Information Theoretic Aspects of the Control and the Mode Estimation of Stochastic Systems

by

Nuno Miguel Lara Cintra Martins

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2004

Author ....
Department of Electrical Engineering and Computer Science
August 31st, 2004

Certified by ....
Munther A. Dahleh
Professor of EECS
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Information Theoretic Aspects of the Control and Mode Estimation of Stochastic Systems

by

Nuno Miguel Lara Cintra Martins

Submitted to the Department of Electrical Engineering and Computer Science
on August 31st, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In this thesis, we investigate three problems: the first broaches the control under information constraints in the presence of uncertainty; in the second we derive a new fundamental limitation of performance in the presence of finite capacity feedback; while the third studies the estimation of Hidden Markov Models. **Problem 1:** We study the stabilizability of uncertain stochastic systems in the presence of finite capacity feedback. We consider a stochastic digital link that sends words whose size is governed by a random process. Such link is used to transmit state measurements between the plant and the controller. We derive necessary and sufficient conditions for internal and external stabilizability of the feedback loop. In addition, stability in the presence of uncertainty in the plant is analyzed using a small-gain argument. **Problem 2:** We address a fundamental limitation of performance for feedback systems, in the presence of a communication channel. The feedback loop comprises a discrete-time, linear and time-invariant plant, a channel, an encoder and a decoder which may also embody a controller. We derive an inequality of the form $L_{-} \geq \sum \max\{0, \log(|\lambda_i(A)|)\} - C_{channel}$, where L_{-} is a measure of disturbance rejection, A is the open loop dynamic matrix and $C_{channel}$ is the Shannon capacity of the channel. Our measure L_{-} is non-negative and smaller L_{-} indicates better rejection (attenuation), while $L_{-} = 0$ signifies no rejection. Additionally, we prove that, under a stationarity assumption, L_{-} admits a log-sensitivity integral representation. **Problem 3:** We tackle the problem of mode estimation in switching systems. From the theoretical point of view, our contribution is twofold: creating a framework that has a clear parallel with a communication paradigm and deriving an analysis of performance. In our approach, the switching system is viewed as an encoder of the mode, which is interpreted as the message, while a probing signal establishes a random code. Using a distortion function, we define an uncertainty ball where the estimates are guaranteed to lie with probability arbitrarily close to 1. The radius of the uncertainty ball is directly related to the entropy rate of the switching process.

Thesis Supervisor: Munther A. Dahleh
Title: Professor of EECS

Acknowledgments

I would like to thank Prof. Munther A. Dahleh for accepting me as a student in his group and for being a wonderful advisor over these years. He has taught me how to view a problem from different angles. His intuition is impressive and his ability to extract the essence of a problem is something he was able to instill in me. Prof. Dahleh has also taught me how to appreciate and strive for simplicity and quality. His moral support has been essential, especially in the more stressful moments. He is the one who best understands my work and his appreciation of it is a tremendous motivation. Anything I say about Prof. Michael Athans is not enough to make him justice. He has changed my life completely by convincing me to pursue a dream. Prof. Michael Athans is the main responsible for my decision to come to the U.S.A. and to study at MIT. He has also been a mentor and a dear friend. I will never forget what he did for me. His sense of mission as a Professor is inspiring and an example for me. I am grateful to Prof. Nicola Elia for the rich collaboration and help. He was the responsible for steering my research towards one of the main topics of my Thesis. I have learned a lot from him. Prof. Sanjoy Mitter for motivating me to work in interesting problems. He has played an important role in my education by suggesting interesting ideas and for giving valuable advice. Discussions with Prof. Mitter are always very inspiring and informative. His sense of mission as a Professor is also a great example for me. I also would like to thank him for his support in several occasions. It was also an honor to have Prof. John Tsitsiklis in my Thesis committee. I am grateful for the time he found for our meetings in his busy schedule. To Prof. John Doyle, I would like to thank for his motivation and support. He has the gift of understanding quickly. Prof. John Doyle not only understood my work, but he was already able to call my attention to the several implications of it. He has been tremendously supportive and extremely inspiring. Prof. George Verghese for his support since I came to MIT. His feedback and support were also very important as a member of my committee. The wise advice of Prof. Verghese has made me one of the many students that he has helped. I would like to thank Prof. Richard Melrose for being a dedicated Professor and for teaching me serious Mathematics. His motivation and help were very important in many occasions. Prof. Jeffrey Shapiro was the best academic advisor one can possible get. His dedication and genuine interest for my career were absolutely essential. I am indebted to Prof. Dimitri Bertsekas for his moral support when I arrived in Boston. Prof. Saligrama Venkatesh has also played a major role in my education at MIT. I have learned a lot from him. He was the one who motivated me to study Information Theory. He is a tremendous mind teaser and I feel very fortunate to have had the opportunity to work with him. I also would like to thank Prof. Eric Feron for his help. His moral support has also been very important over these years. The help of Sekhar Tatikonda was also important, by making available some pre-prints of his most recent results. I also would like to thank Prof. Alex Megretski for interesting discussions. The dedicated help from Doris Inslee and Fifa Monserrate were also critical.

In the other side of the Atlantic Ocean, I also would like to thank Prof. João Lemos for being a mentor and a dear friend. I would like to express my gratitude to Prof. João Sentieiro for his support and to Prof. Agostinho Rosa, Prof. Maria João Martins, Prof. António Serralheiro and Prof. António Pascoal for their motivation and help. To my colleagues at LIDS I would like to express my gratitude for creating a fantastic atmosphere of collaboration. In particular to my office-mates Georgios Kotsalis, Neal Jameson, Sridevi Sarma and Soosan Beheshti. How could I have survived without the discussions with Georgios over late night coffee? Or the laughing contests with Neal? From my dear friend Jorge Gonçalves I have received support, motivation, help and some of the most happy moments in Boston. I also would like to thank Ola Ayaso for patiently printing and handing in my Thesis while I was in Lisbon. Chung Yao Kao was a also a fantastic colleague with whom I had so many interesting discussions. To Miguel and Helga Arsenio, Madalena Costa, Rafaela Arriaga and Tiago Ribeiro I just say that they were my family in Boston. To my Portuguese friends, I would like thank João Soares, Miguel Barão and Joaquim Filipe for their unconditional friendship.

Finally and most of all to my dearest ones: My mother Maria do Céu Martins, my father João Cintra Martins, my brother Victor Martins and my aunt Fátima Lara are the basis of who I am. And to Amorzinho ... I owe my heart. It is thanks to her that I am able to move on every day. Amorzinho is my true inspiration. So that you know it is you: Napoleon - Gazpacho - DarguiDarga.

I also would like to thank the PRAXIS XXI program from the Portuguese Foundation for Science and Technology and the European Social Fund for their Financial Support through the fellowship BD/19630/99. I also would like to thank Prof. Diogo de Lucena and the Gulbenkian Foundation for some financial support in my first year at MIT.

Contents

List of Figures

Chapter 1

Introduction

In an attempt to delineate research directions for this century, investigators have recently gathered all over the world in workshops and study groups. The main reports that came out [19, 15] alluded to new challenges, both in control and estimation. Among the consentient ideas was the further study of control and system identification, under an information theoretic viewpoint. The adoption of such approach is backed by:

- The need to analyze and design feedback systems in the presence of finite capacity feedback. This problem is also intimately related to the state estimation of a dynamic system when the state information must be conveyed through a finite capacity channel.
- The potential of using the information theoretic framework and results to study estimation and system identification.

In this thesis, we identify three problems which are related to the items listed above. We describe the first in section 1.1, while the second is expounded in section 1.2 and the third is included in section 1.3. The remaining chapters explore each of these problems in detail. In particular, Chapter 2 refers to problem 1, while Chapters 3 and 4 address problems 2 and 3, respectively. The Chapters are, mostly, self contained as they are extended versions of published papers. The chapters are organized according to the following list:

- Chapter 2 [44]: Feedback Stabilization of Uncertain Systems Using a Stochastic Digital Link.

- Chapter 3 [42] Fundamental Limitations of Disturbance Attenuation in the Presence of Finite Capacity Feedback.
- Chapter 4 [41]: An Information Theoretic Approach to the Modal Estimation of FIR Linear Systems

1.1 Feedback Stabilization of Uncertain Systems

Motivated by applications, such as remote feedback, control in the presence of information constraints has received considerable attention. Certainly, the exploration of such problems is exciting as they foster the interaction between the disciplines of Information Theory and Control. Finite capacity feedback results from the use of an analog communication channel or a digital link as a way to transmit information about the state of the plant. It can also be the abstraction of computational constraints created by several systems sharing a common decision center.

1.1.1 Problem Statement

We wish to study the stabilizability of uncertain stochastic systems in the presence of finite capacity feedback. Motivated by the structure of communication networks, we consider a stochastic digital link that sends words whose size is governed by a random process. Such link is used to transmit state measurements between the plant and the controller. We intend to derive necessary and sufficient conditions for robust internal and external stabilizability of the feedback loop. We expect that stability in the presence of uncertainty in the plant can be analyzed using a small-gain argument.

In Figure 1-1, we depict the system interconnection that we have adopted in our study. One of the motivations for the proposed formulation is the control of systems entrenched in uncertain environments that change stochastically. Stochasticity in the environment impacts the link and the plant by introducing randomness and uncertainty. Estimation and modeling errors are also accounted for as uncertainty in the plant.

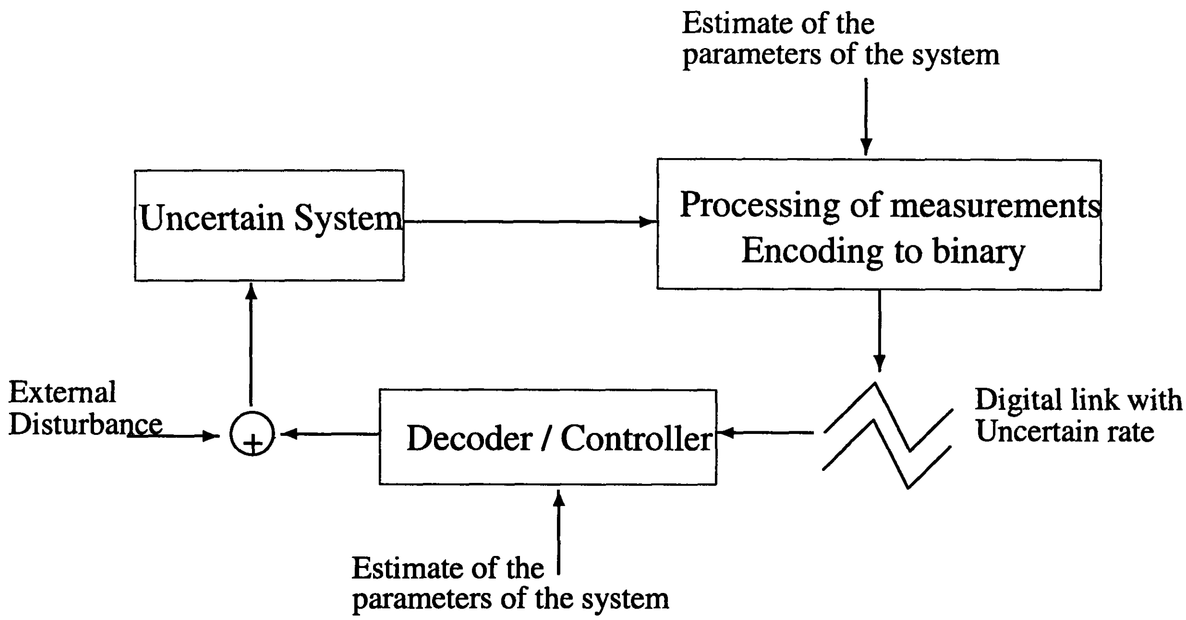

Figure 1-1: Control under information constraints paradigm

1.1.2 Discussion of Previous Publications

Various publications in this field have introduced necessary and sufficient conditions for the stabilizability of unstable plants in the presence of data-rate constraints. The construction of a stabilizing controller requires that the data-rate of the feedback loop is above a non-zero critical value [59, 60, 50, 54, 39]. Different notions of stability have been investigated, such as containability [66, 67], moment stability [54] and stability in the almost sure sense [59]. The last two are different when the state is a random variable. That happens when disturbances are random or if the communication link is stochastic. In [59] it is shown that the necessary and sufficient condition for almost sure stabilizability of finite dimensional linear and time-invariant systems is given by an inequality of the type $\mathcal{C} > \mathcal{R}$. The parameter $\mathcal{C}$ represents the average data-rate of the feedback loop and $\mathcal{R}$ is a quantity that depends on the eigenvalues of A, the dynamic matrix of the system. If a well defined channel is present in the feedback loop then $\mathcal{C}$ may be taken as the Shannon Capacity. If it is a digital link then $\mathcal{C}$ is the average transmission rate. Different notions of stability may lead to distinct requirements for stabilization. For tighter notions of stability, such as in the m-th moment sense, the knowledge of $\mathcal{C}$ may not suffice. More informative notions, such as higher moments or any-time capacity [54], are necessary. Results for the problem of state estimation in the presence of information constraints can be found in [66], [55] and [38].

1.1.3 Main Contributions of Chapter 2

In Chapter 2, we study the moment stabilizability of a class of uncertain time-varying stochastic systems in the presence of a stochastic digital link. In contrast with [49], we consider systems whose time-variation is governed by an identically and independently distributed (i.i.d.) process which may be defined over a continuous and unbounded alphabet. We also provide complementary results to [49, 21, 23, 35] because we use a different problem formulation where we consider external disturbances and uncertainty on the plant and a stochastic digital link.

