\tag{6.78}
$$

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.

That is, $\mathcal{C}_n \geq R_{T,r}(D_v)$ 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 channels with feedback, the directed information from channel inputs to outputs must be used instead of the mutual information in (6.78), in which this capacity is the same as the capacity without feedback, i.e., $\mathcal{C}_n$.

Proof of Lemma 4.4.2.

From Lemma 6.2.24, it follows that

$$
\begin{aligned}
\inf_{P(d\tilde{Y}^{T-1};y^{T-1})\in\mathcal{M}_{DC}} \frac{1}{T} I(Y^{T-1};\tilde{Y}^{T-1}) &\geq \frac{1}{T} H_S(f_{Y^{T-1}}) - \max_{h\in G_D} H_S(h), \\
\forall P(dY^{T-1}) &= f_{Y^{T-1}} dY^{T-1};\, f_{Y^{T-1}} \in \mathcal{D}_{SU}^{T}.
\end{aligned}
\tag{6.79}
$$

Further, when $\int_{\Re^d} e^{s\rho(\xi)} d\xi < \infty$, $\forall s<0$, the maximizer $h^*(\xi) \in G_D$ is given by (4.61). Subsequently,

$$
\begin{aligned}
\frac{1}{T} R_{T,r}(D_v) &\triangleq \inf_{P(d\tilde{Y}^{T-1};y^{T-1})\in\mathcal{M}_{DC}} \sup_{P(dY^{T-1})\in\mathcal{M}_{SU}^{T}} \frac{1}{T} I(Y^{T-1};\tilde{Y}^{T-1}) \\
&\geq \sup_{P(dY^{T-1})\in\mathcal{M}_{SU}^{T}} \inf_{P(d\tilde{Y}^{T-1};y^{T-1})\in\mathcal{M}_{DC}} \frac{1}{T} I(Y^{T-1};\tilde{Y}^{T-1}) \\
&\geq \sup_{f_{Y^{T-1}}\in\mathcal{D}_{SU}^{T}} \frac{1}{T} H_S(f_{Y^{T-1}}) - \max_{h\in G_D} H_S(h).
\end{aligned}
\tag{6.80}
$$

Proof of Theorem 4.4.3.

(Uniform Observability). Consider the case where the class of sources is described by a class of probability measures, that is, $P(dY^{T-1}) = f_{Y^{T-1}}dY^{T-1} \in \mathcal{M}_{SU}^T = \{P(dY^{T-1}) = f_{Y^{T-1}}dY^{T-1}; f_{Y^{T-1}} \in \mathcal{D}_{SU}^T\}$. Assume for a given control signal, there exist an encoder and decoder such that uniform observability in probability in the sense of Definition 4.2.2 is obtained. From (4.6), it follows that for a fixed $\delta \geq 0$ and $D_v \geq 0$, there exists $T(\delta, D_v) \in \mathbf{N}_+$ such that, $\forall T \geq T(\delta, D_v)$

$$\sup_{P(dY^{T-1}) \in \mathcal{M}_{SU}^T} \frac{1}{T} \sum_{t=0}^{T-1} \Pr(||Y_t - \tilde{Y}_t|| > \delta) \leq D_v. \tag{6.81}$$

This implies that

$$\frac{1}{T} \sum_{t=0}^{T-1} \Pr(||Y_t - \tilde{Y}_t|| > \delta) \leq D_v, \quad \forall P(dY^{T-1}) \in \mathcal{M}_{SU}^T. \tag{6.82}$$

Next, define 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 $\rho(.,.)$ is defined by $\rho(Y, \tilde{Y}) = \begin{cases} 1, & ||Y - \tilde{Y}|| > \delta \\ 0, & ||Y - \tilde{Y}|| \leq \delta \end{cases}$. Then, for $T \geq T(\delta, D_v)$

$$\begin{aligned} E\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) &= \frac{1}{T} \sum_{t=0}^{T-1} \Pr(||Y_t - \tilde{Y}_t|| > \delta) \leq D_v, \\ &\qquad \forall P(dY^{T-1}) \in \mathcal{M}_{SU}^T. \end{aligned} \tag{6.83}$$

That is, a robust rate distortion with distortion value D_v is obtained for $T \geq T(\delta, D_v)$. Then, by Theorem 4.4.1 and Lemma 4.4.2, the capacity and robust rate distortion must for all $T \geq T(\delta, D_v)$ satisfy

$$\begin{aligned} \lim_{T \to \infty} \frac{1}{T} \mathcal{C}_T &\geq \lim_{T \to \infty} \frac{1}{T} H_r(f^*_{Y^{T-1}}) - \max_{h \in G_D} H_S(h) \\ \mathcal{C} &\geq \mathcal{H}_r(\mathcal{Y}) - \max_{h \in G_D} H_S(h). \end{aligned} \tag{6.84}$$

Since among all distributions with the same covariance, the Gaussian distribution has the biggest entropy, $h^*(\xi) \in G_D$ that maximizes $H_S(h)$ is a Gaussian distributed which satisfies the boundary conditions of G_D. That is, $h^*(\xi) \sim N(0, \Gamma_g)$, in which Γ_g satisfies (4.64). Consequently, by substituting $\max_{h \in G_D} H_S(h) = H_S(h^*) = \frac{1}{2} \log_e[(2\pi e)^d \det \Gamma_g]$ in (6.84), the lower bound (4.63) is obtained.

Following the same procedure, a necessary condition for uniform observability in r-mean is given by (4.65) in which, from Remark 6.2.25, iii, $\max_{h\in G_D} H_S(h) = \frac{d}{r} - \log_e\left(\frac{r}{dV_d\Gamma(\frac{d}{r})}\left(\frac{d}{rD_v}\right)^{\frac{d}{r}}\right)$ nats per time step.

For the case when the class of sources is described by the class of PSD's, by considering the class of source probability measures induced by this class of PSD's and following the same procedure, it can be shown that (4.63) and (4.65) are also necessary conditions for uniform observability in probability and r-mean, respectively.

(Robust Stability). Follows similarly by considering the rate distortion between Y^{T-1} and Γ^{T-1}.

6.2 Concepts and Results From the Literature

Tools from probability theory, system and control, and information/communication theory are used extensively throughout the thesis to derive new results. Therefore, in this section we summarized known concepts, measures, and results from probability theory, information theory, and system and control that we used them throughout the thesis.

6.2.1 Concepts and Results From Probability Theory

In this section, we recall some of the well known concepts and results from probability theory.

Definition 6.2.1 *(Stochastic Kernel) Consider the Borel measurable spaces* $(\mathcal{A}, \mathcal{B}(\mathcal{A}))$ *and* $(\hat{\mathcal{A}}, \mathcal{B}(\hat{\mathcal{A}}))$ *where R.V.'s* X *and* Y *are taking values in* $\mathcal{A}$ *and* $\hat{\mathcal{A}}$*, respectively. A stochastic kernel denoted by* $P(dX; y)$ *is a mapping* $P : \mathcal{B}(\mathcal{A}) \times \hat{\mathcal{A}} \rightarrow [0, 1]$ *which satisfies*
i) For every $y \in \hat{\mathcal{A}}$*, the set function* $P(.; y)$ *is a probability measure on* $\mathcal{B}(\mathcal{A})$*;*
ii) For every $x \in \mathcal{B}(\mathcal{A})$*, the function* $P(dX; .)$ *is* $\mathcal{B}(\hat{\mathcal{A}})$ *measurable.*

Example 6.2.2 *(Gaussian R.V.'s) A R.V.,* $Y : (\Omega, \mathcal{F}) \rightarrow (\Re^d, \mathcal{B}(\Re^d))$ *with* $K = Cov(Y) > 0$ *is called Gaussian and it is denoted by* $Y \sim N(m, K)$ *if it has density function* $f_Y(y)$ *given by*

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^d \det(K)}} e^{-\frac{(Y-m)' K^{-1}(Y-m)}{2}} \tag{6.85}$$

where e is the base of natural logarithm and $m = E[Y]$. *For the case of* $K = Cov(Y) = 0$, *there is an alternative description for the Gaussian R.V. using the characteristic function [64].*

A sequence of R.V.'s $Y^T = (Y_0, Y_1, ..., Y_T)$ *is called jointly Gaussian if*

$$f_{Y^T}(y^T) = \frac{1}{\sqrt{(2\pi)^{Td}\det(K)}} e^{-\frac{(Y-m)' K^{-1}(Y-m)}{2}} \tag{6.86}$$

where $m = E[(Y_0', Y_1', ..., Y_T')']$ *and* $K = Cov((Y_0' \; Y_1' \; ... \; Y_T')')$.

Definition 6.2.3 *(Random Process) Let* $(\Omega, \mathcal{F}, P)$ *be a probability space,* $[0, T] \subset \Re$ *an index set and* $(\Re^d, \mathcal{B}(\Re^d))$ *a Borel measurable space.*

A stochastic process or Random Process (R.P.) on $(\Omega, \mathcal{F}, P)$ *is a function*

$$Y : [0, T] \times \Omega \rightarrow \Re^d \tag{6.87}$$

such that $\forall t \in [0, T]$

$$Y(t, .) : \Omega \rightarrow \Re^d \tag{6.88}$$

is a R.V.

For convenience we denote continues time R.P.'s by $y(t)$ *and discrete time R.P's by* Y_t.

In the study of R.P. $\{Y_t; t \in [0, T]\}$ *(resp.* $\{y(t); t \in [0, T]\}$*) defined on the probability space* $(\Omega, \mathcal{F}, P)$, *it is important to keep track of past, present and future information, through the time dependent component of the R.P. For this reason, we shall equip the sample space* $(\Omega, \mathcal{F})$ *with a monotone increasing family,* $\{\mathcal{F}_t; t \in [0, T]\}$, *of the sub-σ-field of* $\mathcal{F}$ *called filtration, meaning that* $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$, $0 \leq s \leq t \leq T$. *With respect to filtration* $\{\mathcal{F}_t; t \in [0, T]\}$, *we can define the σ-field of events prior to* $t \in (0, T]$ *by* $\mathcal{F}_{t_-} \triangleq \sigma(\bigcup_{s<t} \mathcal{F}_s)$, *and the σ-field of events immediately after* $t \in [0, T)$, *by* $\mathcal{F}_{t_+} \triangleq \sigma(\bigcap_{\epsilon \geq 0} \mathcal{F}_{t+\epsilon})$. *We say that* $\{\mathcal{F}_t; t \in [0, T]\}$ *is right-continuous if* $\mathcal{F}_{t_+} = \mathcal{F}_t$ *and left continuous if* $\mathcal{F}_{t_-} = \mathcal{F}_t$. *The space* $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}_{t\geq 0})$ *is called a filtered probability space. Finally, we say that* $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}_{t\geq 0})$ *satisfies the usual conditions if* $(\Omega, \mathcal{F}, P)$ *is complete,* $\mathcal{F}_0$ *contains all* P *sets of measure zero in* $\mathcal{F}$, *and* $\{\mathcal{F}_t\}_{t\geq 0}$ *is right-continuous.*

Definition 6.2.4 *Consider the continuous or discrete time R.P.,* $\{Y_t; t \in [0, T]\}$ *defined on the measurable space* $(\Omega, \mathcal{F})$ *with filtration* $\{\mathcal{F}_t; t \in [0, T]\}$, *which takes values on Borel measurable space* $(\Re^d, \mathcal{B}(\Re^d))$.