Our work provides a unified framework for the necessary and sufficient conditions for robust stabilizability by establishing that the average transmission rate must satisfy $C > R + \alpha + \beta$, where $\alpha, \beta \geq 0$ are constants that quantify the influence of randomness in the link and the plant, respectively. As a consequence, C must be higher than R to compensate for randomness both in the plant and the digital link. The conclusion that $C > R$ does not suffice in the presence of a stochastic link was originally derived by [55]. We quantify such difference for stochastic digital links. The work of [55] was an important motivation for our work and the treatment of the nominal stabilization of first order linear systems, using a parameterized notion of capacity[1] for general channels, can found there. If the plant and the link are deterministic, we get $\beta, \alpha = 0$ which is consistent with the condition $C > R$ derived by [59]. We also show that model uncertainty in the plant can be tolerated. By using an appropriate measure, we prove that an increase in C leads to higher tolerance to uncertainty. All of our conditions for stability are expressed as simple inequalities where the terms depend on the description of uncertainty in the plant as well as the statistics of the system and the digital link. A different approach to deal with robustness, with respect to transmission rates, can be found in [37].

[1] Denoted as Anytime Capacity

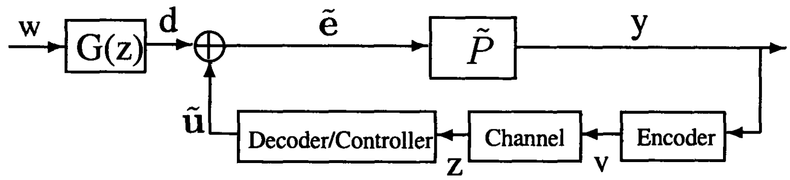

Figure 1-2: Structure of the Feedback Interconnection

1.2 Fundamental Limitations in the Presence of Finite Capacity Feedback

So far, research in this field has primarily directed its attention to stabilization. The basic framework is depicted in Fig 1-2 and comprises a plant, a channel, an encoder and a decoder, which implicitly embeds a controller. Measurements of the plant's output must be encoded and sent through the channel. The information, received at the other end of the channel, is decoded and used to generate a control signal. It has been shown that stabilization, of a linear and time-invariant plant, requires [59, 60, 50] that $C_{channel}$, the channel's Shannon capacity, is larger than $\sum \max\{0, \log(|\lambda_i(A)|)\}$, where A is the dynamic matrix of the state-space representation of the plant. For certain channels, the condition $C_{channel} > \sum \max\{0, \log(|\lambda_i(A)|)\}$ is sufficient for stabilization in the almost sure sense [60], but it may not suffice for moment stability[54]. Stabilization of nonlinear systems has also been studied by [51] and [40]. The work by [22] has used the integral of the log-sensitivity, as seen by the noise in an additive Gaussian channel, to establish that the optimal encoding/decoding scheme can be constructed using standard optimal control theory. Another recent area of investigation is the analysis in the presence of disturbances and uncertainty (Chapter 2).

Understanding the fundamental limitations of performance in a feedback system is critical for effective control design. One of the most well known trade-offs is the water-bed effect for linear feedback systems, which results from Bode's integral formula[13]. In such classical theory, the transfer function, between the disturbance $\mathbf{d}$ and $\tilde{\mathbf{e}} = \tilde{\mathbf{u}} + \mathbf{d}$ (see Fig 1-2), is denoted as sensitivity and is represented by $S(z)$. Bode's result, for a strictly proper loop gain, is expressed as:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}[\log|S(e^{j\omega})|]_{-}d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi}[\log|S(e^{j\omega})|]_{+}d\omega = \sum \max\{0, \log(|\lambda_i(A)|)\} \quad (1.1)$$

where $[\log|S(e^{j\omega})|]_{-} = \min\{0, \log|S(e^{j\omega})|\}$ and $[\log|S(e^{j\omega})|]_{+} = \max\{0, \log|S(e^{j\omega})|\}$. It implies that sensitivity can't be small at all frequencies,i.e., reduction of $\int_{-\pi}^{\pi}[\log|S(e^{j\omega})|]_{-}d\omega$ is achieved at the expense of increase in $\int_{-\pi}^{\pi}[\log|S(e^{j\omega})|]_{+}d\omega$.

1.2.1 Main Contributions of Chapter 3

Recent publications [27, 71] have provided new versions[2] of (1.1). The work by [71] has introduced a Bode-like integral inequality for non-linear systems, which is derived based on information theoretic principles.

In Chapter 3, we derive a fundamental limitation involving the directed information rate[3] [47, 60] at the channel, denoted by $\bar{I}_{\infty}(\mathbf{v} \to \mathbf{z})$. Our results show that the following must hold:

$$\frac{1}{2}L_{-} + \bar{I}_{\infty}(\mathbf{v} \to \mathbf{z}) \geq \sum \max\{0, \log(|\lambda_i(A)|)\}$$

where L_{-} is a measure of disturbance rejection. Such measure satisfies $L_{-} \leq 0$, where $L_{-} = 0$ means no-rejection and small L_{-} attests disturbance attenuation. We show that, under stationarity assumptions, L_{-}becomes an integral and our condition can be expressed as:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}[\log|S(e^{j\omega})|]_{-}d\omega + \bar{I}_{\infty}(\mathbf{v} \to \mathbf{z}) \geq \sum \max\{0, \log(|\lambda_i(A)|)\} \quad (1.2)$$

By means of an argument similar to the water-bed effect, the inequality (1.2) asserts that attenuation, when measured by $\int_{-\pi}^{\pi}[\log|S(e^{j\omega})|]_{-}d\omega$, has to be repaid by a higher information rate in the channel. Since $\bar{I}_{\infty}(\mathbf{v} \to \mathbf{z}) \leq C_{channel}$, we infer that the trade-off (1.2) creates a fundamental limitation.

Using information theoretic arguments and assuming stationarity, we also derive the Bode integral formula. Our derivations require a linear and time-invariant plant, but the

[2]Recently John Doyle (CALTECH) and Jorge Gonçalves have also derived a version of Bode's fundamental limitations for general feedback systems, by means of the properties of the Fourier series

[3]This quantity is represented as $\bar{I}_{\infty}(\mathbf{v} \to \mathbf{z})$ and will be precisely defined in section 3.2.

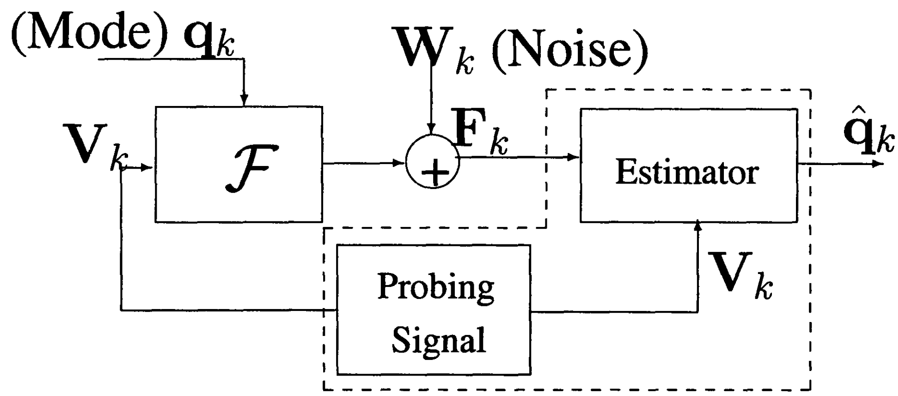

Figure 1-3: Diagram of the Estimation Setup

encoder, the channel and the decoder/controller can be any causal operators.

1.3 Estimation of a Class of Hidden Markov Models

Consider the estimation setup depicted in figure 1-3. That is the archetypal scenario where an external observer intends to estimate $\mathbf{q}_k$, the mode of operation of the system $\mathcal{F}$. The estimator generates, or has access to, a probing signal $\mathbf{V}_k$ which is used to stimulate the system so that the noisy observations of the output are informative. In this framework $\mathbf{V}_k$ is i.i.d., zero mean and Gaussian with covariance matrix Σ_V, which is considered one of the design parameters. Additionally, we assume that $\mathcal{F}$ is known. The reason to adopt $\mathbf{V}_k$ i.i.d. is to eliminate the possibility that the agent, represented by $\mathcal{F}$, can predict the probing signal.

The aforementioned scheme corresponds to problem 2, according to the classification in [53], but it can also be designated, in the context of Hybrid Systems [33], as *mode estimation*. In [53], [4] and [68, 33] one finds applications that illustrate the flexibility of this framework and suggest why its employment is so vast. Besides failure detection, other usances are the mode estimation of systems in adversarial environments, where such information might be critical in the prediction of attack maneuvers.

1.3.1 Problem Statement

Let $\mathbb{A} = \{1, \ldots, M\}$ be a finite set, which we denote as alphabet, and $\mathbf{q}_k$ be a Markovian process, which we designate as mode sequence, taking values in $\mathbb{A}$. Consider also a function

$$\mathcal{F} : \left(\mathbb{R}^{n^V}\right)^{\alpha} \times \mathbb{A} \rightarrow \mathbb{R}^{n^F}$$

which is linear in $\left(\mathbb{R}^{n^V}\right)^{\alpha}$. The function $\mathcal{F}$ is used to describe the input-output behavior of the system in Figure 1-3 according to the following Hidden Markov Model (HMM):

$$\mathbf{F}_k = \mathcal{F}(\mathbf{V}_k, \cdots, \mathbf{V}_{k-\alpha+1}, \mathbf{q}_k) + \mathbf{W}_k \tag{1.3}$$

where $\mathbf{V}_k \in \mathbb{R}^{n^V}$ and $\mathbf{W}_k \in \mathbb{R}^{n^F}$ are independent and identically distributed Gaussian random processes that represent the input and measurement noise, respectively. We select this finite impulse response (FIR) structure because it can be used as an approximation for a broad class of stable systems, while providing a simple analysis framework. By adopting this setup, we avoid the need to estimate the continuous state that would be necessary in the formulation adopted by [33]. That allows us to concentrate on the mode estimation alone as suggested by [58]. In physical applications, $\mathcal{F}$ is a passive element that reacts to the probing signal. Typical examples are when $\mathcal{F}$ represents the reflection of an acoustic or electromagnetic probing signal. The changes in the reflection characteristics are a function of the mode and are represented by the function $\mathcal{F}$.

If the transition probabilities of $\mathbf{q}_k$ are available, then we can adopt a Maximum A Posteriori (MAP) estimator. For each k, such estimator leads to the minimal probability of error $\mathcal{P}\left((\mathbf{q}_1, \ldots, \mathbf{q}_k) \neq (\hat{\mathbf{q}}_1, \ldots, \hat{\mathbf{q}}_k)\right)$. On the other hand, since such quantity goes to one, as k tends to infinity, we conclude that it is not useful as a measure of estimation fidelity.

Another approach is to minimize the probability of error $\mathcal{P}\left(\mathbf{q}_k \neq \hat{\mathbf{q}}_k\right)$ at every step k. If the transition probabilities are provided then this method can be efficiently implemented. In [24], the estimator equations, for such paradigm, are elegantly derived by means of a change of measure. A comprehensive discussion of the choice of fidelity measures can also be found in [53].

In previous publications, the parameter estimation of time-varying systems was also studied in a deterministic framework. They corroborate the intuitive idea that, for a fixed signal to noise ratio, *slow varying* systems can be estimated more accurately. More specifically, [70] shows that time-variation generates uncertainty in a robust estimation sense. The natural extension to the present framework would be the use of $\mathcal{P}(\mathbf{q}_k \neq \hat{\mathbf{q}}_k)$ as a measure of estimation fidelity. But, the reality is that the determination of the probability of error, in the presence of sources with memory, is very difficult. A common approach is to use Monte Carlo simulations [12]. The aforementioned way of determining estimation fidelity does not allow a proper study of the influence of the probing signal, nor it reveals how the statistic properties of $\mathbf{q}_k$ and $\mathbf{V}_k$ affect fidelity.

Along these lines, we would like to address the following questions:

- Decide for a meaningful measure of distortion and derive a worst case analysis for the achievable estimation fidelity. Such result must reflect the memory of $\mathbf{q}_k$. Intuitively, one should expect that if $\mathbf{q}_k$ is *almost periodic* (deterministic) then high fidelity can be achieved even when the signal to noise ratio is small. That should be, mutatis mutandis, analogous to the result where [70] relates the degree of time variation with the magnitude of estimation uncertainty. In addition, we want to determine what is the essential feature one must know about the source in order to provide such analysis.

- Investigate how the probing signal affects estimation fidelity, when evaluated in terms of distortion measures which are associated to the frequency of error.

1.3.2 Preview of the Main Contributions

In Chapter 4, we study a measure of distortion $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$, which is related to the concept of divergence as in [3]. The distortion $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$ has several desirable properties which are explored in [43]. In particular, $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$ establishes a topology in the space of sequences. The covariance matrix of the probing signal is one of the parameters that define $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$ and, as such, it will shape the topology. In particular, scaling up the covariance matrix of $\mathbf{V}_k$ leads to a finer topology.

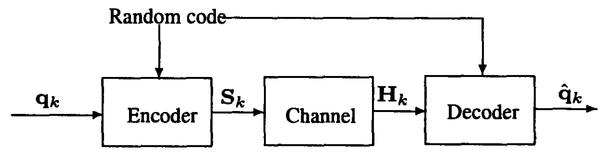

Figure 1-4: Framework for a communication system

An essential feature of $\mathcal{D}_d$ is the one expressed in the following theorem, proven in chapter 4. We also stress that the proof of such result holds even if the transition probabilities of $\mathbf{q}_k$ are not known.

Let r^q be the entropy rate [20] of the process $\mathbf{q}_k$, given by:

$$r^q = \mathcal{E}[-\ln(p^q(\mathbf{q}_k|\mathbf{q}_{k-1}))] \tag{1.4}$$

Theorem 1.3.1 *Let $\hat{\mathbf{q}}_{1,k}$ be determined according to the decoding process described in the definition 4.3.4. For any given $\delta > 0$, there exists $k \in \mathbb{N}$ and $\epsilon \in (0,1)$ such that:*

$$\mathcal{P}\left(\mathcal{D}_d(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \ln(m)r^q + \frac{n^F}{2} + \delta\right) < \delta, \ \textit{if}\, r^q \geq 0 \tag{1.5}$$

$$\mathcal{P}\left(\mathcal{D}_d(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \delta\right) < \delta, \ \textit{if}\, r^q = 0 \tag{1.6}$$

The worst case analysis quantified in Theorem 1.3.1 enables the use of $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$ to answer the following question: **Is it possible to reliably communicate[4] through the sequence $\mathbf{q}_k$?** From Theorem 1.3.1 we find that the answer is yes, provided that we can partition the typical sets of $\mathbf{q}_{1,k}$ with balls of size $2\ln(m)r^q + n^F$, as measured by $\mathcal{D}_d$.

1.3.3 Comparative Analysis with Previous Work

In Chapter 4 we introduce a suitable measure of distortion and derive theorem 1.3.1. We provide a unified probabilistic worst case analysis that analytically unveils the importance of the covariance matrix of $\mathbf{V}_k$ as well as the entropy rate of $\mathbf{q}_k$.

[4]In a recent conversation with Anant Sahai, he suggested that this question could be phrased as:"Can $\mathcal{F}$ transmit information through its behavior, as represented by the sequence $\mathbf{q}_k$?". Anant Sahai would call that a Poltergeist channel to emphasize the fact that $\mathcal{F}$ is communicating as a passive element.

There is an extensive list of articles in Signal Processing, Automatic Control and Information Theory that broach the estimation and tracking of HMM. Although these communities are, with a few exceptions, self-referential, the disconnect is artificial. There exists a great deal of intersection in the objectives and motivations. In common they lack a suitable analysis of performance and that is what distinguishes our approach. We briefly discuss the relevant publications according to the following classification:

- **Information Theoretic Approaches to the Estimation and tracking of HMM** The similarities, between the estimation setup in Fig 1-3 and the communication scheme in Fig 1-4, suggest that the study of estimation fidelity, in the present context, is related to Rate-Distortion theory [7]. This view has already been explored in the area of pattern recognition, as a way to study the effects of quantization [68]. The same problem was broached in [36] for general HMM with discrete measurements. In the aforementioned publication, the rate-distortion tradeoff has to be determined point-wise by numerically solving two non-linear equations. The use of numerical methods in the determination of the Rate-Distortion tradeoff is a common practice since its introduction by [11]. The use of entropy-based distortion measures and its relation to the probability of error has also received attention more recently in [25]. When compared to these publications, our work makes a contribution for the constrained coding established by the linear system in Fig 1-3. By suggesting how the input shapes the distortion function, we also provide a unified framework to study the influence of the probing signal. The design of probing signals was studied in a similar framework, for the time-invariant case, by [48].

- **Identification of Time-Varying Systems** Several publications have addressed the trade-off between tracking and disturbance rejection [29] or, alternatively, the trade-off between the *rate* of time-variation and uncertainty [70, 6]. The latter one is the deterministic analogous of the rate-distortion tradeoff. Also, in the scope of the design of probing signals, [10] refers that the state of the art is not satisfactory. A good review of past results can also be found in [17], where it becomes apparent the lack a unified framework as well as the use of assumptions such as *slow variation*.

- **Design of sub-optimal estimators for Hybrid Linear Systems** Since the first contributions came out [18, 1] several new algorithms have attempted to tackle the exponential complexity of the hypothesis set. A collection of the main results can be found in [62, 4, 24, 57]. These results rely on approximations that make the study of estimation fidelity very difficult [12] and [58] suggests that more attention should be devoted to the mode estimation problem.

Chapter 2

Feedback Stabilization of Uncertain Stochastic Systems Using a Stochastic Digital Link

2.1 Introduction

In order to focus on the fundamental issues and keep clarity, we start by deriving our results for first order linear systems. Subsequently, we provide an extension to a class of multi-state linear systems. As pointed out in [49], non-commutativity creates difficulties in the study of arbitrary time-varying stochastic systems. Results for the fully-observed Markovian case over finite alphabets, in the presence of a deterministic link, can be found in [49].

Besides the introduction, this Chapter has 5 sections: section 2.2 comprises the problem formulation and preliminary definitions; in section 2.3 we prove sufficiency conditions by constructing a stabilizing feedback scheme for first order systems; section 2.4 contains the proof of the necessary condition for stability; some of the quanties, introduced in the paper, are given a detailed interpretation in section 2.5 and section 2.6 extends the sufficient conditions to a class of multi-state linear systems.

The following notation is adopted:

- Whenever that is clear from the context we refer to a sequence of real numbers $x(k)$

simply as x. In such cases we may add that $x \in \mathbb{R}^\infty$.

- Random variables are represented using boldface letters, such as $\mathbf{w}$
- if $\mathbf{w}(k)$ is a stochastic process, then we use $w(k)$ to indicate a specific realization. According to the convention used for sequences, we may denote $\mathbf{w}(k)$ just as $\mathbf{w}$ and $w(k)$ as w.
- the expectation operator over $\mathbf{w}$ is written as $\mathcal{E}[\mathbf{w}]$
- if E is a probabilistic event, then its probability is indicated as $\mathcal{P}(E)$
- we write $\log_2(.)$ simply as $\log(.)$
- if $x \in \mathbb{R}^\infty$, then

$$\|x\|_1 = \sum_{i=0}^{\infty} |x(i)|$$

$$\|x\|_\infty = \sup_{i \in \mathbb{N}} |x(i)|$$

Definition 2.1.1 *Let $\varrho \in \mathbb{N}_+ \bigcup \{\infty\}$ be an upper-bound for the memory horizon of an operator. If $G_f : \mathbb{R}^\infty \to \mathbb{R}^\infty$ is a causal operator then we define $\|G_f\|_{\infty(\varrho)}$ as:*

$$\|G_f\|_{\infty(\varrho)} = \sup_{k \geq 0, x \neq 0} \frac{|G_f(x)(j)|}{\max_{j \in \{k-\varrho+1,\dots,k\}} |x(j)|} \tag{2.1}$$

Note that, since G_f is causal, $\|G_f\|_{\infty(\infty)}$ is just the infinity induced norm of G_f:

$$\|G_f\|_{\infty(\infty)} = \|G_f\|_\infty = sup_{x \neq 0} \frac{\|G_f(x)\|_\infty}{\|x\|_\infty}$$

2.2 Problem Formulation

We study the stabilizability of uncertain stochastic systems under communication constraints. Motivated by the type of constraints that arise in most computer networks, we consider the following class of stochastic links:

Definition 2.2.1 *(Stochastic Link) Consider a link that, at every instant k, transmits* $\mathbf{r}(k)$ *bits. We define it to be a stochastic link, provided that* $\mathbf{r}(k) \in \{0, \dots, \bar{r}\}$ *is an independent and identically distributed (i.i.d.) random process satisfying:*

$$\mathbf{r}(k) = C - \mathbf{r}^{\delta}(k) \tag{2.2}$$

where $\mathcal{E}[\mathbf{r}^{\delta}(k)] = 0$ *and* $C \geq 0$*. The term* $\mathbf{r}^{\delta}(k)$ *represents a fluctuation in the transfer rate of the link.*

More specifically, the link is a stochastic truncation operator $\mathcal{F}_k^l : \{0,1\}^{\bar{r}} \to \bigcup_{i=0}^{\bar{r}} \{0,1\}^i$ *defined as:*

$$\mathcal{F}_k^l(b_1, \dots, b_{\bar{r}}) = (b_1, \dots, b_{\mathbf{r}(k)}) \tag{2.3}$$

where $b_i \in \{0,1\}$*.*

Given $x(0) \in [-\frac{1}{2}, \frac{1}{2}]$ and $\bar{d} \geq 0$, we consider nominal systems of the form:

$$\mathbf{x}(k+1) = \mathbf{a}(k)\mathbf{x}(k) + \mathbf{u}(k) + \mathbf{d}(k) \tag{2.4}$$

with $|\mathbf{d}(k)| \leq \bar{d}$ and $\mathbf{x}(i) = 0$ for $i < 0$.

2.2.1 Description of Uncertainty in the Plant

Let $\varrho \in \mathbb{N}_+ \bigcup \{\infty\}$, $\bar{z}_f \in [0,1)$ and $\bar{z}_a \in [0,1)$ be given constants, along with the stochastic process $\mathbf{z}_a$ and the operator $G_f : \mathbb{R}^{\infty} \to \mathbb{R}^{\infty}$ satisfying:

$$|\mathbf{z}_a(k)| \leq \bar{z}_a \tag{2.5}$$

$$G_f \text{ causal and } \|G_f\|_{\infty(\varrho)} \leq \bar{z}_f \tag{2.6}$$

Given $x(0) \in [-\frac{1}{2}, \frac{1}{2}]$ and $\bar{d} \geq 0$, we study the existence of stabilizing feedback schemes for the following perturbed plant (see Figure 1):

$$\mathbf{x}(k+1) = \mathbf{a}(k)\left(1 + \mathbf{z}_a(k)\right)\mathbf{x}(k) + \mathbf{u}(k) + G_f(\mathbf{x})(k) + \mathbf{d}(k) \tag{2.7}$$

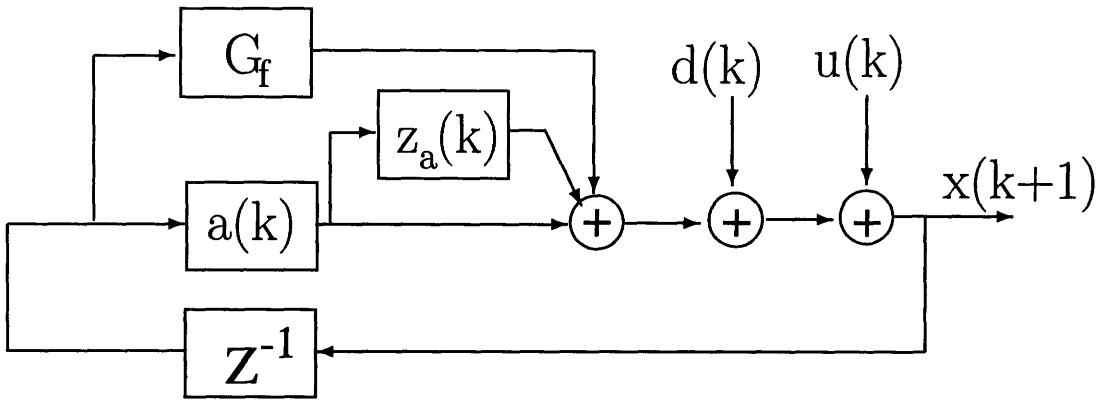

Figure 2-1: Diagram of the Plant with Model Uncertainty

where the perturbation processes $\mathbf{z}_a$ and $G_f(\mathbf{x})$ satisfy (2.5)-(2.6). Notice that $\mathbf{z}_a(k)$ may represent uncertainty in the knowledge of $\mathbf{a}(k)$, while $G_f(\mathbf{x})(k)$ is the output of the feedback uncertainty block G_f. We chose this structure because it allows the representation of a wide class of model uncertainty. It is also allows the construction of a suitable stabilizing scheme.

Example 2.2.1 *If* $G_f(\mathbf{x})(k) = \mu_0\mathbf{x}(k) + \ldots + \mu_{n-1}\mathbf{x}(k-n+1)$ *then* $\|G_f\|_{\infty(\varrho)} = \sum |\mu_i|$ *for* $\varrho \geq n$.

In general, the operator G_f may be nonlinear and time-varying.

2.2.2 Statistical Description of $\mathbf{a}(k)$

The process $\mathbf{a}(k)$ is i.i.d. and independent of $\mathbf{r}(k)$ and $x(0)$, meaning that it carries no information about the link nor the initial state. In addition, for convenience, we use the same representation as in (2.2) and write:

$$\log(|\mathbf{a}(k)|) = \mathcal{R} + \mathbf{l}_a^\delta(k) \tag{2.8}$$

where $\mathcal{E}[\mathbf{l}_a^\delta(k)] = 0$. Notice that $\mathbf{l}_a^\delta(k)$ is responsible for the stochastic behavior, if any, of the plant. Since $\mathbf{a}(k)$ is ergodic, we also assume that $\mathcal{P}\left(a(k) = 0\right) = 0$, otherwise the system is trivially stable. Such assumption is also realistic if we assume that (2.7) comes from the discretization of a continuous-time system.

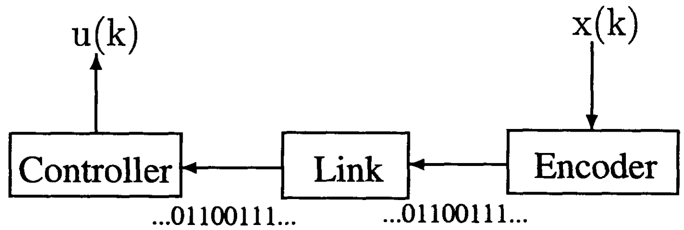

Figure 2-2: Structure of the Feedback Interconnection

2.2.3 Functional Structure of the Feedback Interconnection

We assume that the feedback loop has the structure depicted in Figure 2, also designated as information pattern [65]. The blocks denoted as encoder and controller are stochastic operators whose domain and image are uniquely determined by the diagram. At any given time k, we assume that both the encoder and the controller have access to $a(0), \ldots, a(k)$ and $r(k-1)$ as well as the constants ϱ, $\bar{z}_f$, $\bar{z}_a$ and $\bar{d}$. The encoder and the controller are described as:

- The encoder is a function $\mathcal{F}_k^e : \mathbb{R}^{k+1} \to \{0,1\}^{\bar{r}}$ that has the following dependence on observations:

$$\mathcal{F}_k^e(x(0), \ldots, \mathbf{x}(k)) = (\mathbf{b}_1, \ldots, \mathbf{b}_{\bar{r}}) \tag{2.9}$$

- The control action results from a map, not necessarily memoryless, $\mathcal{F}_k^c : \bigcup_{i=0}^{\bar{r}} \{0,1\}^i \to \mathbb{R}$ exhibiting the following functional dependence:

$$\mathbf{u}(k) = \mathcal{F}_k^c(\vec{\mathbf{b}}(k)) \tag{2.10}$$

where $\vec{b}(k)$ are the bits successfully transmitted through the link, i.e.:

$$\vec{\mathbf{b}}(k) = \mathcal{F}_k^l(\mathbf{b}_1, \ldots, \mathbf{b}_{\bar{r}}) = (\mathbf{b}_1, \ldots, \mathbf{b}_{\mathbf{r}(k)}) \tag{2.11}$$

As such, $\mathbf{u}(k)$ can be equivalently expressed as

$$\mathbf{u}(k) = (\mathcal{F}_k^c \circ \mathcal{F}_k^l \circ \mathcal{F}_k^e)(x(0), \ldots, \mathbf{x}(k))$$

Definition 2.2.2 *(**Feedback Scheme**) We define a feedback scheme as the collection of a* *controller $\mathcal{F}_k^c$ and an encoder $\mathcal{F}_k^e$.*

2.2.4 Problem Statement and M-th Moment Stability

Definition 2.2.3 ***(Worst Case Envelope)*** *Let $\mathbf{x}(k)$ be the solution to (2.7) under a given feedback scheme. Given any realization of the random variables $\mathbf{r}(k)$, $\mathbf{a}(k)$, $G_f(\mathbf{x})(k)$, $\mathbf{z}_a(k)$ and $\mathbf{d}(k)$, the worst case envelope $\bar{\mathbf{x}}(k)$ is the random variable whose realization is defined by:*

$$\bar{x}(k) = \sup_{x(0)\in[-\frac{1}{2},\frac{1}{2}]} |x(k)| \tag{2.12}$$

Consequently, $\bar{\mathbf{x}}(k)$ is the smallest envelope that contains every trajectory generated by an initial condition in the interval $x(0) \in [-\frac{1}{2}, \frac{1}{2}]$. We adopted the interval $[-\frac{1}{2}, \frac{1}{2}]$ to make the text more readable. All results are valid if it is replaced by any other symmetric bounded interval.