The R.P. $\{Y_t; t \in [0,T]\}$ is called $\{\mathcal{F}_t; t \in [0,T]\}$-adapted if for all $t \in [0,T]$, Y_t is an $\mathcal{F}_t$ measurable R.V. That is, for all $t \in [0,T]$, the mapping

$$Y(t,.) : (\Omega, \mathcal{F}) \to (\Re^d, \mathcal{B}(\Re^d)) \tag{6.89}$$

is $\mathcal{F}_t/\mathcal{B}(\Re^d)$-measurable.

One of the well known and frequently used continuous time stochastic process is Brownian motion. In the following we recall one dimensional standard Brownian motion.

Definition 6.2.5 *(Standard Brownian Motion)[65] Let $\{w(t) \in \Re;\ t \in I \subset \Re\}$ be a R.P. defined on $I \times \Omega$. $w(t)$ is said to be standard Wiener process or standard Brownian motion, if*

i) $w(0) = 0$, a.s. (i.e., $P(w \in \Omega; w(0) = 0) = 1$),

ii) $w(t)$ has independent increments $\{w(t) - w(\tau)\}$ for all $\tau, t \in I$ with $\tau \leq t$, and this set of increments is stationary with correlation function $E[|w(t) - w(\tau)|^2] = |t - \tau|$ for all $t, \tau \in I$,

iii) The trajectory of $w(t)$ is continuous a.s. with respect to the time variable $t \in I$,

iv) $E[w(t)] = 0$.

Next, we list some of the important properties of standard Brownian motion.

Remark 6.2.6 *(Properties of Brownian Motion)([65], Theorem 3.2, pp. 68) Let $(\Omega, \mathcal{F}, P)$ be a probability space, $I \subset \Re$ a time interval, and $w(t) \in \Re$ a standard Wiener process defined on $I \times \Omega$. Then,*

i) $w(t)$ is a Gaussian process.

ii) For each $n = 1, 2, ...$, the increments $\{w(t_i) - w(t_{i-1}) : t_i \in I, t_0 < t_1 < ... < t_n\}$ are orthogonal in the Hilbert space of second order R.V.'s in the sense that $E[(w(t_i) - w(t_{i-1}))(w(t_j) - w(t_{j-1}))] = 0$, $\forall i \neq j$.

iii) The covariance of $w(t)$ is given by $Cov(w(t), w(\tau)) = \min(t, \tau)$, $t, \tau \in I$.

Remark 6.2.7 *i)If two R.V.'s are independent then they are uncorrelated and the reverse implication holds only if they are Gaussian.*

ii) An i.i.d. Gaussian process is an i.i.d. process in which for every $\{t_1, t_2, ..., t_n\} \subset [0,T]$, it is jointly Gaussian distributed.

iii) R.V.'s X, Y, Z are said to be conditionally independent or forms a Markov chain denoted by $X \to Y \to Z$ if knowing Y, X and Z are independent. Please note that $X \to Y \to Z$ implies $Z \to Y \to X$ ([40], pp. 32).

In the following lemma a relation between the Power Spectral Density (PSD's) of the input and output of a stable linear system with transfer function $H(z)$ when the input is stationary process, is given.

Lemma 6.2.8 *([66], pp. 15) If an $m \times 1$ stationary process X_t with PSD, $S_X(e^{jw})$ is applied to an $p \times m$ stable linear system with transfer matrix $H(z)$ to yield an output process Y_t, we have the following relation between the PSD's of X_t and Y_t.*

$$S_Y(e^{jw}) = H(e^{jw})S_X(e^{jw})H'(e^{-jw}) \tag{6.90}$$

where $'$ denotes matrix transpose.

6.2.2 Information Theoretic Measures and Results

Classical Information Theory

One of the elementary information theoretic measures is Shannon (differential) entropy which is a quantity to measure the amount of uncertainty conveyed by a R.V.

Definition 6.2.9 *(Shannon Entropy) ([40], pp. 224) The Shannon entropy of a R.V. $Y \in \Re^d$ with PDF, f_Y is defined by*

$$H_S(Y) = H_S(f_Y) \triangleq - \int_{\Re^d} f_Y(y) \log f_Y(y) dy \tag{6.91}$$

where $\log(.)$ *denotes the logarithm of base* 2.

Other forms of entropy are the Rényi entropy [67] defined by

$$H_R(Y) = H_R(f_Y) \triangleq \frac{1}{1-\alpha} \log \int_{\Re^d} f_Y^\alpha(y) dy, \qquad \alpha > 0, \alpha \neq 1, \tag{6.92}$$

and the Tsallis entropy [68] defined by

$$H_T(Y) = H_T(f_Y) \triangleq \frac{1}{1-\alpha}\left(\int_{\Re^d} f_Y^\alpha(y) dy - 1\right). \tag{6.93}$$

The Rényi and Tsallis entropies give as special case the Shannon entropy. Specifically, $H_S(f_Y) = \lim_{\alpha\to 1} H_R(f_Y) = \lim_{\alpha \to 1} H_T(f_Y)$.

If R.V.'s X and Y have the joint density function $f_{X,Y}$, the conditional entropy, $H_S(X|Y)$, is defined as follow.

Definition 6.2.10 *(Conditional Entropy) ([40], pp. 230) Let R.V.'s X and Y have the joint density function $f_{X,Y}$ with conditional density function $f_{X|Y}$. Then,*

$$H_S(X|Y) \triangleq -\int f_{X,Y}(x,y)\log f_{X|Y}(x,y)dxdy. \tag{6.94}$$

The conditional entropy $H_S(X|Y)$ measures the amount of uncertainty conveyed by R.V. X when R.V. Y is known.

For a sequence of R.V.'s, the Shannon entropy rate is a quantity to measure the rate of growing of entropy with respect to the length of sequence. It is defined as follow.

Definition 6.2.11 *(Shannon Entropy Rate) ([40], pp. 63) For a sequence of R.V.'s $\{Y_t; t \in \mathbf{N}_+ \triangleq \{0,1,2,...\}\}$, the Shannon entropy rate is defined by*

$$\mathcal{H}_S(\mathcal{Y}) \triangleq \lim_{T\to\infty}\frac{1}{T}H_S(Y^{T-1}),\ Y^{T-1} \triangleq (Y_0,Y_1,...,Y_{T-1}) \tag{6.95}$$

provided the limit exists.

In addition to the Shannon entropy rate, we have the conditional entropy rate. The conditional Shannon entropy rate is defined by ([40], pp. 64)

$$\bar{\mathcal{H}}_S(\mathcal{Y}) \triangleq \lim_{T\to\infty} H_S(Y_T|Y^{T-1}),\ Y^{T-1} = (Y_0,Y_1,...,Y_{T-1}) \tag{6.96}$$

provided the limit exists.

From ([40], Theorem 4.2.1, pp. 64) follows that for a stationary discrete stochastic process, the limits in (6.95) and (6.96) exist; and they are equal, i.e., $\mathcal{H}_S(\mathcal{Y}) = \bar{H}_S(\mathcal{Y})$.

We have the following properties for Shannon entropy.

Remark 6.2.12 *(Shannon Entropy Properties)*

i) Conditioning reduces the entropy. That is, $H_S(X|Y) \leq H_S(X)$ with equality if and only if X and Y are independent ([40], pp. 232).

ii) Chain rule for Shannon entropy: $H_S(Y^T) = \sum_{t=0}^{T} H_S(Y_t|Y^{t-1})$ ($H_S(Y_0|Y^{-1}) = H_S(Y_0)$) ([40], pp. 232).

iii) $H_S(AY) = H_S(Y) + \log|\det(A)|$ ([40], pp. 234).

iv) Let $(Y_0,Y_1,...,Y_{T-1})$; $Y_i \in \Re^d$, $0 \leq i \leq T-1$ have a multivariate normal distribution with mean m and covariance $\Gamma_{Y^{T-1}}$, that is, $Y^{T-1} \sim N(m,\Gamma_{Y^{T-1}})$. Then, $H_S(Y^{T-1}) = \frac{1}{2}\log\left((2\pi e)^{Td}\det(\Gamma_{Y^{T-1}})\right)$ ([40], Theorem 9.4.1, pp. 230).

v) Let the R.V. $Y \in \Re^d$ has zero mean and covariance Γ_Y. Then, $H_S(Y) \leq \frac{1}{2}\log\big((2\pi e)^d\det(K$

)) *with equality if and only if* $Y \sim N(0,K)$ *([40], Theorem 9.6.5, pp. 234).*
vi) $H_S(f_Y)$ *is a concave function of density function* f_Y *(follows from [69]).*
vii) Translation does not change the Shannon entropy. That is, $H_S(Y+c) = H_S(Y)$ *for a fixed c ([40], Theorem 9.6.3, pp. 233).*

We now introduce mutual information which is a measure of the amount of information that one random variable conveys about another [40].

Definition 6.2.13 *(Mutual Information) Consider two R.V.'s* X *and* Y *with probability measure* $P(dX)$ *and* $P(dY)$ *respectively. Then, the mutual information* $I(X;Y)$ *between* X *and* Y *is defined by*

$$I(X;Y) \triangleq \int \log\Big(\frac{P(dY;X)}{P(dY)}\Big) P(dY;X)P(dX) \tag{6.97}$$

where $P(dY;x)$ *is stochastic kernel and* $\frac{P(dY;x)}{P(dY)}$ *is Radon-Nikodym derivative [77].*

We have the following important properties for mutual information.

Remark 6.2.14 *(Mutual Information Properties) i)* $I(X;Y) \geq 0$ *([70], Lemma 1.4.1) with equality if and only if* X *and* Y *are independent.*
ii) (Data Processing Inequality) Let the R.V.'s X, Y, Z *be absolutely continuous with respect to the Lebesgue measure; and form a Markov chain (Remark 6.2.7, iii), that is,* $X \rightarrow Y \rightarrow Z$*. Then,* $I(X;Y) \geq I(X;Z)$ *([40], pp. 32).*
iii) $I(X;Y) = I(Y;X)$ *(by definition).*
iv) $I(X;Y) = H_S(X) - H_S(X|Y) = H_S(Y) - H_S(Y|X)$ *([40], pp. 231).*

Next, we have the following results from [52, 71] which relates the mutual information of continuous time R.P.'s to the mean square estimation error; and subsequently, to PSD.