Our problem consists in determining necessary and sufficient conditions that guarantee the existence of a stabilizing feedback scheme. The results must be derived for the following notion of stability.

Definition 2.2.4 ***(m-th Moment Robust Stability)*** *Let $m > 0$, $\varrho \in \mathbb{N}_+ \bigcup \{\infty\}$, $\bar{z}_f \in [0,1)$, $\bar{z}_a \in [0,1)$ and $\bar{d} \geq 0$ be given. The system (2.7), under a given feedback scheme, is m-th moment (robustly) stable provided that the following holds:*

$$\begin{cases} \lim_{k\to\infty} \mathcal{E}\left[\bar{\mathbf{x}}(k)^m\right] = 0 & \text{if } \bar{z}_f = \bar{d} = 0 \\ \exists b \in \mathbb{R}_+ \text{ s.t. } \limsup_{k\to\infty} \mathcal{E}\left[\bar{\mathbf{x}}(k)^m\right] < b & \text{otherwise} \end{cases} \tag{2.13}$$

The first limit in (2.13) is an internal stability condition while the second establishes external stability. The constant b must be such that $\limsup_{k\to\infty} \mathcal{E}\left[\bar{\mathbf{x}}(k)^m\right] < b$ holds for all allowable perturbations $\mathbf{z}_a$ and $G_f(\mathbf{x})$ satisfying (2.5)-(2.6).

2.2.5 Motivation for our Definition of Stochastic Link and Further Comments on the Information Pattern

Consider that we want to use a wireless medium to transmit information between nodes A and B. In our formulation, node A represents a central station which measures the state of the plant. The goal of the transmission system is to send information, about the state, from node A to node B, which represents a controller that has access to the plant. Notice that node A maybe a communication center which may communicate to several other nodes, but we assume that node B only communicates with node A. Accordingly, we will concentrate on the communication problem between nodes A and B only, without loss of generality.

Definition 2.2.5 *(Basic Communication Scheme) We assume an external synchronization variable, denoted as k. The interval between k and k+1 is of T seconds, of which $T_T < T$ is reserved for transmission. We also denote the number of bits in each packet as Π, excluding headers. In order to submit an ordered set of packets for transmission, we consider the following basic communication protocol, at the media access control level:*

(Initialization) *A variable denoted by $c(k)$ is used to count how many packets are sent in the interval $t \in [kT, kT+T_T]$. We consider yet another counter p, which is used to count the number of periods for which no packet is sent. Such variable is initialized as $p = 0$.*

(For node A)

(Synchronization) *If k changes to $k := k + 1$ then step 1 is activated.*

- ***Step1*** *The packets to be submitted for transmission are numbered according to the order of priority; 0 is the highest priority. The order of each packet is included in the header of the packet. The first packet has an extra header, comprising the pair $(c(k-p-1), p)$. The variable $c(k)$ is initialized to $c(k) = 0$ and p is incremented to $p := p + 1$. Move to step 2.*

- ***Step 2:*** *Stands by until it can send the packet $c(k)$. If such opportunity occurs, move to step 3.*

- ***Step 3:*** *Node A sends the packet $c(k)$ to node B and waits for an ACK signal from node B. If node B receives an ACK signal then $c(k) := c(k) + 1$, $p = 0$ and move* *back to step 2. If time-out then go back to Step 2.*

The time-out *decision may be derived from several events: a fixed waiting time; a random timer (CSMA/CA) or a new opportunity to send a packet (CSMA).*

(For node B)

- ***Step 1:*** *Node B stands by until it receives a packet from node A. Once a packet is received, check if it is a first packet: if so, extract* $(\mathbf{c}(k-p-1), p)$ *and construct* $\mathbf{r}_d(i)$, *with* $i \in \{k-p-1, \ldots, k-1\}$, *according to:*

$$\begin{cases} \mathbf{r}_d(k-p-1) = \mathbf{c}(k-p-1)\Pi, \ \mathbf{r}_d(i) = 0, i \in \{k-p, \ldots, k-1\} & \text{if } p \geq 1 \\ \mathbf{r}_d(k-1) = \mathbf{c}(k-1)\Pi & \text{otherwise} \end{cases}$$

  where Π *is the size of the packets, excluding the header. If the packet is not duplicated then make the packet available to the controller. Move to step 2.*

- ***Step 2:*** *Wait until it can send an ACK signal to node A. Once ACK is sent, move to step 1.*

The scheme of definition 2.2.5 is the simplest version of a class of media access control (MAC) protocols, denoted as *CSMA*. A recent discussion and source of references about (CSMA) is [32]. Such scheme also describes the MAC operation for a wireless communication network between two nodes. Also, we adopt the following strong assumptions:

- Every time node A sends a packet to node B: either it is sent without error or it is lost. This assumption means that we are not dealing with, what is commonly referred to as, a *noisy channel.*

- Every ACK will go through before k changes to $k+1$. This assumption is critical to guarantee that no packets are wasted. Notice that node B can use the whole interval $t \in (kT+T_T, (k+1)T)$ to send the last ACK. During this period, the controller is not expecting new packets. The controller will generate $\mathbf{u}(k)$ using the packets that were sent in the interval $t \in [kT, kT+T_T]$. Consequently, such ACK is not important in the generation of $\mathbf{u}(k)$. It will be critical only for $\mathbf{u}(i)$ for $i > k$.

We adopt k, the discrete-time unit, as a reference. According to the usual framework of digital control, k will correspond to the discrete time unit obtained by partitioning the continuous time in periods of duration T. Denote by $T_T < T$ the period allocated for transmission. Now, consider that the aim of a discrete-time controller is to control a continuous-time linear system, which admits[1] a discretization of the form $\mathbf{x}_c((k+1)T) = \mathbf{A}(k)\mathbf{x}_c(kT) + \mathbf{u}(k)$. The discretization is such that $u(k)$ represents the effect of the control action over $t \in (kT + T_T, (k+1)T)$. What is left is reserved for the transmission of information about $\mathbf{x}(k) = \mathbf{x}_c(kT)$, the state of the plant at the sampling instant $t = kT$. Whenever k changes, we construct a new queue and assume that the cycle of definition 2.2.5 is re-initialized.

We denote by $\mathbf{r}(k)$ the random variable which represents the total number of bits that are transmitted in the time interval $t \in [kT, kT + T_T]$. The $\mathbf{r}(k)$ transmitted bits are used by the controller to generate $\mathbf{u}(k)$. Notice that our scheme does not pressupose an extra delay, because the control action will act, in continuous time, in the interval $t \in (kT + T_T, (k+1)T)$.

Synchronization between the encoder and the decoder

Denote by $\mathbf{r}_e(k)$ the total number of bits that the encoder has successfully sent between k and $k+1$, i.e., the number of bits for which the encoder has received an ACK. The variable $\mathbf{r}_e(k)$ is used by the encoder to keep track of how many bits were sent. The corresponding variable at the decoder is represented as $\mathbf{r}_d(k)$. From definition 2.2.5 we infer that $\mathbf{r}_d(k-1)$ may not be available at all times. On the other hand, we emphasize that the following holds:

$$\mathbf{c}(k) \neq 0 \implies \mathbf{r}_d(i) = \mathbf{r}_e(i) \text{ for } i \in \{0, \ldots, k-1\} \tag{2.14}$$

In section 2.3, the stabilizing control is constructed in a way that: if no packet goes through between k and $k+1$, i.e., $c(k) = 0$ then $\mathbf{u}(k) = 0$. That shows that $\mathbf{r}_d(k-1)$ is not available only when it is not needed. That motivated us to adopt the simplifying assumption that $\mathbf{r}(k-1) = \mathbf{r}_e(k-1) = \mathbf{r}_d(k-1)$.

[1] A controllable linear and time-invariant system admits a discretization of the required form. If the system is stochastic an equivalent condition has to be imposed

Encoding and Decoding for First Order Systems

Given the transmission scheme described above, the only remaining degrees of freedom are how to encode the measurement of the state and how to construct the queue. From the proofs of theorems 2.4.1, 2.3.2 and 2.3.4, we infer that a necessary and sufficient condition for stabilization is the ability to transmit, between nodes A and B, an estimate of the state $\hat{\mathbf{x}}(k)$ with an accuracy lower-bounded[2] by $\mathcal{E}[|\hat{\mathbf{x}}(k) - \mathbf{x}(k)|^m] < 2^{-R}$, where $R > 0$ is a given constant that depends on the state-space representation of the plant. Since the received packets preserve the original order of the queue, we infer that the *best* way to construct the queues, at each k, is to compute the binary expansion of $\mathbf{x}(k)$ and position the packets so that the bits corresponding to higher powers of 2 are sent first. The lost packets will always[3] be the *less important*. The abstraction of such procedure is given by the truncation operator of definition 2.2.1. The random behavior of $\mathbf{r}(k)$ arises from random time-out, the existence of collisions generated by other nodes trying to communicate with node A or from the fading that occurs if node B is moving. The fading fenomena may also occur from interference.

2.3 Sufficiency Conditions for the Robust Stabilization of First Order Linear Systems

In this section, we derive constructive sufficient conditions for the existence of a stabilizing feedback scheme. We start with the deterministic case in subsection 2.3.1, while 2.3.2 deals with random $\mathbf{r}$ and $\mathbf{a}$. We stress that our proofs hold under the framework of section 2.2. The strength of our assumptions can be accessed from the discussion in section 2.2.5.

The following definition introduces the main idea behind the construction of a stabilizing feedback scheme.

Definition 2.3.1 *(Upper-bound Sequence) Let* $\bar{z}_f \in [0,1)$, $\bar{z}_a \in [0,1)$, $\bar{d} \geq 0$ *and* $\varrho \in$

[2]This observation was already reported in [60]

[3]The situation were the packets lost are in random positions is characteristic of large networks where packets travel through different routers.

$\mathbb{N}_+ \bigcup \{\infty\}$ *be given. Define the upper-bound sequence as:*

$$\mathbf{v}(k+1) = |\mathbf{a}(k)| 2^{-\mathbf{r}_e(k)} \mathbf{v}(k) + \bar{z}_f \max\{\mathbf{v}(k-\varrho+1), \ldots, \mathbf{v}(k)\} + \bar{d}, \tag{2.15}$$

where $v(i) = 0$ for $i < 0$, $v(0) = \frac{1}{2}$ and $\mathbf{r}_e(k)$ is an effective rate given by:

$$\mathbf{r}_e(k) = -\log(2^{-\mathbf{r}(k)} + \bar{z}_a) \tag{2.16}$$

Definition 2.3.2 *Following the representation for $\mathbf{r}(k)$ we also define C_e and $\mathbf{r}_e^\delta(k)$ such that:*

$$\mathbf{r}_e(k) = C_e - \mathbf{r}_e^\delta(k) \tag{2.17}$$

where $\mathcal{E}[\mathbf{r}_e^\delta(k)] = 0$.

We adopt $v(0) = \frac{1}{2}$ to guarantee that $|x(0)| \leq v(0)$. If $x(0) = 0$ then we can select $v(0) = 0$. Notice that the multiplicative uncertainty $\bar{z}_a$ acts by reducing the effective rate $\mathbf{r}_e(k)$. After inspecting (2.16), we find that $\mathbf{r}_e(k) \leq \min\{\mathbf{r}(k), -\log(\bar{z}_a)\}$. Also, notice that:

$$\bar{z}_a = 0 \implies \mathbf{r}_e(k) = \mathbf{r}(k), \mathbf{r}_e^\delta(k) = \mathbf{r}^\delta(k) \text{ and } C = C_e \tag{2.18}$$

Definition 2.3.3 ***(Stabilizing feedback scheme)*** *We make use of the sequence specified in definition 2.3.1. Notice that $\mathbf{v}(k)$ can be constructed at the controller and the encoder because both have access to ϱ, $\bar{z}_f$, $\bar{z}_a$, $\bar{d}$, $\mathbf{r}(k-1)$ and $\mathbf{a}(k-1)$.*

The feedback scheme is defined as:

- ***Encoder:*** *Measures $x(k)$ and computes $b_i \in \{0, 1\}$ such that:*

$$(b_1, \ldots, b_{\bar{r}}) = arg \max_{\sum_{i=1}^{\bar{r}} b_i \frac{1}{2^i} \leq \left(\frac{\mathbf{x}(k)}{2\mathbf{v}(k)} + \frac{1}{2}\right)} \sum_{i=1}^{\bar{r}} b_i \frac{1}{2^i} \tag{2.19}$$

  Place $(b_1, \ldots, b_{\bar{r}})$ for transmission. For any $r(k) \in \{0, \ldots, \bar{r}\}$, the above construction provides the following centroid approximation $\hat{x}(k)$ for $x(k) \in [-v(k), v(k)]$:

$$\hat{x}(k) = 2v(k)\left(\sum_{i=1}^{r(k)} b_i \frac{1}{2^i} + \frac{1}{2^{r(k)+1}} - \frac{1}{2}\right) \tag{2.20}$$

which satisfies $|x(k) - \hat{x}(k)| \leq 2^{-r(k)}v(k)$.