Lemma 6.2.15 *Consider the following Ito differential equation* $dy(t) = f(t,y(t),x(t))dt + dw(t)$*, where* $y(t), w(t) \in \Re^d$*,* $x(t) \in \Re^q$*,* $w(t)$ *is Brownian motion with* $E|w(t)-w(\tau)| = N_0|t-\tau|$ *and it is independent of the process* $y(t) \in \Re^d$*. Let* $\int_0^T E[f'(t,y(t),x(t))f(t,y(t),x(t))]dt < \infty$ *and* $y(0)=0$ *a.s. Then, for* $y = \{y(s); 0 \leq s \leq T\}$ *and* $x = \{x(s); 0 \leq s \leq T\}$*, we have [71]*

$$I_T(y;x) = \frac{1}{2}E\Big[\int_0^T (f(t,y(t),x(t)) - \hat{f}(t,y(t)))'(f(t,y(t),x(t)) - \hat{f}(t,y(t)))dt\Big], \tag{6.98}$$

where $I_T(.;.)$ denotes the mutual information between sample paths x and y and $\hat{f}(t, y(t)) = E[f(t, y(t), x(t))\big|\mathcal{F}^y_{0,t}]$; $\mathcal{F}^y_{0,t} \triangleq \sigma\{y(s); 0 \leq s \leq t\}$, where $\sigma\{.\}$ denotes the σ-algebra.
Further, for the case of $d = 1$, from explicit results for the mean square error for linear causal filtering in white noise [52] ([72], Section 7), we have

$$\lim_{t\to\infty} E[|f(t, y(t), x(t)) - \hat{f}(t, y(t))|^2] = \frac{N_0}{2\pi}\int_{-\infty}^{+\infty} \log_e(1 + N_0^{-1}S_f(w))dw \tag{6.99}$$

where $S_f(w)$ is the PSD of P.P. $f(t)$ (provided it is stationary).

Furthermore, the conditional mutual information of R.V.'s X and Y given Z is defined by

$$I(X;Y|Z) \triangleq \int \log(\frac{P(dY;X,Z)}{P(dY;Z)})P(dY;X,Z)P(dX;Z)P(dZ). \tag{6.100}$$

Theorem 6.2.16 *[53] Consider the model given by (3.1), (3.2), shown in Figure 3.1. The mutual information between the state of plant $x = \{x(s); 0 \leq s \leq T\}$, and the channel output $y = \{y(s); 0 \leq s \leq T\}$, conditional on the channel state $\theta = \{\theta(s); 0 \leq s \leq T\}$ is given by the following equivalent expressions*

$$i)\ I_T(x;y|\theta) \triangleq E_{x,y,\theta}\Big[\log_e \frac{P(dy;x,\theta)}{P(dy;\theta)}\Big] \tag{6.101}$$

$$ii)\ I_T(x;y|\theta) = \frac{1}{2}E_\theta \int_0^T z^2(t,\theta(t))E\Big[|f(t,x,\tilde{x},\theta)|^2 - |\hat{f}(t,\tilde{x},\theta)|^2\Big|\mathcal{F}^\theta_{0,t}\Big]dt \tag{6.102}$$

where $E_{x,y,\theta}[.]$ represents expectation with respect to the sample paths x, y, θ, $E_\theta[.]$ denotes expectation with respect to the sample path θ, and $\hat{f}(t,\tilde{x},\theta) = E[f(t,x,\tilde{x},\theta)\big|\mathcal{F}^{y,\theta}_{0,t}]$. Here, $P(dy;x,\theta)$ and $P(dy;\theta)$ are stochastic kernels and $\frac{P(dy;x,\theta)}{P(dy;\theta)}$ is Radon-Nikodym derivative.

Proof: ii) By using the methodology of [51] or Lemma 6.2.15 applied to mutual information, we deduce (6.102).
Next, we describe the communication channel; and subsequently we define the channel capacity.

Definition 6.2.17 *A communication channel at time $t \in \mathbf{N}_+$ is modeled by a sequence of stochastic kernels (channel law) $\{P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1})\}_{t\in\mathbf{N}_+}$, where $Z^t = z^t$ is the specific realization of the channel input, and $\tilde{Z}^{t-1} = \tilde{z}^{t-1}$, is the specific realization of the previous channel outputs.*
i) A communication channel is called memoryless if the stochastic kernel satisfies $P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1}) = P(d\tilde{Z}_t; z_t)$ which means that $\tilde{Z}_t$ is conditionally independent of $\tilde{Z}^{t-1}$ and Z^{t-1} given

Z_t.

ii) A communication channel is used without feedback if the stochastic kernel satisfies. $P(dZ_t; z^{t-1}, \tilde{z}^{t-1}) = P(dZ_t; z^{t-1})$.

iii) A feedback communication channel with memory is called non-anticipative or causal if the stochastic kernel satisfies $P(d\tilde{Z}_t; z^n, \tilde{z}^{t-1}) = P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1})$, $\forall n > t$.

iv) A communication channel in which the channel input and output are restricted to the finite alphabet sets and it processes the successive channel input independently of one another, is called a Discrete Memoryless Channel (DMC) [40].

v) An Additive White Gaussian Noise (AWGN) channel is described by $\tilde{Z}_t = Z_t + \tilde{W}_t$, *where* $\tilde{Z}_t \in \Re^d$ *and* $Z_t \in \Re^d$ *are the channel input and output at time* t *and the orthogonal process* $\{\tilde{W}_t\}_{t\in\mathbf{N}_+}$ *(i.e.,* $E[\tilde{W}_i\tilde{W}_j'] = 0$, $i \neq j$*) is a zero mean Gaussian process. This channel is subject to the power constraint* $E[Z_t'Z_t] \leq P_t$.

vi) A digital noiseless channel with transmission rate $\mathcal{R}$ *is a memoryless channel with channel input and output alphabet with size* $2^{\mathcal{R}}$ *which includes the strings of binaries with length* $\mathcal{R}$, *in which* $P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1}) = \delta_{\{\tilde{Z}_t - Z_t\}}$ *where* $\delta_{\{.\}}$ *is the Dirac measure [9].*

vii) An erasure channel with transmission rate $\mathcal{R}$ *and packet erasure probability* α *is a memoryless channel with channel input alphabet of size* $2^{\mathcal{R}}$ *which includes the string of binaries with length* $\mathcal{R}$*; but string of binaries with length* $\mathcal{R}$ *or erasure symbol as output, in which* $P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1}) = (1-\alpha)\delta_{\{\tilde{Z}_t - Z_t\}} + \alpha\delta_{\{\tilde{Z}_t - e_r\}}$, *where* e_r *stands for the erasure symbol [9].*

viii) A delayed digital noiseless channel with transmission delay $\mathcal{R}$ *and time delay* Δ *is a noiseless digital channel, in which* $P(d\tilde{Z}_t; z^t, \tilde{z}^{t-1}) = \delta_{\{\tilde{Z}_t - Z_{t-\Delta}\}}$ *[9].*

Next, we define information channel capacity, in which for some channels (e.g., DMC's and AWGN without feedback) represents (operational) channel capacity which is the highest rate per channel use at which information can be sent with arbitrary low probability of error.

Definition 6.2.18 *(Information Capacity)[40] Consider a communication channel and let* Z^{n-1} *and* $\tilde{Z}^{n-1}$ *be the channel input and output sequences, respectively. Let* $\mathcal{M}_{CI}$ *denotes the set of joint probability measures* $P(dZ^{n-1})$ *which satisfy certain channel input power constraint. The Shannon information capacity for the time horizon* n *(i.e.,* n *channel uses) is defined by*

$$\mathcal{C}_n \triangleq \sup_{P(dZ^{n-1})\in\mathcal{M}_{CI}} I(Z^{n-1}; \tilde{Z}^{n-1}). \tag{6.103}$$

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

For DMC's without feedback and AWGN channel without feedback, the information channel capacity of Definition 6.2.18 represents the operational capacity [40]; or simply the channel capacity.

Example 6.2.19 *i) AWGN Channel: The capacity of an AWGN channel (see Definition 6.2.17, v) under the power constraint* $E[Z_t^2] \leq P$*;* $Z_t \in \Re$ *when* $E[\tilde{W}_t^2] = W_c$ *for* n *channel uses (see Definition 6.2.18) is* $\mathcal{C}_n = \frac{n}{2}\log(1+\frac{P}{W_c})$ *bits per* n *channel uses; and subsequently* $\mathcal{C} = \frac{1}{2}\log(1+\frac{P}{W_c})$ *bits per channel use [9] ([40], Theorem 10.1.1, pp. 242).*
ii) Digital Noiseless Channel: The capacity of a digital noiseless channel with rate $\mathcal{R}$ *(Definition 6.2.17, vi) for* n *channel uses is* $\mathcal{C}_n = n\mathcal{R}$ *bits per* n *channel uses; and subsequently the capacity is* $\mathcal{C} = \mathcal{R}$ *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 (bits per time step) is also* $\mathcal{C} = \mathcal{R}$ *[9]. Here,* $\mathcal{R}$ *is the transmission data rate measured in bits per time step. We have similar results for the capacity of the delayed digital noiseless channel.*
iii) Erasure Channel: The capacity of an erasure channel with rate $\mathcal{R}$ *(Definition 6.2.17, vii) and packet erasure probability* α *for* n *channel uses is* $\mathcal{C}_n = (1-\alpha)n\mathcal{R}$ *bits per channel uses; and subsequently the capacity is* $\mathcal{C} = (1-\alpha)\mathcal{R}$ *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 (bits per time step) is also* $\mathcal{C} = (1-\alpha)\mathcal{R}$ *[9].*

Lemma 6.2.20 *[53]. Consider the model given by (3.1), (3.2) subject to the power constraint* $E[|f(t,x,\tilde{x},\theta)|^2 \big| \mathcal{F}_{0,t}^{\theta}] \leq P$ *and when the channel state information is known for transmitter and receiver. Then, the mutual information between the state of plant* $x = \{x(s); 0 \leq s \leq T\}$*, and the channel output* $y = \{y(s); 0 \leq s \leq T\}$*, conditional on the channel state* $\theta = \{\theta(s); 0 \leq s \leq T\}$ *(see Theorem 6.2.16 for definition) is given by,*

$$\frac{1}{T} I_T(x;y|\theta) \leq \frac{1}{2T} P \int_0^T E[z^2(t,\theta(t))]dt. \qquad (6.104)$$

Moreover, the channel capacity for communication channel (3.2) subject to the power constraint $E[|f(t,x,\tilde{x},\theta)|^2 \big| \mathcal{F}_{0,t}^{\theta}] \leq P$*, is given by (see Definition (3.4.1))*

$$\mathcal{C}_f = \liminf_{T\to\infty} \frac{P}{2T} \int_0^T E[z^2(t,\theta(t))]dt. \qquad (6.105)$$

Proof: According to (6.102) and by considering the instantaneous power constraint

$$I_T(x;y|\theta) \leq \frac{1}{2}E_\theta \int_0^T z^2(t,\theta(t))E\Big[|F(t,x,\tilde{x},\theta)|^2\Big|\mathcal{F}^\theta_{0,t}\Big]dt \leq \frac{P}{2}\int_0^T E[z^2(t,\theta(t))]dt. \quad (6.106)$$

Then, following the same methodology as in ([51], Section 16.4), it is shown that the above upper bound determines the channel capacity. Namely, the signal x that reaches the channel capacity is a white Gaussian noise.