- ***Controller:*** *From the* $\bar{r}$ *bits placed for transmission in the stochastic link, only* $\mathbf{r}(k)$ *bits go through. Compute* $\mathbf{u}(k)$ *as:*

$$\mathbf{u}(k) = -\mathbf{a}(k)\hat{\mathbf{x}}(k) \tag{2.21}$$

As expected, the transmission of state information through a finite capacity medium requires quantization. The encoding scheme of definition 2.3.3 is not an exception and is structurally identical to the ones used by [14, 59], where sequences were already used to upper-bound the state of the plant.

The following lemma suggests that, in the construction of stabilizing controllers, we may choose to focus on the dynamics of the sequence $v(k)$. That simplifies the analysis in the presence of uncertainty because the dynamics of $v(k)$ is described by a first-order difference equation.

Lemma 2.3.1 *Let* $\bar{z}_f \in [0,1)$, $\bar{z}_a \in [0,1)$ *and* $\bar{d} \geq 0$ *be given. If* $\mathbf{x}(k)$ *is the solution of (2.7) under the feedback scheme of definition 2.3.3, then the following holds:*

$$\bar{\mathbf{x}}(k) \leq \mathbf{v}(k)$$

for all $\varrho \in \mathbb{N}_+ \bigcup \{\infty\}$, *every choice* $G_f \in \Delta_{f,\varrho}$ *and* $|\mathbf{z_a}(k)| \leq \bar{z}_a$, *where*

$$\Delta_{f,\varrho} = \{G_f : \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty : \|G_f\|_{\infty(\varrho)} \leq \bar{z}_f\} \tag{2.22}$$

Proof: We proceed by induction, assuming that $\bar{x}(i) \leq v(i)$ for $i \in \{0, \dots, k\}$ and proving that $\bar{x}(k+1) \leq v(k+1)$. From (2.7), we get:

$$|x(k+1)| \leq |a(k)||x(k) + \frac{u(k)}{a(k)}| + |z_a(k)||a(k)||x(k)| + |G_f(x)(k)| + |d(k)| \tag{2.23}$$

The way the encoder constructs the binary expansion of the state, as well as (2.21), allow us to conclude that

$$|x(k) + \frac{u(k)}{a(k)}| \leq 2^{-r(k)}v(k)$$

Now we recall that $|z_a(k)| \leq \bar{z}_a$, $|G_f(x)(k)| \leq \bar{z}_f \max\{v(k-\varrho+1), \ldots, v(k)\}$ and that $|d(k)| \leq \bar{d}$, so that (2.23) implies:

$$|x(k+1)| \leq |a(k)|(2^{-r(k)} + \bar{z}_a)v(k) + \bar{z}_f \max\{v(k-\varrho+1), \ldots, \mathrm{v}(k)\} + \bar{d} \quad (2.24)$$

The proof is concluded once we realize that $|x(0)| \leq v(0)$.

□

2.3.1 The Deterministic Case

We start by deriving a sufficient condition for the existence of a stabilizing feedback scheme in the deterministic case, i.e., $\mathrm{r}(k) = \mathcal{C}$ and $\log(|\mathrm{a}(k)|) = \mathcal{R}$. Subsequently, we move for the stochastic case where we derive a sufficient condition for stabilizability.

Theorem 2.3.2 ***(Sufficiency conditions for Robust Stability)*** *Let $\varrho \in \mathbb{N}_+ \bigcup\{\infty\}$, $\bar{z}_f \in [0, 1)$, $\bar{z}_a \in [0, 1)$ and $\bar{d} \geq 0$ be given and $h(k)$ be defined as*

$$h(k) = 2^{k(\mathcal{R}-\mathcal{C}_e)},\ k \geq 0$$

where $\mathcal{C}_e = r_e = -\log(2^{-\mathcal{C}} + \bar{z}_a)$.

Consider that $\mathrm{x}(k)$ is the solution of (2.7) under the feedback scheme of definition 2.3.3 as well as the following conditions:

- **(C 1)** $\mathcal{C}_e > \mathcal{R}$
- **(C 2)** $\bar{z}_f \|h\|_1 < 1$

If conditions **(C 1)** *and* **(C 2)** *are satisfied then the following holds for all $|\mathrm{d}(t)| \leq \bar{d}$, $G_f \in \Delta_{f,\varrho}$ and $|\mathrm{z}_a(k)| \leq \bar{z}_a$:*

$$\bar{x}(k) \leq \|h\|_1 \left(\bar{z}_f \frac{\|h\|_1 \bar{d} + \frac{1}{2}}{1 - \|h\|_1 \bar{z}_f} + \bar{d}\right) + h(k)\frac{1}{2} \quad (2.25)$$

where $\Delta_{f,\varrho}$ is given by:

$$\Delta_{f,\varrho} = \{G_f : \mathbb{R}^\infty \to \mathbb{R}^\infty : \|G_f\|_{\infty(\varrho)} \leq \bar{z}_f\} \tag{2.26}$$

Proof: From definition 2.3.1, we know that, for arbitrary $\varrho \in \mathbb{N}_+ \bigcup\{\infty\}$, the following is true:

$$v(k+1) = 2^{\mathcal{R}-\mathcal{C}_e} v(k) + \bar{z}_f \max\{v(k-\varrho+1), \ldots, v(k)\} + \bar{d} \tag{2.27}$$

Solving the difference equation gives:

$$v(k) = 2^{k(\mathcal{R}-\mathcal{C}_e)} v(0) + \sum_{i=0}^{k-1} 2^{(k-i-1)(\mathcal{R}-\mathcal{C}_e)} \left(\bar{z}_f \max\{v(i-\varrho+1), \ldots, v(i)\} + \bar{d}\right), \ k \geq 1 \tag{2.28}$$

which, using $\|\Pi_k v\|_\infty = \max\{v(0), \ldots, v(k)\}$, leads to:

$$v(k) \leq \|h\|_1 (\bar{z}_f \|\Pi_k v\|_\infty + \bar{d}) + 2^{k(\mathcal{R}-\mathcal{C}_e)} v(0) \tag{2.29}$$

But we also know that $2^{k(\mathcal{R}-\mathcal{C}_e)}$ is a decreasing function of k, so that:

$$\|\Pi_k v\|_\infty \leq \|h\|_1 (\bar{z}_f \|\Pi_k v\|_\infty + \bar{d}) + v(0) \tag{2.30}$$

which implies:

$$\|\Pi_k v\|_\infty \leq \frac{\|h\|_1 \bar{d} + v(0)}{1 - \|h\|_1 \bar{z}_f} \tag{2.31}$$

Direct substitution of (2.31) in (2.29) leads to:

$$v(k) \leq \|h\|_1 \left(\bar{z}_f \frac{\|h\|_1 \bar{d} + v(0)}{1 - \|h\|_1 \bar{z}_f} + \bar{d}\right) + 2^{k(\mathcal{R}-\mathcal{C}_e)} v(0) \tag{2.32}$$

The proof is complete once we make $v(0) = \frac{1}{2}$ and use lemma 2.3.1 to conclude that $\bar{x}(k) \leq v(k)$.

□

2.3.2 Sufficient Condition for the Stochastic Case

The following lemma provides a sequence, denoted by $v_m(k)$, which is an upper-bound for the m-th moment of $\bar{\mathbf{x}}(k)$. We show that v_m is propagated according to a first-order difference equation that is suitable for the analysis in the presence of uncertainty.

Lemma 2.3.3 *(M-th moment boundedness) Let* $\varrho \in \mathbb{N}_+$, $\bar{z}_f \in [0,1)$, $\bar{z}_a \in [0,1)$ *and* $\bar{d} \geq 0$ *be given along with the following set:*

$$\Delta_{f,\varrho} = \{G_f : \mathbb{R}^\infty \to \mathbb{R}^\infty : \|G_f\|_{\infty(\varrho)} \leq \bar{z}_f\} \tag{2.33}$$

Given m, consider the following sequence:

$$v_m(k) = h_m(k)v_m(0) + \sum_{i=0}^{k-1} h_m(k-i-1)\left(\varrho^{\frac{1}{m}}\bar{z}_f \max\{v_m(i-\varrho+1),\dots,v_m(i)\} + \bar{d}\right) \tag{2.34}$$

where $k \geq 1$, $v_m(i) = 0$ for $i < 0$, $v_m(0) = \frac{1}{2}$ and $h_m(k)$ is the impulse response given by:

$$h_m(k) = \left(\mathcal{E}[2^{m(\log(|a(k)|)-r_e(k))}]\right)^{\frac{k}{m}}, \; k \geq 0 \tag{2.35}$$

and $r_e(k) = -\log\left(2^{-r(k)} + \bar{z}_a\right)$. If $\mathbf{x}(k)$ is the solution of (2.7) under the feedback scheme of definition 2.3.3, then the following holds

$$\mathcal{E}[\bar{\mathbf{x}}(k)^m] \leq v_m(k)^m$$

for all $|\mathbf{d}(t)| \leq \bar{d}$, $G_f \in \Delta_{f,\varrho}$ and $|\mathbf{z}_a(k)| \leq \bar{z}_a$.

Proof: Since lemma 2.3.1 guarantees that $\bar{x}(k+1) \leq v(k+1)$, we only need to show that $\mathcal{E}[\mathbf{v}(k+1)^m]^{\frac{1}{m}} \leq v_m(k+1)$. Again, we proceed by induction by noticing that $v(0) = v_m(0)$ and by assuming that $\mathcal{E}[\mathbf{v}(i)^m]^{\frac{1}{m}} \leq v_m(i)$ for $i \in \{1,\dots,k\}$. The induction hypothesis is proven once we establish that $\mathcal{E}[\mathbf{v}(k+1)^m]^{\frac{1}{m}} \leq v_m(k+1)$. From definition

2.3.1, we know that:

$$\mathcal{E}[\mathbf{v}(k+1)^m]^{\frac{1}{m}} = \mathcal{E}[(2^{\log(|\mathbf{a}(k)|)-\mathbf{r}_e(k)}\mathbf{v}(k) + \bar{z}_f \max\{\mathbf{v}(k-\varrho+1),\ldots,\mathbf{v}(k)\} + \bar{d})^m]^{\frac{1}{m}} \tag{2.36}$$

Using Minkowsky's inequality [31] as well as the fact that $\mathbf{v}(i)$ is independent of $\mathbf{a}(j)$ and $\mathbf{r}_e(j)$ for $j \geq i$, we get:

$$\mathcal{E}[\mathbf{v}(k+1)^m]^{\frac{1}{m}} \leq$$
$$\mathcal{E}[2^{m(\log(|\mathbf{a}(k)|)-\mathbf{r}_e(k))}]^{\frac{1}{m}}\mathcal{E}[\mathbf{v}(k)^m]^{\frac{1}{m}} + \bar{z}_f\mathcal{E}[\max\{\mathbf{v}(k-\varrho+1),\ldots,\mathbf{v}(k)\}^m]^{\frac{1}{m}} + \bar{d} \tag{2.37}$$

which, using the inductive assumption, implies the following inequality:

$$\mathcal{E}[\mathbf{v}(k+1)^m]^{\frac{1}{m}} \leq \mathcal{E}[2^{m(\log(|\mathbf{a}(k)|)-\mathbf{r}_e(k))}]^{\frac{1}{m}}v_m(k) + \varrho^{\frac{1}{m}}\bar{z}_f \max\{v_m(k-\varrho+1),\ldots,v_m(k)\} + \bar{d} \tag{2.38}$$

where we used the fact that, for arbitrary random variables $\mathbf{s}_1,\ldots,\mathbf{s}_n$, the following holds:

$$\mathcal{E}[\max\{|\mathbf{s}_1|,\ldots,|\mathbf{s}_n|\}^m] \leq \mathcal{E}[\sum_{i=1}^{n}|\mathbf{s}_i|^m] \leq n\max\{\mathcal{E}[|\mathbf{s}_1|^m],\ldots,\mathcal{E}[|\mathbf{s}_n|^m]\}$$

The proof follows once we notice that the right hand side of (2.38) is just $v_m(k+1)$.