Theorem 6.2.21 *(Coding Theorem) [53]. Suppose the received signal is defined by (3.2), and the source by (3.1). Then, the encoder, which achieves the channel capacity (6.105), the optimal decoder, and the corresponding error variance, are respectively, given by*

$$f^*(t,x,\tilde{x}^*,\theta) = \sqrt{\frac{P}{V^*(t,y,\theta)}}(x(t) - \tilde{x}^*(t,y,u,\theta)) \quad (6.107)$$

$$d\tilde{x}^*(t,y,u,\theta) = A(t)\tilde{x}^*(t,y,u,\theta)dt + N(t)u(t)dt + z(t,\theta(t))\sqrt{PV^*(t,y,\theta)}dy(t), \quad (6.108)$$

$$V^*(t,y,\theta) = V^*(0)e^{(2\int_0^t A(s)ds - \int_0^t z^2(s,\theta(s))Pds)} + \int_0^t G^2(s)e^{(2\int_s^t A(u)du - \int_s^t z^2(u,\theta(u))Pdu)}ds, \quad (6.109)$$

where $\tilde{x}^*(0) = \bar{x}_0$, and $V^*(0) = \bar{V}_0$.

Proof: We first find (f_0^*, f_1^*) that minimizes the conditional error variance $V(t,y,\theta)$ described by the differential equation (3.15). Then, by substituting the corresponding encoder, decoder, and conditional error variance given in (6.107)-(6.109), into the conditional mutual information of Theorem 6.2.16, we deduce that the upper bound of Lemma 6.2.20 is achieved in capacity given by (6.105) (the proof is a variant of the one given in [51], Section 16.4). Next, we recall information rate distortion function in which for certain sources (e.g., memoryless stationary sources) represents the (operational) rate distortion function which is the minimum bit rate under which a reliable data reproduction up to a given distortion value is possible.

Definition 6.2.22 *(Shannon Information Rate Distortion)[49] Let Y^{T-1} and $\tilde{Y}^{T-1}$ be sequences of length T of the source and the reproduction 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 value and* $\rho_T \in [0, \infty)$ *is the distortion measure. Then, the information rate distortion for time horizon* T *is defined by*

$$R_T(D_v) \triangleq \inf_{\{P(d\tilde{Y}^{T-1}; y^{T-1}) \in \mathcal{M}_{DC}\}} I(Y^{T-1}; \tilde{Y}^{T-1}) \tag{6.110}$$

and subsequently, the information rate distortion is

$$R(D_v) \triangleq \lim_{T \to \infty} \frac{1}{T} R_T(D_v), \tag{6.111}$$

provided the limit exists.

It can be shown that ([49], Section 4.2) the stochastic kernel which achieves the minimum of the rate distortion function is given by

$$P^*(d\tilde{Y}^{T-1}; y^{T-1}) = \frac{e^{s(D_v)\rho_T(y^{T-1}, \tilde{y}^{T-1})} P(d\tilde{Y}^{T-1})}{\int_{\Re^{Td}} e^{s(D_v)\rho_T(y^{T-1}, \tilde{y}^{T-1})} P(d\tilde{Y}^{T-1})}, \quad s(D_v) \leq 0, \tag{6.112}$$

where $s(D_v)$ is the solution of $s(D_v) = \frac{dR(D_v)}{dD_v}$.
In control applications the sequential rate distortion is important since it assumes instantaneous (causal) reconstruction. That is, the current output symbol at the communication end, is determined by looking into the past and current values of source messages. To the best of our knowledge, this measure first introduced in [45] as follows.

Definition 6.2.23 *(Sequential Rate Distortion) [45] Sequential rate distortion for time horizon* T *is given by*

$$R_T^{SRD}(D_v) \triangleq \inf_{\left\{\{P(d\tilde{Y}_t; \tilde{y}^{t-1}, y^t)\}_{t=0}^{T-1} \in \mathcal{M}_{DC}^{SRD}\right\}} I(Y^{T-1}; \tilde{Y}^{T-1}) \tag{6.113}$$

where $\mathcal{M}_{DC}^{SRD} = \left\{\{P(d\tilde{Y}_t; \tilde{y}^{t-1}, y^t)\}_{t=0}^{T-1}; E\rho(Y_t; \tilde{Y}_t) \leq D_v, \forall t\right\}$.

Next, we present the Shannon lower bound which is equivalent to the Shannon rate distortion function under certain conditions.

Lemma 6.2.24 *(Shannon Lower Bound). Let* Y^{T-1}, $Y_t \in \Re^d$, $0 \leq t \leq T-1$ *be a sequence with length* T *produced by the source* $P(dY^{T-1}) = f_{Y^{T-1}} dY^{T-1}$. *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* $\rho(y_t, \tilde{y}_t) = \rho(y_t - \tilde{y}_t)$:

$\Re^d \rightarrow [0,\infty)$ *is continuous. Then,*

i) A lower bound for $\frac{1}{T}R_T(D_v)$ *is given by*

$$\frac{1}{T}R_T(D_v) \geq \frac{1}{T}H_S(Y^{T-1}) - \max_{h \in G_D} H_S(h), \tag{6.114}$$

where $h(\xi)$ *is a density function which belongs to the set* G_D*, in which it is defined by* $G_D \triangleq \{h : \Re^d \rightarrow [0,\infty); \int_{\Re^d} h(\xi)d\xi = 1,\ \int_{\Re^d} \rho(\xi)h(\xi)d\xi \leq D_v,\quad \xi \in \Re^d\}$*. Moreover, when* $\int_{\Re^d} e^{s\rho(\xi)}d\xi < \infty$ *for all* $s < 0$*, then* $h^*(\xi) \in G_D$ *that maximizes* $H_S(h)$ *is*

$$h^*(\xi) = \frac{e^{s\rho(\xi)}}{\int_{\Re^d} e^{s\rho(\xi)}d\xi}, \quad \int_{\Re^d} \rho(\xi)h^*(\xi)d\xi = D_v. \tag{6.115}$$

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

$$R(D_v) \geq \mathcal{H}_S(\mathcal{Y}) - \max_{h \in G_D} H_S(h) \triangleq R_S(D_v). \tag{6.116}$$

ii) Suppose the difference distortion measure $\rho(.)$ *satisfies the following conditions a)* $\rho(\xi) = 0$ *if and only if* $\xi = 0$*,* $\lim_{||\xi|| \rightarrow 0} \rho(\xi) = 0$*, and* $\lim_{||\xi|| \rightarrow \infty} \rho(\xi) = \infty$*. b) There exists a monotone increasing function* $\Psi : [0,\infty) \rightarrow [0,\infty)$ *such that* $\Psi(u) > 0$ *if* $u > 0$ *and* $\rho(\xi) \geq \Psi(||\xi||)$ *for all* $\xi \in \Re^d$*. c)* $\int_{\Re^d} e^{s\rho(\xi)}d\xi < \infty$ *for all* $s < 0$*. d) There exists an* $c > 0$ *such that* $\rho(\xi + \tilde{\xi}) \leq c\rho(\xi) + c\rho(\tilde{\xi})$ *for all* $\xi, \tilde{\xi} \in \Re^d$*. e)* $\mathcal{H}_S(\mathcal{Y}) > -\infty$*. f) There exists an* $\xi^* \in \Re^d$ *such that* $E\rho(\xi - \xi^*) < \infty$*,* $\forall \xi \in \Re^d$*.*

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

Proof: Follows from [43] by considering the method proposed in ([49], pp. 140) or [73].

Remark 6.2.25 *i) A sufficient condition for the existence of* $R(D_v)$ *is stationarity of the source [49].*

ii) For distortion measure $\rho(y, \tilde{y}) = ||y - \tilde{y}||^r$ *(for* $y \in \Re^d$*,* $||.||$ *is the Euclidian norm), in the limit as* $D_v \rightarrow 0$*, the Shannon lower bound is equal to the rate distortion function [43].*

iii) For $\rho(y_t, \tilde{y}_t) = ||y_t - \tilde{y}_t||^r$*,* $\max_{h \in G_D} H_S(h) = \log e^{\frac{d}{r}} - \log(\frac{r}{dV_d\Gamma(\frac{d}{r})}(\frac{d}{rD_v})^{\frac{d}{r}})$ *bits per time step, where* $\Gamma(.)$ *is the gamma function and* V_d *is the volume of the unit sphere (e.g.,* $V_d = Vol(S_d); S_d \triangleq \{\xi \in \Re^d; ||\xi|| \leq 1\}$*) [43].*

Next, we present a solution to the rate distortion function $R_T(D_v)$ associated with the orthogonal Gaussian Random Process (R.P.) and single letter mean square distortion measure.

Example 6.2.26 *(Rate Distortion For a Parallel Gaussian Source) Let $Y_t \in \Re \sim N(0, \Lambda_t)$, $t \in \{0, 1, ..., T-1\}$ be independent (orthogonal) Gaussian random variables and consider the following single letter mean square distortion measure $\rho_T(y^{T-1}; \tilde{y}^{T-1}) = \frac{1}{T}\sum_{t=0}^{T-1}(y_t - \tilde{y}_t)^2$. Then, under the assumption that $D_v < \min_{t\in\{0,1,...,T-1\}} \Lambda_t$, the rate distortion function associated with this source and distortion measure is given by*

$$R_T(D_v) = \sum_{t=0}^{T-1} \frac{1}{2} \log \frac{\Lambda_t}{D_v} \tag{6.117}$$

where this distortion is achieved by the following minimizing stochastic kernel

$$P^*(d\tilde{K}^{T-1}; k^{T-1}) = (\prod_{t=0}^{T-1} q^*_{\tilde{K}_t|K_t})d\tilde{K}^{T-1}; \quad q^*_{\tilde{K}_t|K_t} \sim N(\beta_t k_t, \beta_t D_v), \ \beta_t \triangleq 1 - \frac{D_v}{\Lambda_t}. \tag{6.118}$$

Proof: Follows from ([40], Theorem 13.3.3) by considering the single letter distortion measure. (6.118) also follows from ([49], pp. 99-100).

Next, we recall relative entropy [70]. Relative entropy plays a key role in the definition of every rate function. The relative entropy between two probability measures $P(dY)$ and $\Pi(dY)$ defined on the measurable space $(\Re^d, \mathcal{B}(\Re^d))$ is defined as follows.

Definition 6.2.27 *(Relative Entropy)([70], pp. 32) The relative entropy between probability measures $P(dY)$ and $\Pi(dY)$ defined on the measurable space $(\Re^d, \mathcal{B}(\Re^d))$ is a mapping into extended real line defined as follows.*

i) If $P(dY)$ is absolutely continuous with respect to $\Pi(dY)$, that is, $P(dY) << \Pi(dY)$ ($<<$ denotes the absolute continuity relation), the relative entropy is defined by

$$H(P||\Pi) = \int_{\Re^d} \log(\frac{P(dY)}{\Pi(dY)})P(dY), \quad \log(\frac{P(dY)}{\Pi(dY)}) \in L_1(P(dY)) \tag{6.119}$$

where $L_1(P(dY))$ denotes the set of integrable functions with respect to probability measure $P(dY)$ which are defined on the measurable space $(\Re^d; \mathcal{B}(\Re^d))$; and $\frac{P(dY)}{\Pi(dY)}$ denotes Radon Nikodym Derivative (RND) [77].

ii) Otherwise, we set $H(P||\Pi) = \infty$.

Remark 6.2.28 *(Properties of Relative Entropy) Let $P(dY) \in \mathcal{M}_1(\Re^d)$ and $\Pi(dY) \in \mathcal{M}_1(\Re^d)$ be two probability measures on the Borel measurable space $(\Re^d, \mathcal{B}(\Re^d))$ ($\mathcal{M}_1(\Re^d)$ is the set of all countably additive probability measures defined on the measurable space $(\Re^d, \mathcal{B}(\Re^d))$). Then, the relative entropy $H(.||.)$ has the following properties.*

i)(Legendre-Fenchel Transformation)[74] For every $\Psi(y)$ bounded from below and $P(dY)$, $\Pi(dY)$

$$\log_e \int e^{\Psi(y)} \Pi(dy) = \sup_{P(dY);H(P||\Pi)<\infty} \{ \int \Psi(y) P(dY) - H(P||\Pi) \} \tag{6.120}$$

where $\log_e()$ denotes logarithm of natural number.
Moreover, if $\Psi(y)e^{\Psi(y)} \in L_1(\Pi(dY))$, then the supremum in (6.120) is attained at $P^*(dY)$ given by

$$\frac{P^*(dY)}{\Pi(dY)} = \frac{e^{\Psi(y)}}{\int e^{\Psi(y)} \Pi(dY)}. \tag{6.121}$$

ii) (Convexity) ([70], pp.37) $H(P||\Pi)$ is a convex, lower semi-continuous function of (P, Π). In addition, for a fixed $\Pi(dY)$, $H(.||\Pi)$ is strictly convex on the set $\{P(dY); H(P||\Pi) < \infty\}$.
iii) (Compactness) ([70], pp. 37) For a fixed $\Pi(dY)$, $H(.||\Pi)$ has a compact level sets. That is, for each $R_c < \infty$, the set $\{P(dY); H(P||\Pi) \leq R_c\}$ is a compact subset of $\mathcal{M}_1(\Re^d)$, where $\mathcal{M}_1(\Re^d)$ is the set of all countably additive probability measures defined on the measurable space $(\Re^d, \mathcal{B}(\Re^d))$.
iv) The definition of relative entropy can be extended to two arbitrary positive measures. In particular, if $f_Y, g_Y \in L_1(\Re^d, \Re^+)$ ($L_1(\Re^d, \Re^+)$ is the set of integrable functions defined on the space $\Re^d$ which take values in $\Re^+$), then $P(dY) = f_Y dY$ and $\Pi(dY) = g_Y dY$ are positive measures associated with f_Y and g_Y and $H(P||\Pi)= H(f_Y||g_Y)= \int \log_e(\frac{f_Y}{g_Y}) f_Y dy$. Furthermore, using ([40], p. 29, Theorem 2.7.1), it follows that for a fixed $g_Y \in L_1(\Re^d, \Re^+)$, $H(f_Y||g_Y)$ is strictly convex function of $f_Y \in L_1(\Re^d, \Re^+)$.

Next, we recall the information transmission theorem which relates the channel capacity (per source message) to the rate distortion function for reliable communication.

Theorem 6.2.29 *(Direct Information Transmission Theorem) ([49], pp. 72) For all $\epsilon > 0$ the discrete (finite alphabet) memoryless source $\{Y_t, P(dY_t)\}$ can be reproduced with fidelity $D_v + \epsilon$ at the output of an DMC of capacity $C > R(D_v) + \epsilon$ bits per source message.*

Under conditional independence assumption between source, channel input-output, and reconstruction of the source, we have the converse of information transmission theorem as follow.

Theorem 6.2.30 *(Converse Information Transmission Theorem) ([49], pp. 72) It is impossible to reproduce the discrete memoryless source of $\{Y_t, P(dY_t)\}$ with fidelity D_v at the receiving end of any DMC of capacity $C < R(D_v)$ bits per source message.*

Remark 6.2.31 *The extension of these results to the case of continuous memoryless sources can be found in ([75], Theorem 9.6.3, pp. 473) in which an extension of this result to continuous amplitude stationary ergodic source is also given in ([75], Theorem 9.8.3, pp. 500). The Converse Information Transmission Theorem is also valid for AWGN channels.*

Finally, we close this section by recalling some interesting results from analysis that we used throughout thesis.

Lemma 6.2.32 *(Cesaro Mean) ([40], Theorem 4.2.3, pp. 64) If* $a_n \to a$ *and* $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$*, then* $b_n \to a$.

Lemma 6.2.33 *(Lagrange Duality)([76], pp. 224) Let* f *be a real valued convex function defined on convex subset* Ω *of a vector space* X*, and let* G *be a convex mapping of* X *into a normed space* Z*. Suppose there exists an* x_1 *such that* $G(x_1) < \theta$ *and that* $\inf\{f(x) : G(x) \leq \theta; x \in \Omega\}$ *is finite.*
Then,

$$\inf_{G(x)\leq\theta; x\in\Omega} f(x) = \max_{z^*\geq\theta} \inf_{x\in\Omega}[f(x) + z^*(G(x) - \theta)] \tag{6.122}$$

where z^* *belongs to the dual space of* Z *and the maximum on the right side is achieved by some* $z_0^* \geq \theta$.

Lemma 6.2.34 *(Lebesgue's Dominated Convergence Theorem)([77], Theorem 2.2.4, pp. 72) Let* $(\mathcal{Y}, \mathcal{F}, \mu)$ *be a measurable space,* g *be a* $[0, +\infty]$*-valued integrable function on* $\mathcal{Y}$*, and* f *and* $f_1, f_2, ...$ *be* $[-\infty, +\infty]$*-valued* $\mathcal{F}$ *measurable functions on* $\mathcal{Y}$ *such that the relations*
i) $f(y) = \lim_{n\to\infty} f_n(y)$
and
ii) $|f_n(y)| \leq g(y), \quad n = 1, 2, ...$
hold at almost every y *in* $\mathcal{Y}$*. Then,* f *and* f_1*,* f_2*, ... are integrable and* $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$.

Information Theory and Uncertainty

In this section, we recall some of information theoretic measures and results such as Sakrison (robust) rate distortion which have been developed to study reliable communication in the presence of uncertainty in the source of information.
An information source is uncertain if the source probability measure $P(dY^T)$ is unknown,

but it belongs to the known class $\mathcal{M}_{SU}^T$. In the presence of uncertainty in the source, the rate distortion function is defined as follows.

Definition 6.2.35 *(Sakrison Information Rate Distortion) [63] Let Y^{T-1} and $\tilde{Y}^{T-1}$ be sequences of length T of the source and the corresponding reproduction of the source messages, respectively; and let $P(dY^{T-1}) \in \mathcal{M}_{SU}^T$ denote the probability measure of Y^{T-1} which belongs to the class $\mathcal{M}_{SU}^T$. Let also $\mathcal{M}_{DC} \triangleq \{P(d\tilde{Y}^{T-1}; y^{T-1}); E\rho_T(Y^{T-1}, \tilde{Y}^{T-1}) \leq D_v\}$ be 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 time horizon T and for the family $\mathcal{M}_{SU}^T$ is defined by

$$R_{T,r}(D_v) \triangleq \inf_{P(d\tilde{Y}^{T-1}; y^{T-1}) \in \mathcal{M}_{DC}} \sup_{P(dY^{T-1}) \in \mathcal{M}_{SU}^T} I(Y^{T-1}; \tilde{Y}^{T-1}) \tag{6.123}$$

and subsequently, the robust information rate distortion in each time step is given by

$$R_r(D_v) = \lim_{T \to \infty} \frac{1}{T} R_{T,r}(D_v). \tag{6.124}$$

If the uncertain source is memoryless, distortion measure is single letter, and the uncertainty set is a compact set, then the above information definition for rate distortion represents the minimum bit rate for uniform reliable data reconstruction of the uncertain source, up to distortion value D_v.

Remark 6.2.36 *An extension of Theorems 6.2.29 and 6.2.30 to the case of uncertain continuous memoryless source when the source probability measure belongs to a compact set $\mathcal{M}_{SU}^T$ and the distortion measure is single letter, is given in [63], in which it is shown that*

$$R_{T,r}(D_v) = \sup_{P(dY^{T-1}) \in \mathcal{M}_{SU}^T} \inf_{P(d\tilde{Y}^{T-1}; y^{T-1}) \in \mathcal{M}_{DC}} I(Y^{T-1}; \tilde{Y}^{T-1}). \tag{6.125}$$

6.2.3 System and Control Theory Results

Review of System Theoretic Concepts

In this section we recall elementary concepts such as controllability, Stabilizability, and detectability. Throughout, we denote continuous time variables by the lowercase letters (e.g., $y(t)$) and the discrete time variables by the capital letters (e.g., Y_t). The system matrices are also shown by the capital letters.

Definition 6.2.37 *Consider the following time-invariant system*

$$\dot{x}(t) = Ax(t) + Nu(t), \quad x(t) \in \Re^q, \quad u(t) \in \Re^o. \tag{6.126}$$

For time-invariant system (6.126) we have the following condition for controllability ([56], Theorem 1.23, pp. 55).
The q-dimensional time-invariant system (6.126) is controllable if and only if the column vectors of the controllable matrix $P \triangleq (N \quad AN \quad A^2N \quad . \quad . \quad . \quad A^{q-1}N)$ *span the q-dimensional space. That is,* $Rank(P) = q$*. For this case* (A, N) *is called a controllable pair. For discrete time-invariant analogous system, we have similar definition for controllability.*

Next, we define stable and unstable systems.