□

Theorem 2.3.4 ***(Sufficient Condition)*** *Let* m , $\varrho \in \mathbb{N}_+$, $\bar{z}_f \in [0,1)$, $\bar{z}_a \in [0,1)$ *and* $\bar{d} \geq 0$ *be given along with the quantities bellow:*

$$\beta(m) = \frac{1}{m}\log\mathcal{E}\left[2^{m\mathbf{l}_a^\delta(k)}\right]$$

$$\alpha_e(m) = \frac{1}{m}\log\mathcal{E}\left[2^{m\mathbf{r}_e^\delta(k)}\right]$$

$$h_m(k) = 2^{k(\mathcal{R}+\beta(m)+\alpha_e(m)-\bar{\mathcal{C}}_e)},\ k \geq 0$$

where $\mathbf{r}_e^\delta$ *comes from (2.17). Consider that* $\mathbf{x}(k)$ *is the solution of (2.7) under the feedback scheme of definition 2.3.3 as well as the following conditions:*

- **(C 3)** $C_e > \mathcal{R} + \beta(m) + \alpha_e(m)$

- **(C 4)** $\varrho^{\frac{1}{m}} \bar{z}_f \|h_m\|_1 < 1$

If conditions **(C 3)** *and* **(C 4)** *are satisfied, then the following holds for all* $|\mathbf{d}(t)| \leq \bar{d}$, $G_f \in \Delta_{f,\varrho}$ *and* $|\mathbf{z}_a(k)| \leq \bar{z}_a$*:*

$$\mathcal{E}[\bar{\mathbf{x}}(k)^m]^{\frac{1}{m}} \leq \|h_m\|_1 \left(\varrho^{\frac{1}{m}} \bar{z}_f \frac{\|h_m\|_1 \bar{d} + \frac{1}{2}}{1 - \varrho^{\frac{1}{m}} \bar{z}_f \|h_m\|_1} + \bar{d} \right) + h_m(k) \frac{1}{2} \tag{2.39}$$

where $\Delta_{f,\varrho}$ *is given by:*

$$\Delta_{f,\varrho} = \{ G_f : \mathbb{R}^\infty \to \mathbb{R}^\infty : \|G_f\|_{\infty(\varrho)} \leq \bar{z}_f \} \tag{2.40}$$

Proof: Using v_m from lemma 2.3.3, we arrive at:

$$v_m(k) \leq h_m(k) v_m(0) + \|h_m\|_1 \left(\varrho^{\frac{1}{m}} \bar{z}_f \|\Pi_k v_m\|_\infty + \bar{d} \right) \tag{2.41}$$

where we use $\|\Pi_k v_m\|_\infty = \max\{v_m(0), \ldots, v_m(k)\}$. But from (2.41), we conclude that:

$$\|\Pi_k v_m\|_\infty \leq v_m(0) + \|h_m\|_1 \left(\varrho^{\frac{1}{m}} \bar{z}_f \|\Pi_k v_m\|_\infty + \bar{d} \right) \tag{2.42}$$

or equivalently:

$$\|\Pi_k v_m\|_\infty \leq \frac{v_m(0) + \|h_m\|_1 \bar{d}}{1 - \|h_m\|_1 \varrho^{\frac{1}{m}} \bar{z}_f} \tag{2.43}$$

Substituting (2.43) in (2.41), gives:

$$v_m(k) \leq h_m(k) v_m(0) + \|h_m\|_1 \left(\varrho^{\frac{1}{m}} \bar{z}_f \frac{v_m(0) + \|h_m\|_1 \bar{d}}{1 - \|h_m\|_1 \varrho^{\frac{1}{m}} \bar{z}_f} + \bar{d} \right) \tag{2.44}$$

The proof follows from lemma 2.3.3 and by noticing that $h_m(k)$ can be rewritten as:

$$h_m(k) = \left(\mathcal{E}[2^{m(\log(|\mathbf{a}(k)|) - \mathbf{r}_e(k))} \right)^{\frac{k}{m}} = 2^{k(\mathcal{R} + \beta(m) + \alpha_e(m) - C_e)}, \ k \geq 0$$

□

2.4 Necessary Conditions for the Existence of Stabilizing Feedback Schemes

Consider that $\bar{z}_a = \bar{z}_f = \bar{d} = 0$. We derive necessary conditions for the existence of an internally stabilizing feedback scheme. We emphasize that the proofs in this section use the m-th moment stability as a stability criteria and that they are valid regardless of the encoding/decoding scheme. They follow from a counting argument[4] which is identical to the one used by [60]. Necessary conditions for stability were also studied for the Gaussian channel in [61] and for other stochastic channels in [54, 55]. A necessary condition for the almost sure stability of general stochastic channels is given by [59]. We include our treatment, because it provides necessary conditions for m-th moment stability, which are inequalities involving directly the defined quantities $\alpha(m)$ and $\beta(m)$. Such quantities are an important aid on the derivation of the conclusions presented in section 2.5. In section 2.6.8, we show that the necessary condition of Theorem 2.4.1 is not conservative.

We derive the necessary condition for the following class of state-space representations:

$$\mathbf{x}(k) = \mathbf{U}(k)\mathbf{x}(k) + B\mathbf{u}(k) \tag{2.45}$$

where $x(k) \in \mathbb{R}^n$, $u(k) \in \mathbb{R}^{n_b}$, $B \in \mathbb{R}^{n \times n_b}$ and $\mathbf{U}(k)$ is a block upper-triangular matrix of the form:

$$\mathbf{U}(k) = \begin{bmatrix} \mathbf{a}(k)\mathbf{Rot}(k) & \cdots & \cdots & \\ 0 & \mathbf{a}(k)\mathbf{Rot}(k) & \ddots & \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 0 & \mathbf{a}(k)\mathbf{Rot}(k) \end{bmatrix} \tag{2.46}$$

and $\mathbf{Rot}$ is a sequence of random rotation matrices satisfying $det(Rot(k)) = 1$. We also assume that $\mathbf{Rot}$ is independent of $\mathbf{r}$.

Theorem 2.4.1 *Let* $\mathbf{x}(k)$ *be the solution of the state-space equation (2.45) along with* $\alpha(m)$

[4] We also emphasize that this proof is different from what we had originally. The present argument was suggested by a reviewer of one of our publications

and $\beta(m)$ given by:

$$\alpha(m) = \frac{1}{m}\log(E[2^{m\mathbf{r}^{\delta}(k)}]) \tag{2.47}$$

$$\beta(m) = \frac{1}{m}\log(E[2^{m\mathbf{l}_a^{\delta}(k)}]) \tag{2.48}$$

and the norm on the vector $x(k)$ be represented as:

$$\|x(k)\|_{\infty} = \max_i |[x(k)]_i|, \tag{2.49}$$

where $[x(k)]_i$ are components of the vectors $x(k)$.

If the state satisfies the following:

$$\sup_k E[\sup_{x(0)\in[-1/2,1/2]^n} \|\mathbf{x}(k)\|_{\infty}^{m}] < \infty \tag{2.50}$$

then the following must hold:

$$C - \alpha(\frac{m}{n}) > n\beta(m) + nR \tag{2.51}$$

Proof: Consider a specific realization of $\mathbf{Rot}$, $\mathbf{r}$ and $\mathbf{a}$ along with the following sets:

$$\bar{\Omega}_k = \{a(k)Rot(k)x(0) : x(0) \in [-1/2, 1/2]^n\} \tag{2.52}$$

$$\Omega_k(u(k)) = \{a(k)Rot(k)x(0) : x(0) \in [-1/2, 1/2]^n, \mathcal{F}(x(0), k) = u(k)\} \tag{2.53}$$

where $u(k)$ is a function of $x(0)$ and k, according to $u(k) = \mathcal{F}(x(0), k)$.

Since $x(k)$ is given by (2.45) and $u(k)$ can take, at most, $2^{\sum_{i=1}^{k} r(i)}$ values, we find that:

$$\frac{Vol(\bar{\Omega}_k)}{\max_{u(k)} Vol(\Omega_k(u(k)))} \leq 2^{\sum_{i=1}^{k} r(i)} \tag{2.54}$$

Computing bounds for the volumes, we get:

$$Vol(\bar{\Omega}_k) = 2^{\sum_{i=1}^{k} \log|det(U(i))|} \tag{2.55}$$

$$Vol(\Omega_k(u(k))) \leq v^n(k) \tag{2.56}$$

where $v(k)$ is given by:

$$v(k) = 2 \sup_{x(0)\in[-1/2,1/2]^n} \|x(k)\|_\infty = 2 \sup_{x(0)\in[-1/2,1/2]^n} \|a(k)Rot(k)x(0) + \mathcal{F}(x(0),k)\|_\infty$$

Consequently, using (2.54) we infer that:

$$2^{\sum_{i=1}^{k} \log|det(U(i))|} 2^{-\sum_{i=1}^{k} r(i)} \leq 2^n v^n(k) \tag{2.57}$$

By taking expectations, the m-th moment stability assumption leads to:

$$\limsup_{k\to\infty} (\mathcal{E}[2^{\frac{m}{n}\log|det(\mathbf{U}(k))|}]\mathcal{E}[2^{-\frac{m}{n}\mathbf{r}(k)}])^k \leq 2^m \limsup_{k\to\infty} \mathcal{E}[\mathbf{v}^m(k)] < \infty \tag{2.58}$$

which implies that:

$$C > \alpha(\frac{m}{n}) + n\beta(m) + nR \tag{2.59}$$

where we used the fact that $\mathcal{E}[2^{\frac{m}{n}\log|det(\mathbf{U}(k))|}]\mathcal{E}[2^{-\frac{m}{n}\mathbf{r}(k)}] < 1$ must hold and that:

$$\log|det(\mathbf{U}(k))| = n\log|\mathbf{a}(k)|$$

□

Corollary 2.4.2 *Let* $\mathbf{x}(k)$ *be the solution of the following linear and time-invariant system:*

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + B\mathbf{u}(k) \tag{2.60}$$

If the state satisfies the following:

$$\sup_k E[\sup_{x(0)\in[-1/2,1/2]^n} \|\mathbf{x}(k)\|_\infty^m] < \infty \tag{2.61}$$

then the following must hold:

$$C - \alpha(\frac{m}{n_{unstable}}) > \sum_{i=1}^{n} \max\{\log|\lambda_i(A)|, 0\} \tag{2.62}$$

where $n_{unstable}$ is the number of unstable eigenvalues.

Proof: The proof is a direct adaptation of the proof of Theorem 2.4.1. □

2.5 Properties of the measures $\alpha(m)$ and $\beta(m)$

Consider that **a** and **r** are stochastic processes, that there is no uncertainty in the plant and no external disturbances, i.e., $\bar{z}_f = \bar{z}_a = \bar{d} = 0$. In such situation, (2.7) can be written as:

$$\mathbf{x}(k+1) = \mathbf{a}(k)\mathbf{x}(k) + \mathbf{u}(k) \tag{2.63}$$

For a given m, the stability condition of definition 2.2.4 becomes:

$$lim_{k \to \infty} \mathcal{E}[\bar{\mathbf{x}}(k)^m] = 0 \tag{2.64}$$

If **v** is a real random variable then Jensen's inequality [20] implies:

$$\mathcal{E}[2^{\mathbf{v}}] \geq 2^{\mathcal{E}[\mathbf{v}]}$$

where equality is attained if and only if **v** is a deterministic constant.

As such, $\log(\mathcal{E}[2^{\mathbf{v}}]2^{-\mathcal{E}[\mathbf{v}]}) \geq 0$ can be used as a measure of "randomness" which can be taken as an alternative to variance. Notice that such quantity may be more informative than variance because it depends on higher moments of **v**. We use this concept to interpret our results and express our conditions in a way that is amenable to a direct comparison with other publications. Along these lines, the following are randomness measures for $\log(|\mathbf{a}(k)|)$ and $\mathbf{r}(k)$:

$$\beta(m) = \frac{1}{m} \log(\mathcal{E}[2^{m\mathbf{l}_a^\delta(k)}]) \tag{2.65}$$

$$\alpha(m) = \frac{1}{m}\log(\mathcal{E}[2^{m\mathbf{r}^{\delta}(k)}]) \tag{2.66}$$

where $\mathbf{l}_a^{\delta}(k)$ and $\mathbf{r}^{\delta}(k)$ are given by:

$$\log(|\mathbf{a}(k)|) = \mathcal{E}[\log(|\mathbf{a}(k)|)] + \mathbf{l}_a^{\delta}(k) = R + \mathbf{l}_a^{\delta}(k) \tag{2.67}$$

$$\mathbf{r}(k) = \mathcal{E}[\mathbf{r}(k)] - \mathbf{r}^{\delta}(k) = C - \mathbf{r}^{\delta}(k) \tag{2.68}$$

The following equivalence is a direct consequence of the necessary and sufficient conditions proved in theorems 2.4.1 and 2.3.4:

$$\exists \text{ feedback scheme s.t. } \lim_{k\to\infty} \mathcal{E}[\bar{\mathbf{x}}(k)^m] = 0 \Longleftrightarrow \mathcal{C} > \mathcal{R} + \alpha(m) + \beta(m) \tag{2.69}$$

After examining (2.69), we infer that $\alpha(m)$ and $\beta(m)$ encompass the influence of m on the stability condition, while C and R are independent of m. The condition (2.69) suggests that $\alpha(m)$ is the right intuitive measure of quality, of a stochastic link, for the class considered in this Chapter.

The following are properties of $\alpha(m)$ and $\beta(m)$:

- Note that Jensen's inequality implies that $\alpha(m) \geq 0$ and $\beta(m) \geq 0$, where equality is achieved only if the corresponding random variable is deterministic. Accordingly, (2.69) shows that randomness in $\mathbf{r}(k)$ implies that $\mathcal{C} > R + \alpha(m)$ is necessary for stabilization. The fact that randomness in the channel creates the need for capacity larger than R, was already established, but quantified differently, in [54]. In addition, we find that randomness in the system adds yet another factor $\beta(m)$.

- by means of a Taylor expansion and taking limits, we get

$$\lim_{m\searrow 0} \alpha(m) = \lim_{m\searrow 0} \beta(m) = 0 \tag{2.70}$$

Under the above limit, the necessary and sufficient condition (2.69) becomes $C > R$. That is the weakest condition of stability and coincides with the one derived by [59] for almost sure stability. By means of (2.69) and (2.70) we can also conclude that if $C > R$, i.e. the feedback scheme is almost surely stabilizable [59], then it is m-th moment stabilizable for some $m > 0$.

- the opposite limiting case, gives

$$\lim_{m \to \infty} \alpha(m) = C - r_{min} \tag{2.71}$$

$$\lim_{m \to \infty} \beta(m) = \log(a^{sup}) - R \tag{2.72}$$

  where

$$r_{min} = \min\{r \in \{0, \ldots, \bar{r}\} : \mathcal{P}(\mathrm{r}(k) = r) \neq 0\}$$

$$a^{sup} = \sup\{\tilde{a} : \mathcal{P}(|\mathrm{a}(k)| \geq \tilde{a}) \neq 0\}$$

- $\alpha(m)$ and $\beta(m)$ are non-decreasing functions of m

From the previous properties of $\alpha(m)$ and $\beta(m)$ we find that

- a feedback scheme is stabilizing for all moments, i.e., $\forall m, \sup_k \mathcal{E}[\bar{\mathrm{x}}(k)^m] < \infty$ if and only if $r_{min} > \log(a^{sup})$.

- if $r_{min} = 0$ then there exists m_0 such that $\forall m \geq m_0, \sup_k \mathcal{E}[\bar{\mathrm{x}}(k)^m] = \infty$. This is the case of the erasure channel suggested by [54]. This conclusion was already reported in [54] (see Example 2.5.2).

- similarly, if $\log(a^{sup}) = \infty$ then there exists m_0 such that $\forall m \geq m_0, \sup_k \mathcal{E}[\bar{\mathrm{x}}(k)^m] = \infty$ (see example 2.5.1). Notice that a^{sup} can be larger than one and still $\mathcal{E}[\mathrm{a}(k)^m] < 1$ for some m. Even more, in example 2.5.1, we have $a^{sup} = \infty$ and $\mathcal{E}[|\mathrm{a}(k)|^m] < \infty$

Example 2.5.1 *Consider that* $|\mathrm{a}|$ *is log-normally distributed, i.e.,* $\log|\mathrm{a}(k)|$ *is normally distributed. An example where* $\mathrm{a}(k)$ *is log-normally distributed is given by [16]. If* $Var(|\mathrm{a}(k)|)$ *is the variance of* $|\mathrm{a}(k)|$ *then* $\beta(m)$ *is given by:*

$$\beta(m) = \frac{m}{2} \log\left(1 + \frac{Var(|\mathrm{a}(k)|)}{\left(\mathcal{E}\left[|\mathrm{a}(k)|\right]\right)^2}\right) \tag{2.73}$$

where the expression is obtained by direct integration. Note that $\beta(m)$ grows linearly with m. It illustrates a situation where, given $Var(|\mathrm{a}(k)|) > 0$, $\mathcal{C}$ and $\alpha(m)$, there always exist m large enough such that the necessary and sufficient condition $\mathcal{C} > \mathcal{R} + \beta(m) + \alpha(m)$ is violated.

The above analysis stresses the fact that feedback, using a stochastic link, acts by increasing m^{max} for which $\forall m \leq m^{max} \sup_k \mathcal{E}[\bar{\mathrm{x}}(k)^m] < \infty$. In some cases one may get $m^{max} = \infty$.

The Exponential Statistic

Directly from (2.65) and (2.66), we derive the equivalence below:

$$\mathcal{C} > \mathcal{R} + \alpha(m) + \beta(m) \Longleftrightarrow \mathcal{E}[|\mathrm{a}(k)|^m]\mathcal{E}[2^{-m\mathrm{r}(k)}] < 1 \tag{2.74}$$

The equivalences expressed in (2.69) and (2.74) show that all the information we need to know about the link is $\alpha(m)$ and $\mathcal{C}$ or, equivalently, $\mathcal{E}[2^{-m\mathrm{r}(k)}]$.

Example 2.5.2 *(From [54]) The binary erasure channel is a particular case of the class of stochastic links considered. It can be described by taking $r(k) = 1$ with probability $1 - p_e$ and $r(k) = 0$ with probability of erasure p_e. In that case, $\mathcal{E}[2^{-m\mathrm{r}(k)}] = 2^{-m}(1 - p_e) + p_e$. After working through the formulas, one may use (2.74) and (2.69) to get the same result as in [54]. In particular, the necessary and sufficient condition for the existence of a stabilizing feedback, for the time-invariant system with $\mathrm{a}(k) = a$, is given by*

$$0 \leq p_e < 1 - \frac{|a|^m - 1}{|a|^m(1 - 2^{-m})}$$

2.5.1 Determining the decay of the probability distribution function of $\bar{\mathrm{x}}$

In this subsection, we explore (2.69) as way to infer the decay of the probability distribution of $\bar{\mathrm{x}}(k)$.

From Markov's inequality (pp. 80 of [9]), we have that:

$$\forall m > 0, \forall k, \mathcal{P}(\bar{\mathbf{x}}(k) > \vartheta) \leq \vartheta^{-m}\mathcal{E}[\bar{\mathbf{x}}(k)^m] \tag{2.75}$$

On the other hand, for any given m, if $\bar{\mathbf{x}}(k)$ has a probability density function then:

$$\exists \varepsilon, \delta > 0, \forall k \geq 0, \forall \vartheta > 0, \mathcal{P}(\bar{\mathbf{x}}(k) > \vartheta) < \varepsilon\vartheta^{-(m+\delta+1)} \implies \limsup_{k\to\infty} \mathcal{E}[\bar{\mathbf{x}}(k)^m] < \infty \tag{2.76}$$

As such, we infer that (2.69) and (2.75)-(2.76) lead to:

$$\mathcal{C} > \mathcal{R} + \alpha(m) + \beta(m) \implies \exists \varepsilon > 0, \forall k, \forall \vartheta, \mathcal{P}(\bar{\mathbf{x}}(k) > \vartheta) \leq \varepsilon\vartheta^{-m} \tag{2.77}$$

$$\mathcal{C} < \mathcal{R} + \alpha(m) + \beta(m) \implies \forall \varepsilon, \delta > 0, \exists k, \exists \vartheta, \mathcal{P}(\bar{\mathbf{x}}(k) > \vartheta) \geq \varepsilon\vartheta^{-(m+\delta+1)} \tag{2.78}$$

2.5.2 Uncertainty Interpretation of the Statistical Description of the Stochastic Link

We suggest that $\alpha(m)$ can be viewed not only as a measure of the quality of the link, in the sense of how $\mathbf{r}(k)$ is expected to fluctuate over time, but it can also be modified to encapsulate a **description of uncertainty**. To be more precise, consider that Δ_l is an uncertainty set of stochastic links and that the "nominal" link has a deterministic data-rate $\mathbf{r}^o(k) = \mathcal{C}$. The elements of Δ_l are the following probability mass functions:

$$\Delta_l \subset \{p_l : \{0, \ldots, \bar{r}\} \to [0,1] : \sum_{i=0}^{\bar{r}} p_l(i) = 1, \sum_{i=0}^{\bar{r}} i \times p_l(i) = \mathcal{C}\}$$

where $p_l \in \Delta_l$ represents a stochastic link by specifying its statistics, i.e., $\mathcal{P}(r(k) = i) = p_l(i)$. The following is a measure of uncertainty in the link:

$$\bar{\alpha}(m) = \sup_{p_l \in \Delta_l} \frac{1}{m} \log(\mathcal{E}[2^{m\mathbf{r}^\delta(k)}]) \tag{2.79}$$

In this situation, (2.69) implies that the following is a necessary and sufficient condition for the existence of a feedback scheme that is stabilizing for all stochastic links in the uncertainty set Δ_l:

$$\mathcal{C} - \mathcal{R} - \beta(m) > \bar{\alpha}(m)$$

The authors suggest that $\mathcal{C} - \mathcal{R} - \beta(m) > \bar{\alpha}(m)$ should be viewed as a stability margin condition well adapted to this type of uncertainty.

If the plant and the link are time-invariant then $\mathcal{C} - \mathcal{R} > 0$ is necessary and sufficient for stabilizability. Stability is preserved for any stochastic link in Δ_l characterized by $\bar{\alpha}(m) < \mathcal{C} - \mathcal{R}$. This shows that the results by [59, 60] are robust to stochastic links with average transmission rate $\mathcal{C}$ and $\alpha(m) > 0$ sufficiently small.

2.5.3 Issues on the Stabilization of Linearizable Non-Linear Systems

In this section, we prove that a minimum rate must be guaranteed at all times in order to achieve stabilization in the sense of Lyapunov[5]. The fact that the classical erasure channel cannot be used to achieve stability in the sense of Lyapunov could already be inferred from [55]. Consider that the following is a state-space representation which corresponds to the linearization of a non-linear system around an equilibrium point:

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + B\mathbf{u}(k) \tag{2.80}$$

$$\mathbf{y}(k) = C\mathbf{x}(k) \tag{2.81}$$

where $x(k) \in \mathbb{R}^n$

If the linearized system is stable in the sense of Lyapunov then (2.80) must also be stable in the sense of Lyapunov. Consequently, that implies that:

$$\sup_{k} \sup_{x(0)\in[-1/2,1/2]^n} \|x(k)\|_\infty < \infty \tag{2.82}$$

where $\|x(k)\|_\infty = max_i|[x(k)]_i|$ and $[x(k)]_i$ are the components of $x(k)$. But (2.82) im-

[5] Also denoted as $\epsilon - \delta$ stability

plies that $\mathbf{x}(k)$ is stable for all moments, so Corollary 2.4.2 leads to:

$$\forall m, C - \alpha(\frac{m}{n_{unstable}}) > \sum_i \max\{\log|\lambda_i(A)|, 0\} \tag{2.83}$$

which also implies that:

$$r_{min} > \sum_i \max\{\log|\lambda_i(A)|, 0\} \tag{2.84}$$

where we have used (2.71). As a consequence, local stabilization imposes a minimum rate which has to be guaranteed at all times. The classical packet-erasure channel is characterized by $r_{min} = 0$ and, as such, it cannot be used for stabilization in the sense of Lyapunov. This is an important issue in the control of non-linear systems because, frequently, it is necessary to keep the state in a bounded set. That may arise from a physical limitation or as a way to stay in a region of model validity. The stabilization of non-linear systems in this framework was studied in [45], for deterministic channels. Since $r_{min} > \sum_i \max\{\log|\lambda_i(A)|, 0\}$, we conclude that it is sufficiently general to consider deterministic links with rate r_{min}, in the study of stabilization in the sense of Lyapunov.

2.6 Sufficient Conditions for a Class of Systems of Order Higher Than One

The results, derived in section 2.3, can be extended, in specific cases, to systems of order higher than one (see section 2.6.1). In the subsequent analysis, we outline how and suggest a few cases when such extension can be attained. Our results do not generalize to arbitrary stochastic systems of order $n > 1$. We emphasize that the proofs in this section are brief as they follow the same structure of the proofs of section 2.3 [6].

We use n as the order of a linear system whose state is indicated by $x(k) \in \mathbb{R}^n$. The following is a list of the adaptations, of the notation and definitions of section 2.1, to the multi-state case:

- if $x \in \mathbb{R}^n$ then we indicate its components by $[x]_i$, with $i \in \{1, \ldots, n\}$. In a similar

[6] The authors suggest the reading of section 2.3 first

way, if M is a matrix then we represent the element located in the i-th row and j-th column as $[M]_{ij}$. We also use $|M|$ to indicate the matrix whose elements are obtained as $[|M|]_{ij} = |[M]_{ij}|$.

- $\mathbb{R}^{n\times\infty}$ is used to represent the set of sequences of n-dimensional vectors, i.e., $x \in \mathbb{R}^{n\times\infty} \implies x(k) \in \mathbb{R}^n, k \in \mathbb{N}$.

- the infinity norm in $\mathbb{R}^{n\times\infty}$ is defined as:

$$\|x\|_\infty = \sup_i \max_j |[x(i)]_j|$$

It follows that if $x \in \mathbb{R}^n$ then $\|x\|_\infty = \max_{j\in\{1,\dots,n\}} |[x]_j|$. Accordingly, if $x \in \mathbb{R}^{n\times\infty}$ we use $\|x(k)\|_\infty = \max_{j\in\{1,\dots,n\}} |[x(k)]_j|$ to indicate the norm of a single vector, at time k, in contrast with $\|x\|_\infty = \sup_i \max_j |[x(i)]_j|$.

- the convention for random variables remains unchanged, e.g., $[\mathrm{x}(k)]_j$ is the jth component of a n-dimensional random sequence whose realizations lie on $\mathbb{R}^{n\times\infty}$

- If H is a sequence of matrices, with $H(k) \in \mathbb{R}^{n\times n}$, then

$$\|H\|_1 = \max_i \sum_{k=0}^{\infty} \sum_{j=1}^{n} |[H(k)]_{ij}|$$

For an arbitrary vector $x \in \mathbb{R}^n$ we use Hx to represent the sequence $H(k)x$. For a particular matrix $H(k)$, we also use $\|H(k)\|_1 = \max_i \sum_{j=1}^{n} |[H(k)]_{ij}|$.

- we use $\vec{1} \in \mathbb{R}^n$ to indicate a vector of ones, i.e., $[\vec{1}]_j = 1$ for $j \in \{1, \dots, n\}$.

2.6.1 Description of the nominal plant and equivalent representations

In this section, we introduce the state-space representation of the nominal discrete-time plant, for which we want to determine robust stabilizability. We also provide a condition, under which the stabilizability, of such state-space representation, can be inferred from the stabilizability of another representation which is more convenient. The condition is stated in Proposition 2.6.1 and a few examples are listed in remark 2.6.2. Such equivalent representation is used in section 2.6.2 as way to obtain stability conditions that depend explicitly on the eigenvalues of the dynamic matrix.

Consider the following nominal state-space realization:

$$\tilde{\mathbf{x}}(k+1) = \tilde{\mathbf{A}}(k)\tilde{\mathbf{x}}(k) + \tilde{\mathbf{u}}(k) + \tilde{\mathbf{d}}(k) \tag{2.85}$$

where $||\tilde{\mathbf{d}}||_\infty \leq \bar{\tilde{d}}$ and $\bar{\tilde{d}}$ is a pre-specified constant.

We also consider that $\tilde{\mathbf{A}}$ is a real Jordan form with a structure given by:

$$\tilde{\mathbf{A}}(k) = diag(\mathbf{J}_1(k), \ldots, \mathbf{J}_{\mathbf{q}(k)}(k)) \tag{2.86}$$

where $J_i(k)$ are real Jordan blocks[34] with multiplicity q_i satisfying $\sum_i q_i = n$.