Definition 6.2.38 *(Stable and Unstable System) Consider the continuous time-invariant system* $\dot{x}(t) = Ax(t) + NU(t)$ *(resp.* $X_{t+1} = AX_t + NU_t$ *for discrete time version). This system is stable if and only if its eigenvalues are all stable (i.e., for continuous time systems all the eigenvalues have negative real parts; and for discrete time systems all eigenvalues are inside the unit circle).*

Now, we are ready to recall the stabilizability concept.

Definition 6.2.39 *Continuous time (resp. discrete time) system (6.126) is stabilizable if and only if there exists a matrix* K *such that* $A+NK$ *is stable matrix (i.e., all the eigenvalues of* $A + NK$ *have negative real part for continuous system; or they are inside unit circle for discrete time system). For this case,* (A, N) *is called a stabilizable pair which simply means that there exists a matrix* K *such that* $A + NK$ *is a stable matrix.*

Next, we recall detectability concept for the following linear time-invariant continuous-time system.

$$\begin{cases} \dot{x}(t) = Ax(t) + Nu(t), & x(0) \in \Re^q, \ u(t) \in \Re^o, \\ y(t) = Cx(t), & y(t) \in \Re^d \end{cases} \tag{6.127}$$

where $x(t)$ is the state of system, $u(t)$ is the control signal, and $y(t)$ is the observation provided by sensors.

Definition 6.2.40 *[23] The continuous time (resp. discrete time analogous) system (6.127) is detectable if and only if there exists a matrix* K *such that* $A' + C'K$ *is stable matrix (i.e.,* *the eigenvalues of $A' + C'K$ have negative real parts for continuous system; or they are inside the unit circle for discrete time system). For this case, (C, A) is called a detectable pair which simply means that there exists a matrix K such that $A' + C'K$ is a stable matrix.*

System Theoretic Results

In this section, we recall some of the frequently used system and control results such as Bode integral formula and the LQG control.

We start this section by Bode integral formula. Bode integral formula relates sensitivity transfer function of single input, single output linear time-invariant systems to unstable eigenvalues of system.

Theorem 6.2.41 *(Discrete Generalized Bode's Theorem)[44] Consider a discrete, single input, single-output, linear, time invariant system with open-loop transfer function $L(z) \triangleq \frac{K\prod_{i=1}^{m}(z-z_i)}{\prod_{i=1}^{n}(z-p_i)}$ where $K \neq 0$ and it is chosen to stabilize the closed-loop system. The p_i's are open-loop poles, some of them being allowed outside the open-unit disk. The sensitivity function of the system is defined as $S(z) \triangleq \frac{1}{1+L(z)}$. Then,*

$$\int_{-\pi}^{\pi} \log|S(e^{jw})|dw = 2\pi(\sum_i \log|p_{u_i}| - \log|\gamma + 1|), \tag{6.128}$$

where p_{u_i}'s are the unstable (i.e., outside the closed-unit disk) open-loop poles, and $\gamma \triangleq \lim_{z\to\infty} L(z)$.

Please recall that if $\gamma = 0$, then $L(z)$ is strictly proper (i.e., $L(z) = \frac{N(z)}{D(z)}$, where the degree of the polynomial $D(z)$ is greater than the degree of polynomial $N(z)$), and if $\gamma \neq 0$, $L(z)$ is proper (i.e., $L(z) = \frac{N(z)}{D(z)}$, where the degree of the polynomial $D(z)$ is greater or equal to the degree of polynomial $N(z)$).

Theorem 6.2.42 *(Revised Generalized Bode's Theorem) [44] Consider a single-input, single-output, linear, continuous, time-invariant system with open loop transfer function $L(s) \triangleq \frac{K\prod_{i=1}^{m}(s-z_i)}{\prod_{i=1}^{n}(s-p_i)}$ where $K \neq 0$ and it is chosen such that the closed loop system is asymptotically stable, and the p_i's are the open loop poles, some of them being allowed on the imaginary axis. Let $v \triangleq n - m$. The sensitivity function is $S(s) \triangleq \frac{1}{1+L(s)}$. Then,*

$$\int_{-\infty}^{\infty} \log_e|S(jw)|dw = \begin{cases} +\infty & if \quad v = 0 \; and \; -2 < K < 0 \\ -\infty & if \quad v = 0 \; and \; K > 0 \; or \; K < -2 \\ 2\pi(\sum_i Re(z_i) - \sum_i Re(p_{s_i})) & if \quad v = 0 \; and \; K = -2 \\ 2\pi\sum_i Re(p_{u_i}) - \gamma\pi & if \quad v = 1 \; where \; \gamma = \lim_{s\to\infty} sL(s) \\ 2\pi\sum_i Re(p_{u_i}) & if \quad v \geq 2 \end{cases} \tag{6.129}$$

where p_{u_i}, p_{s_i} and z_i denote the ith unstable open-loop pole, the ith stable open-loop pole, and the ith open-loop zero, respectively.

Please recall that if $v = 0$, $L(s)$ is proper and if $v \geq 1$, $L(s)$ is strictly proper.
In addition to above results, we have the following result from [78] which relates conditional Shannon entropy rate to the transfer function of a linear discrete time-invariant stable system.

Lemma 6.2.43 *([78], Lemma 1) Let the random process $\{Y_t; t \in \mathbf{N}_+\}$ be obtained by passing the random process $\{U_t; t \in \mathbf{N}_+\}$ through a stable linear discrete time-invariant system described by transfer function $F(z)$. Then,*

$$\bar{\mathcal{H}}(\mathcal{Y}) = \bar{\mathcal{H}}(\mathcal{U}) + \int_{-\pi}^{\pi} \log_e |F(e^{jw})| dw, \tag{6.130}$$

where $\bar{\mathcal{H}}(.)$ denotes the conditional Shannon entropy rate (see (6.96)).

Next, we recall one of the well known mean square estimators, known as Kalman filter.

Definition 6.2.44 *(Discrete Time Kalman Filter) ([23], Theorem 4.1, pp. 158) Consider the following dynamical system*

$$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t\geq 0}) : \begin{cases} X_{t+1} = AX_t + W_t, \quad X_0, \\ Y_t = CX_t + G_t, \end{cases} \tag{6.131}$$

where $X_t \in \Re^q$ is the state of system, X_0 is the initial state in which $\bar{x}_0 = E[X_0]$, $\bar{V}_0 = E\big[(X_0 - \bar{x}_0)(X_0 - \bar{x}_0)'\big]$, $W_t \in \Re^m$ is the process noise with $E[W_t W_k'] = Q\delta_{tj}$ and $E[W_t] = 0$, $Y_t \in \Re^d$ is the observation obtained by sensors, $G_t \in \Re^l$ is the measurement noise with $E[G_t G_k'] = R\delta_{tk}$, $E[G_t] = 0$ and $\{W_t, G_t, X_0\}$ are mutually orthogonal. Then, the solution of the estimation problem $\hat{X}_{t|t-1} \triangleq E[X_t|\sigma\{Y^{t-1}\}]$ is given by the following recursive equation known as the Kalman filter equation.

$$\begin{aligned} \hat{X}_{t+1|t} &= A\hat{X}_{t|t-1} + \Delta_t(Y_t - C\hat{X}_{t|t-1}), \quad \hat{X}_{0|-1} = \bar{x}_0 \\ \Delta_t &= AV_tC'(CV_tC' + R)^{-1}, \end{aligned} \tag{6.132}$$

where V_t is the solution of the following forward recursive equation known as Riccati equation

$$V_{t+1} = AV_tA' - AV_tC'(CV_tC' + R)^{-1}CV_tA' + Q, \quad V_0 = \bar{V}_0, \tag{6.133}$$

where from ([23], Theorem 4.2) under the assumption that (C, A) is detectable and $(A, Q^{\frac{1}{2}})$ is stabilizable, there exists a unique positive semi-definite $V_\infty = \lim_{t\to\infty} V_t$.

Next, we recall one of the well known quadratic regulators, known as LQG control.

Definition 6.2.45 *(Linear Quadratic Gaussian Control) ([23], Theorem 9.2, pp. 697) Consider the following dynamical system*

$$\begin{cases} X_{t+1} = AX_t + NU_t + W_t, \quad X_0, \\ Y_t = CX_t + G_t, \end{cases} \tag{6.134}$$

where $X_t \in \Re^q$, $U_t \in \Re^o$ *is the control signal,* $W_t \in \Re^m$, $Y_t \in \Re^d$, *and* $G_t \in \Re^l$, $X_0 \sim N(\bar{x}_0, \bar{V}_0)$, W_t *i.i.d.* $\sim N(0,Q)$, G_t *i.i.d.* $\sim N(0,R)$, $R > 0$.
Let (C,A) *be detectable and* $(A, Q^{\frac{1}{2}})$ *be stabilizable. Let also* $((C'C)^{\frac{1}{2}}, A)$ *be detectable and* (A,N) *be stabilizable. Then, the stabilizing control law,* $U_t \in \mathcal{U}_t \triangleq \{U_t : \Re^{t(o+d)} \rightarrow \Re^o; U_t \in \mathcal{G}^U_{t-1}\}$; $\mathcal{G}^U_t \triangleq \sigma\{Y_0, ..., Y_t, U_0, ..., U_t\}$, *that minimizes the following cost functional*

$$\lim_{T\rightarrow\infty}\{\inf_{U^{T-1}\in\mathcal{U}_0\times...\times\mathcal{U}_{T-1}} \frac{1}{T}\sum_{t=0}^{T-1} E(||X_t||^2_{C'C} + ||U_t||^2_H)\}, \quad (H > 0) \tag{6.135}$$

is given by

$$U_t = -\Delta\hat{X}_{t|t-1}, \tag{6.136}$$

where

$$\Delta = (H + N'P_\infty N)^{-1}N'P_\infty A, \tag{6.137}$$

in which P_∞ *is the unique positive semi definite solution of the following control Riccati equation*

$$P_\infty = A'P_\infty A - A'P_\infty N(N'P_\infty N + H)^{-1}A'P_\infty N + C'C \tag{6.138}$$

and $\hat{X}_{t|t-1}$ *is the Kalman filter state estimation, given by (see (6.132))*

$$\hat{X}_{t+1|t} = A\hat{X}_{t|t-1} + \Delta_t(Y_t - C\hat{X}_{t|t-1}) + NU_t, \quad \hat{X}_{0|-1} = \bar{x}_0. \tag{6.139}$$

Furthermore, the optimal cost is equal to

$$\begin{aligned} &\lim_{T\rightarrow\infty}\{\inf_{U^{T-1}\in\mathcal{U}_0\times...\times\mathcal{U}_{T-1}} \frac{1}{T}\sum_{t=0}^{T-1} E(||X_t||^2_{C'C} + ||U_t||^2_H)\} \\ &= trac(C'CV_\infty) + trac(AV_\infty C'(CV_\infty C' + R)^{-1}CV_\infty A'P_\infty) \end{aligned} \tag{6.140}$$

where V_∞ *is the unique positive semi definite solution of the following Algebraic Riccati equation*

$$V_\infty = AV_\infty A' - AV_\infty C'(CV_\infty C' + R)^{-1}CV_\infty A' + Q. \tag{6.141}$$

The continuous time version of Definition 6.2.45 is given below.

Definition 6.2.46 *(Linear Quadratic Gaussian Control) ([56], Theorem 5.4, pp. 394) Consider the dynamical system*

$$(\Omega, \mathcal{F}(\Omega), P; \{\mathcal{F}_t\}_{t\geq 0}) : \begin{cases} \dot{x}(t) = A(t)x(t) + N(t)u(t) + w(t), \quad X(0), \\ y(t) = C(t)x(t) + g(t), \end{cases} \tag{6.142}$$

where $x(t) \in \Re^q$, $u(t) \in \Re^o$, $w(t) \in \Re^m$, $y(t) \in \Re^d$, $g(t) \in \Re^l$, $x(0)$ is stochastic vector, with mean $\bar{x}_0$ and covariance matrix $\bar{V}_0$, and $w(t)$, $g(t)$ are zero mean orthogonal white noises. That is, $Cov(w(t), w(\tau)) = V_1(t)\delta(t-\tau)$, $Cov(g(t), g(\tau)) = V_2(t)\delta(t-\tau)$, and $Cov(w(t), g(t)) = 0$.

Then, the stabilizing controller that minimizes the following cost functional

$$J = \frac{1}{T}E\{\int_0^T [x^{'}(t)R_3(t)x(t) + u^{'}(t)R_2(t)u(t)]dt\}, \quad R_2(t) > 0, \; R_3(t) \geq 0 \tag{6.143}$$

is given by $u(t) = -F(t)\hat{x}(t)$, where $F(t) = R_2^{-1}(t)N^{'}(t)P(t)$ in which the non-negative definite matrix $P(t)$ is the solution of the following Riccati equation

$$-\dot{P}(t) = R_3(t) - P(t)N(t)R_2^{-1}(t)N^{'}(t)P(t) + A^{'}(t)P(t) + P(t)A(t), \; P(T) = 0. \tag{6.144}$$

Further, $\hat{x}(t) = E[x(t)|\sigma\{y(s); 0 \leq s \leq t\}]$ is obtained as a solution of the following generalized Kalman filter equation.

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + K(t)(y(t) - C(t)\hat{x}(t)) + N(t)u(t), \quad \hat{x}(0) = \bar{x}_0, \tag{6.145}$$

where $K(t) = V(t)C^{'}(t)V_2^{-1}(t)$ in which the covariance matrix $V(t)$ is the solution of the following Riccati equation

$$\dot{V}(t) = V_1(t) - V(t)C^{'}(t)V_2^{-1}(t)C(t)V(t) + A(t)V(t) + V(t)A^{'}(t), \quad V(0) = \bar{V}_0. \tag{6.146}$$

Furthermore,

$$\begin{aligned} \bar{J} &= \lim_{T\to\infty} \frac{1}{T} trac(\int_0^T [\bar{P}(t)\bar{K}(t)V_2(t)\bar{K}^{'}(t) + \bar{V}(t)R_1(t)]dt) \\ &= \lim_{T\to\infty} \frac{1}{T} trac(\int_0^T [\bar{P}(t)V_1(t) + \bar{V}(t)\bar{F}^{'}(t)R_2(t)\bar{F}(t)]dt) \end{aligned} \tag{6.147}$$

where $\bar{P}(t)$ is the steady state solution of the Riccati equation (6.144), $\bar{V}(t) = \lim_{t\to\infty} V(t)$, $\bar{K}(t) = \bar{V}(t)C^{'}(t)V_2^{-1}(t)$ is the gain corresponding to the steady state solutions of $\bar{V}(t)$ and $\bar{F}(t) = R_2^{-1}(t)N^{'}(t)\bar{P}(t)$.

In addition, in the time-invariant case where $\bar{V}(t) = \bar{V}$, $\bar{P}(t) = \bar{P}$ and thus also $\bar{F}(t) = R_2^{-1}N'\bar{P}$ and $\bar{K}(t) = \bar{V}C'V_2^{-1}$ are constant matrices, the following expression holds

$$\bar{J} = trac(\bar{P}\bar{K}V_2\bar{K}' + \bar{V}R_1) = trac(\bar{P}V_1 + \bar{V}\bar{F}'R_2\bar{F}) \tag{6.148}$$

Please notice that from (6.144)-(6.146) follows that separation principle holds between estimation and control.

Next, in the following lemma, we present conditions under which the Riccati equation (6.144) has a steady state solution.

Lemma 6.2.47 *([56], Theorem 3.7 (c), pp. 238) For the time-invariant analogous system of (6.142) (i.e., when $A(t) = A$, $N(t) = N$ and $C(t) = C$), when $R_3(t) = R_3 > 0$ and $R_2(t) = R_2 > 0$, if the system is controllable, $\bar{P}(t) = \bar{P}$ is a non-negative definite symmetric solution to the following Algebraic Riccati equation.*

$$0 = R_3 - \bar{P}NR_2^{-1}N'\bar{P} + A'\bar{P} + \bar{P}A. \tag{6.149}$$

In addition to above result for existence of the steady state solution, we have the following result for the existence of the solution for the Discrete Algebraic Riccati Equation (DARE).

Lemma 6.2.48 *([79], Theorem 12.7.1, pp. 302) Consider the following DARE*

$$X = A'XA + Q - (C + B'XA)'(R + B'XB)^{-1}(C + B'XA) \tag{6.150}$$

where $A \in \Re^{n \times n}$, $B \in \Re^{n \times m}$, $C \in \Re^{m \times n}$, $Q = Q' \in \Re^{n \times n}$, $R = R' \in \Re^{m \times m}$ and $X \in \Re^{n \times n}$ is a real symmetric matrix which must be found.
If (A, B) is controllable, A and R are invertible, and $\Psi(\eta) > 0$ for some η, $|\eta| = 1$, where $\Psi(z)$ is the rational matrix function given by

$$\Psi(z) = R + B'(z^{-1}I_n - A')^{-1}Q(zI_n - A)^{-1}B, \tag{6.151}$$

DARE (6.150) has a real symmetric solution.

Bibliography

[1] D. F. Delchamps, Stabilizing a Linear System with Quantized State Feedback, *IEEE Transactions on Automatic Control*, vol. 35, No. 8, pp. 916-924, Auguest 1990.

[2] W. S. Wong and R. W. Brockett, Systems with Finite Communication Bandwidth Constraints-Part I: State Estimation Problems, *IEEE Transactions on Automatic Control*, vol. 42, No. 9, pp. 1294-1299, September 1997.

[3] W. S. Wong and R. W. Brockett, Systems with Finite Communication Bandwidth Constraints-II: Stabilization with Limited Information Feedback, *IEEE Transactions on Automatic Control*, vol. 44, No. 5, pp. 1049-1053, May 1999.

[4] Anant Sahai, *Anytime Information Theory*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT., February 2001.

[5] G. N. Nair and R. J. Evans, "Communication-Limited Stabilization of Linear Systems", in *Proceedings of 39th IEEE Conference on Decision and Control*, pp. 1005-1010, Sydney, December 2000.

[6] G. N. Nair and R. J. Evans, "Mean Square Stabilisability of Stochastic Linear Systems with Data Rate Constraints", in *Proceedings of 41st IEEE Conference on Decision and Control*, pp. 1632-1637, Las Vegas, December 2002.

[7] G. N. Nair and R. J. Evans, Stabilizability of Stochastic Linear Systems with Finite Feedback Data Rates, *SIAM Journal of Control and Optimization*, vol. 43, No. 2, pp. 413-436, 2004.

[8] G. N. Nair, S. Dey and R. J. Evans, "Infimum Data Rates for Stabilising Markov Jump Linear Systems", in *Proceedings of 42nd IEEE Conference on Decision and Control*, pp. 1176-1181, Hawaii, December 2003

[9] S. Tatikonda and S. Mitter, Control over Noisy Channels, *IEEE Transactions on Automatic Control*, vol. 49, No. 7, pp. 1196-1201, July 2004.

[10] S. Tatikonda and S. Mitter, Control under Communication Constraints, *IEEE Transactions on Automatic Control*, vol. 49, No. 7, pp. 1056-1068, July 2004.

[11] S. Tatikonda, A. Sahai, and S. Mitter, Stochastic Linear Control over a Communication Channel, *IEEE Transactions on Automatic Control*, vol. 49, No. 9, pp. 1549-1561, Semptember 2004.

[12] Nicola Elia, When Bode Meets Shannon: Control-Oriented Feedback Communication Schemes, *IEEE Transactions on Automatic Control*, vol. 49, No. 9, pp. 1477-1488, September 2004.

[13] K. Li and J. Baillieul, Robust Quantization for Digital Finite Communication Bandwidth (DFCB) Control, *IEEE Transactions on Automatic Control*, vol. 49, No. 9, pp. 1573-1584, September 2004.

[14] G. N. Nair, R. J. Evans, I. M. Y. Mareels and W. Moran, Topological Feedback Entropy and Nonlinear Stabilization, *IEEE Transations on Automatic Control*, vol. 49, No. 9, pp. 1585-1597, September 2004.

[15] D. Liberzon and J. P. Haspanha, Stabilization of Nonlinear Systems with Limited Information Feedback, *IEEE Transactions on Automatic Control*, vol. 50, No. 6, pp. 910-915, June 2005.

[16] N. C. Martins, A. Dahleh, and N. Elia, Feedback Stabilization of Uncertain Systems in the Presence of a Direct Link, *IEEE Transactions on Automatic Control*, vol. 51, No. 3, March 2006.

[17] S. Tatikonda, "Some Scalling Properties of Large Distributed Control Systems," in *Proceedings of 42nd IEEE Conference on Decision and Control*, pp. 3142-3147, Maui, Hawaii, USA, December 2003.

[18] S. Yuksel and T. Basar, "On the Absence of Rate Loss in Decentralized Sensor and Controller Structure for Asymptotic Stability", in *Proceedings of the 2006 American Control Conference*, pp. 5562-5567, Minnesota, June 14-16, 2006.

[19] Alexey S. Matveev, and Andrey V. Savkin, "Stabilization of Multisensors Networked Control Systems with Communication Constraints," in *5th Asian Control Conference*, pp. 1905-1913, 2004.

[20] G. N. Nair, R. J. Evans and P. E. Caines, "Stabilising Decentralised Linear Systems under Data Rate Constraints," in *Proceedings of 43rd IEEE Conference on Decision and Control*, pp. 3992-3997, Atlandis, Paradise Island, Bahamas, December 2004.

[21] K. Li and J. Baillieul, "Problems in Decentralized Sensor-Actuator Networks," in *Proceedings of 44th IEEE Conference on Decision and Control, and European Control Conference*, pp. 3207-3212, Seville, Spain, December 2005.

[22] S. Yuksel and T. Basar, "Coding and Control over Diecrete Noisy Forward and Feedback Channels," in *Proceedings of 44th IEEE Conference on Decision and Control, and the European Control Conference*, pp. 2517-2522, Seville, December 12-15, 2005.

[23] P. E. Caines, *Linear Stochastic Systems*, John Wiley and Sons, 1988.

[24] L. Schenato, B. Sinoploli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of Control and Estimation over Lossy Networks", in *Proceedings of IEEE, Special Issue on Networked Control Systems*, pp. 163-187, vol. 95, No. 1, January 2007.

[25] V. Gupta, D. Spanos, B. Hassibi, and R. M. Murray, Optimal LQG Control accross a Packet-Dropping Link, *Systems and Control Letters*, 2005.

[26] Y. Xu and J. Hespanha, "Estimation under Controlled and Uncontrolled Communications in Networked Control Systems", in *Proceedings of 44th IEEE Conference on Decision and Control*, pp. 842-847, Seville, December 12-15, 2005.

[27] S. Smith and P. Seiler, "Estimation with Lossy Measurements: Jump Estimators for Jump Systems", *IEEE Transactions on Automatic Control*, pp. 1453-1464, vol. 48, No. 12, December 2003.

[28] J. Nilson, B. Bernhardsson, and B. Wittenmark, "Stochastic Analysis and Control of Real-Time Systems with Random Time Delays", *Automatica*, pp. 57-64, vol. 34, No. 1, January 1998.

[29] J. Nilsson, *Real-Time Control Systems with Delays*, Ph.D. Thesis, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden, 1998.

[30] C. N. Hadjicostis and R. Touri, "Feedback Control Utilizing Packet Dropping Network Links", in *Proceedings of the IEEE Conference on Decision and Control*, pp. 1205-1210, Las Vegas, December 2002.

[31] Q. Ling and M. Lemmon, "Optimal Dropout Compensation in Networked Control Systems", *Proceedings of the IEEE Conference on Decision and Control*, Maui, HI, December 2003.

[32] N. Yu, L. Wang, G. Xie, and T. Chu, "Stabilization of Networked Control Systems with Data Packet Dropout via Swithed System Approach", *IEEE International Symposium Computer Aided Control Systems Design*, pp. 362-367, Taipei, Taiwan, September 2004.

[33] P. Seiler and R. Sengupta, An H_∞ Approach to Networked Control, *IEEE Transactions on Automatic Control*, pp. 356-364, vol. 50, No. 3, March 2005.

[34] A. V. Savkin and I. R. Petersen, Robust Filtering with Missing Data and a Deterministic Description of Noise and Uncertainty, *International Journal System and Science*, pp. 373-390, vol. 28, No. 4, 1997.

[35] J. do Val, J. Geromel, and O. Costa, Solutions for the Linear Quadratic Control Problem of Markov Jump Linear Systems, *Journal of Optimization Theory and Applications* pp. 283-311, vol. 103, No. 2, 1999.

[36] O. Costa, Stationary Filter for Linear Minimum Mean Square Error Estimator of Discrete-Time Markovian Jump Systems, *IEEE Transactions on Automatic Control*, pp. 1351-1356, vol. 48, No. 8, August 2002.

[37] Z. Wang, D. Ho, and X. Liu, Variance-Constrained Control for Uncertain Stochastic Systems with Missing Measurements, *IEEE Transactions on System, Man and Cybernatic A: Systems and Humans*, pp. 746-753, vol. 35, No. 5, September 2005.

[38] N. Elia and J. Eisembeis, "Limitation of Linear Control over Packet Drop Networks", *Proceedings of the IEEE Conference on Decision and Control*, pp. 5152-5157, Bahamas, December 2004.

[39] N. Elia, Remote Stabilization over Fading Channels, *Systems and Control Letters*, pp. 237-249, vol. 54, 2005.

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

[41] K. Tsumura and J. Maciejowski, "Stabilizability of SISO Control Systems under Constraints of Channel Capacities", in *Proceedings of 42nd IEEE Conference on Decision and Control*, Hawaii, pp. 193-197, December 2003.

[42] S. Tatikonda, A. Sahai, and S. Mitter, " Control of the LQG Systems Under Communication Constraints", in *Proceedings of 37th IEEE Conference on Decision and Control*, pp. 1165-1170, Tampa, December 1998.

[43] T. Linder and R. Zamir, On the Asymptotic Tightness of the Shannon Lower Bound, *IEEE Transactions on Information Theory*, vol. 40, No. 6, pp. 2026-2031, November 1994.

[44] B. F. Wu, and E. A. Jonckheere, A Simplified Approach to Bode's Theorem for Continuous-Time and Discrete-Time Systems, *IEEE Transactions on Automatic Control*, vol. 37, No. 11, pp. 1797-1802, November 1992.

[45] Sekhar Tatikonda, *Control Under Communication Constraint*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT., September 2000.

[46] J. Massey, "Causality, Feedback and Directed Information", in the *Proceedings of the 1990 Symposium on Information Theory and its Applications*, pp. 303-305, 1990.

[47] E. Baccarelli and R. Cusani, Linear Feedback Communication Systems with Markov Sources: Optimal System Design and Performance Evaluation, *IEEE Transactions on Information Theory*, vol. 41, No. 6, pp. 1868-1876, 1995.

[48] T. M. Cover and S. Bombra, Gaussian Feedback Capacity, *IEEE Transactions on Information Theory*, vol. 35, No. 1, pp. 37-43, 1989.

[49] Toby Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Printice-Hall, 1971.

[50] M. Gastper, B. Rimoldi and M. Vetterli, To Code, or Not to Code: Lossy Source-Channel Communication Revisited, *IEEE Transactions on Information Theory*, vol. 49, No. 5, pp. 1147-1158, May 2003.

[51] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes-Applications II*, Springer-Verlang, New York, 1978.

[52] T. T. Kadota, M. Zakai, and J. Ziv, Mutual Information of the White Gaussian Channel with and without Feedback, *IEEE Transactions on Information Theory*, vol. 17, No. 4, pp. 368-371, November 1971.

[53] C. D. Charalambous and Stojan Denic, "On the Channel Capacity of Wireless Fading Channels", in *Proceedings of 41st IEEE Conference on Decision and Control*, Las Vegas, December 2002.

[54] S. Ihara, Coding Theorems for a Continuous-Time Gaussian Channel with Feedback , *IEEE Transactions on Information Theory*, vol. 40, No. 6, pp. 2041-2045, November 1994.

[55] E. Biglieri, J. Proakis, and S. Shamai (Shitz), Fading Channels: Information-Theoretic and Communications Aspects, *IEEE Transactions on Information Theory*, vol. 44, No. 6, pp. 2619-2692, October 1998.

[56] H. Kwakernnak and R. Sivan, *Linear Optimal Control System*, John Wiley and Sons, 1972.

[57] I. R. Petersen and M. R. James, Performance Analysis and Controller Synthesis for Nonlinear Systems with Stochastic Uncertainty Constraints, *Automatica*, vol. 32, pp. 959-972, 1996.

[58] S. O. R. Moheimani, A. V. Savkin, and I. R. Petersen, A Connection between H^∞ Control and the Absolute Stailizability of Discrete-Time Uncertain Systems, *Automatica*, vol. 31, pp. 1193-1195, 1995.

[59] I. R. Petersen, M. R. James and P. Dupuis, Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints, *IEEE Transactions on Automatic Control*, vol. 45, No. 3, pp. 398-412, March 2000.

[60] A. V. Savkin and I. R. Petersen, Minimax Optimal Control of Uncertain Systems with Structured Uncertainty, *International Journal of Robust Nonlinear Control*, vol. 5, pp. 119-137, 1995.

[61] A. V. Savkin and I. R. Petersen, Nonlinear versus Linear Control in the Absolute Stabilizability of Uncertain Linear Systems with Structured Uncertainty, *IEEE Transactions on Automatic Control*, vol. 40, pp. 122-127, 1995.

[62] A. Megretski and A. Rantzer, System Analysis via Integral Quadratic Constraints, *IEEE Transactions on Automatic Control*, vol. 42, pp. 819-830, 1997.

[63] David J. Sakrison, The Rate Distortion Function for a Class of Sources, *Information and Control*, vol. 15, pp. 165-195, 1969.

[64] C. D. Charalambous, *Stochastic Systems*, Lecture Note, Fall 2001.

[65] G. Chen, Goong Chen, S. H. Hsu, *Linear Stochastic Control Systems*, Boca Roton, 1995.

[66] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation and Control A Unified Approach to H^2 and H^∞ Theories*, SIAM Stadies in Applied and Numerical Mathematics, 1999.

[67] A. Rényi, "On Measures of Entropy and Information", in *Proceedings of 4th Berkeley Symposion on Mathematical Statistics and Probability*, vol. I, Berekeley, CA, pp. 547-561, 1961.

[68] C. Tsallis, Possible Generalization of Boltzmann-Gibbs Statistics, *Journal of Statistics Physics*, vol. 52, p. 479, 1988.

[69] Takayuki Furuta, Simple Proof of the Concavity of Operator Entropy $f(A) = -A \log A$, *Mathematical Inequalities and Applications*, vol. 3, No. 2, pp. 305-306, 2000.

[70] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley, 1997.

[71] T. Duncan, On the Calculation of Mutual Information, *SIAM Journal of Appliled Mathematics*, vol. 19, pp. 215-220, July 1970.

[72] T. Kailath, *Lectures on Linear Least-Squares Estimation*, Springer-Veralg, 1976.

[73] J. T. Pinkston, *Information Rates of Independent Sample Sources*, M.S. Thesis, Department of Electerical Engineering, MIT., Cambridge, Mass. 1966.

[74] P. D. Pra, L. Meneghini, and W. Runggaldier, Some Connections between Stochastic Control and Dynamic Games, *Mathematics of Control, Signals, and Systems*, vol. 9, pp. 303-326, 1996.

[75] R.G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons, INC., 1968.

[76] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley, 1969.

[77] Donald L. Cohn, *Measure Theory*, Boston, 1980.

[78] G. Zang and P. A. Iglesias, Nonlinear Extension of Bode's Integral Based on an Information-Theoretic Interpretation, *Systems and Control Letters*, vol. 50, pp. 11-19, 2003.

[79] P. Lancaster and L. Rodman, *Algebraic Riccati Equations*, Oxford Science Publications, 1995.