The state-space representation of a linear and time-invariant system can always be transformed in a way that $\tilde{A}$ is in real Jordan form. The discretization of a controllable continuous-time and time-invariant linear system can always be expressed in the form (2.85), i.e., with $B = I$. If the system is not controllable, but stabilizable, then we can ignore the stable dynamics and consider only the unstable part which can be written in the form (2.85).

If the system is stochastic then, in general, there is no state transformation leading to $\mathbf{A}(k)$ in real Jordan form. The following is a list of conditions, under which a state-space representation of the form $\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) + \mathbf{u}(k)$ can be transformed in a new one, for which $\tilde{\mathbf{A}}$ is in real Jordan form:

- When the original dynamic matrices are already in real Jordan form. A particular instance of that are the second order stochastic systems with complex poles.

- A collection of systems with a state-space realization of the type $\mathbf{x}(k+1) = \mathbf{J}(k)\mathbf{x}(k) + \mathbf{u}(k)$ which connected in a shift-invariant topology. Here we used the fact that if several *copies* of the same system are connected in a shift-invariant topology then they can be decoupled by means of a time-invariant transformation [46].

Still, the representation (2.85)-(2.86) cannot be studied directly due to the fact that it may have complex eigenvalues. We will use the idea in [60] and show that, under certain conditions, there exists a transformation which leads to a more convenient state-space representation. Such representation has a dynamic matrix which is upper-triangular and has a diagonal with elements given by $|\lambda_i(\tilde{\mathbf{A}}(k))|$.

If we denote $R(\theta)$ as the following rotation:

$$R(\theta) = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix} \tag{2.87}$$

then the general structure of $J_i(k) \in \mathbb{R}^{q_i}$ is:

$$J_i(k) = \begin{cases} \begin{bmatrix} \eta_i(k) & 1 & \cdots & 0 & 0 \\ 0 & \eta_i(k) & 1 & \ddots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \eta_i(k) \end{bmatrix} & \text{if } \eta_i(k) \text{ is real} \\ \begin{bmatrix} |\eta_i(k)|R(\theta_i(k)) & I & \cdots & 0 & 0 \\ 0 & |\eta_i(k)|R(\theta_i(k)) & I & \ddots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & |\eta_i(k)|R(\theta_i(k)) \end{bmatrix} & \text{otherwise} \end{cases} \tag{2.88}$$

We start by following [60] and defining the following matrix:

Definition 2.6.1 *(Rotation dynamics) Let the real Jordan form $\tilde{\mathbf{A}}(k)$ of (2.85) be given by (2.86). We define the rotation dynamics $\mathbf{R}_{\tilde{A}}(k)$ as the following matrix:*

$$\mathbf{R}_{\tilde{A}}(k) = diag(\mathbf{J}_1^R(k), \ldots, \mathbf{J}_q^R(k)) \tag{2.89}$$

where $J_i^R(k) \in \mathbb{R}^{q_i}$ are given by:

$$J_i^R(k) = \begin{cases} sgn(\eta_i(k))I & \textit{if } \eta_i(k) \textit{ is real} \\ diag(R(\theta_i(k)), \ldots, R(\theta_i(k))) & \textit{otherwise} \end{cases} \tag{2.90}$$

For technical reasons, we use the idea of [60] and study the stability of $\mathbf{x}$ given by:

$$\mathbf{x}(k) = \mathbf{R}_{\tilde{A}}(k-1)^{-1} \cdots \mathbf{R}_{\tilde{A}}(0)^{-1} \tilde{\mathbf{x}}(k) \tag{2.91}$$

Remark 2.6.1 *The motivation for such time-varying transformation is that, by multiplying (2.85) on the left by $\mathbf{R}_{\tilde{A}}(k)^{-1} \cdots \mathbf{R}_{\tilde{A}}(0)^{-1}$, the nominal dynamics of $\mathbf{x}$ is given by:*

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) + \mathbf{d}(k) + \mathbf{u}(k) \tag{2.92}$$

where $\mathbf{d}(k) = \mathbf{R}_{\tilde{A}}(k-1)^{-1} \cdots \mathbf{R}_{\tilde{A}}(0)^{-1} \tilde{\mathbf{d}}(k)$, $\mathbf{u}(k) = \mathbf{R}_{\tilde{A}}(k-1)^{-1} \cdots \mathbf{R}_{\tilde{A}}(0)^{-1} \tilde{\mathbf{u}}(k)$ and $\mathbf{A}(k)$ is the following upper-triangular matrix[7]:

$$\mathbf{A}(k) = \mathbf{R}_{\tilde{A}}(k)^{-1} \tilde{\mathbf{A}}(k) = \begin{bmatrix} |\lambda_1(\tilde{\mathbf{A}}(k))| & \cdots & \\ 0 & \ddots & \vdots \\ 0 & 0 & |\lambda_n(\tilde{\mathbf{A}}(k))| \end{bmatrix} \tag{2.93}$$

The following proposition is a direct consequence of the previous discussion:

Proposition 2.6.1 *(Condition for equivalence of representations) Let $\tilde{\mathbf{A}}(k)$ be such that $\mathbf{R}_{\tilde{A}}(k)$ satisfies:*

$$\sup_k \|\mathbf{R}_{\tilde{A}}(k)^{-1} \cdots \mathbf{R}_{\tilde{A}}(0)^{-1}\|_1 \leq \Gamma_1 < \infty \tag{2.94}$$

$$\sup_k \| \left(\mathbf{R}_{\tilde{A}}(k)^{-1} \cdots \mathbf{R}_{\tilde{A}}(0)^{-1}\right)^{-1} \|_1 \leq \Gamma_2 < \infty \tag{2.95}$$

Under the above conditions, the stabilization of (2.92)-(2.93) and the stabilization of (2.85)-(2.86) are equivalent in the sense of (2.96)-(2.97).

$$\limsup_{k \to \infty} \|\mathbf{x}(k)\|_\infty \leq \Gamma_1 \limsup_{k \to \infty} \|\tilde{\mathbf{x}}(k)\|_\infty \leq \Gamma_1 \Gamma_2 \limsup_{k \to \infty} \|\mathbf{x}(k)\|_\infty \tag{2.96}$$

$$\limsup_{k \to \infty} \mathcal{E}[\|\mathbf{x}(k)\|_\infty^m] \leq \Gamma_1^m \limsup_{k \to \infty} \mathcal{E}[\|\tilde{\mathbf{x}}(k)\|_\infty^m] \leq \Gamma_1^m \Gamma_2^m \limsup_{k \to \infty} \mathcal{E}[\|\mathbf{x}(k)\|_\infty^m] \tag{2.97}$$

Remark 2.6.2 *Examples of $\tilde{\mathbf{A}}(k)$ for which (2.94)-(2.95) hold are:*

[7]Here we use an immediate modification of lemma 3.4.1, from [60]. It can be shown that, if $\tilde{\mathbf{A}}(k)$ is a real Jordan form then $\mathbf{R}_{\tilde{A}}(j)^{-1} \tilde{\mathbf{A}}(k) = \tilde{\mathbf{A}}(k) \mathbf{R}_{\tilde{A}}(j)^{-1}$ holds for any j, k. This follows from the fact that $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2) = R(\theta_2)R(\theta_1)$ holds for arbitrary θ_1 and θ_2.

- *all time-invariant* $\tilde{A}$

- $\tilde{\mathbf{A}}(k) = diag(\mathbf{J}_1(k), \ldots, \mathbf{J}_q(k))$ *where* q_i *are invariant.*

- *all 2-dimensional* $\mathbf{A}(k)$. *In this case* $\mathbf{R}_{\tilde{A}}(k)$ *is always a rotation matrix, which includes the identity as a special case. Under such condition, the bounds in (2.94)-(2.95) are given by*

$$\sup_k \|\mathbf{R}_{\tilde{A}}(k)^{-1} \cdots \mathbf{R}_{\tilde{A}}(0)^{-1}\|_1 \leq 2$$

$$\sup_k \| \left(\mathbf{R}_{\tilde{A}}(k)^{-1} \cdots \mathbf{R}_{\tilde{A}}(0)^{-1}\right)^{-1} \|_1 \leq 2$$

2.6.2 Description of uncertainty and robust stability

Let $\varrho \in \mathbb{N}_+ \bigcup\{\infty\}$, $\bar{d} \geq 0$, $\bar{z}_f^x, \bar{z}_f^u \in [0,1)$ and $\bar{z}_a > 0$ be given constants then we study the stabilizability of the following uncertain system:

$$\mathbf{x}(k+1) = \mathbf{A}(k)(I + \mathbf{Z}_a(k))\mathbf{x}(k) + \mathbf{u}(k) + \mathbf{d}(k) + G_f(\mathbf{x}, \mathbf{u})(k) \tag{2.98}$$

$$\mathbf{A}(k) = \begin{bmatrix} |\lambda_1(\mathbf{A}(k))| & \cdots & \\ 0 & \ddots & \vdots \\ 0 & 0 & |\lambda_n(\mathbf{A}(k))| \end{bmatrix} \tag{2.99}$$

where $\|\mathbf{d}\|_\infty \leq \bar{d}$, $|[\mathbf{Z}_a(k)]_{ij}| \leq \bar{z}_a$ and $G_f(\mathbf{x}, \mathbf{u})(k)$ satisfies:

$$\|G_f(\mathbf{x}, \mathbf{u})(k)\|_\infty \leq \bar{z}_f^x \max\{\|\mathbf{x}(k-\varrho+1)\|_\infty, \ldots, \|\mathbf{x}(k)\|_\infty\} + \bar{z}_f^u \max\{\|\mathbf{u}(k-\varrho+1)\|_\infty, \ldots, \|\mathbf{u}(k)\|_\infty\} \tag{2.100}$$

Recall, from Proposition 2.6.1, that the stabilizability of a state-space representation of the form (2.85), satisfying (2.94)-(2.95), can be studied in the equivalent form where (2.99) holds. We emphasize that, in (2.98), we incorporate $\mathbf{u}$ in the description of the feedback uncertainty. As it will be evident from the subsequent discussion, such generalization can be treated with the same techniques used in section 2.3. We decide for including $\mathbf{u}$ in G_f because that allows for a richer description of uncertainty.

A given feedback scheme is robustly stabilizing if it satisfies the following definition.

Definition 2.6.2 *(m-th Moment Robust Stability) Let* $m > 0$, $\varrho \in \mathbb{N}_+ \bigcup \{\infty\}$, $\bar{z}_f^x, \bar{z}_f^u \in [0,1)$, $\bar{z}_a \in [0,1)$ *and* $\bar{d} \geq 0$ *be given. The system (2.98), under a given feedback scheme, is m-th moment (robustly) stable provided that the following holds:*

$$\begin{cases} \lim_{k\to\infty} \mathcal{E}\left[\|\bar{\mathbf{x}}(k)\|_\infty^m\right] = 0 & \text{if } \bar{z}_f^x = \bar{z}_f^u = \bar{d} = 0 \\ \exists b \in \mathbb{R}_+ \text{ s.t. } \limsup_{k\to\infty} \mathcal{E}\left[\|\bar{\mathbf{x}}(k)\|_\infty^m\right] < b & \text{otherwise} \end{cases} \tag{2.101}$$

where $\bar{\mathbf{x}}(k)$ *is given by:*

$$[\bar{\mathbf{x}}(k)]_i = \sup_{x(0)\in[-1/2,1/2]^n} |[\mathbf{x}(k)]_i|$$

2.6.3 Feedback structure and channel usage assumptions

In order to study the stabilization of systems of order higher than one, we assume the existence of a channel allocation $\vec{\mathbf{r}}(k) \in \{0, \dots, \bar{r}\}^n$ satisfying:

$$\sum_{j=1}^{n} [\vec{\mathbf{r}}(k)]_j = \mathbf{r}(k) \tag{2.102}$$

where $\mathbf{r}(k)$ is the instantaneous rate sequence as described in section 2.2.5. We also emphasize that $\mathbf{A}(k)$ and $\vec{\mathbf{r}}(k)$ are i.i.d and independent of each other.

Using the same notation of section 2.1, we define C_j and $[\vec{\mathbf{r}}^\delta(k)]_j$ as:

$$[\vec{\mathbf{r}}(k)]_j = C_j - [\vec{\mathbf{r}}^\delta(k)]_j \tag{2.103}$$

Similarly, we also define $\alpha_i(m)$ as:

$$\alpha_i(m) = \frac{1}{m} \log \mathcal{E}[2^{m[\vec{\mathbf{r}}^\delta(k)]_i}] \tag{2.104}$$

In the general case, the allocation problem is difficult because it entails a change of the encoding process described in section 2.2.5. The encoding must be such that each $[\vec{\mathbf{r}}^\delta(k)]_i$ corresponds to the instantaneous rate of a truncation operator. In section 2.6.8 we solve the allocation problem for a class of stochastic systems in the presence of a stochastic link.

As in the one dimensional case, we assume that both the encoder and the controller have access to $A(0), \ldots, A(k)$ and $\vec{r}(k-1)$ as well as the constants ϱ, $\bar{z}_f^x$, $\bar{z}_f^u$, $\bar{z}_a$ and $\bar{d}$. The encoder and the controller are described as:

- The encoder is a function $\mathcal{F}_k^e : \mathbb{R}^{n\times(k+1)} \rightarrow \{0,1\}^{n\times\bar{r}}$ that has the following dependence on observations:

$$\mathcal{F}_k^e(x(0), \ldots, \mathbf{x}(k)) = (\mathbf{b}_1, \ldots, \mathbf{b}_{\bar{r}}) \tag{2.105}$$

  where $\mathbf{b}_i \in \{0,1\}^n$.

- The control action results from a map, not necessarily memoryless, $\mathcal{F}_k^c : \bigcup_{i=0}^{\bar{r}} \{0,1\}^{n\times i} \rightarrow \mathbb{R}$ exhibiting the following functional dependence:

$$\mathbf{u}(k) = \mathcal{F}_k^c(\vec{\mathbf{b}}(k)) \tag{2.106}$$

  where $\vec{\mathbf{b}}(k)$ is the vector for which, each component $[\vec{\mathbf{b}}(k)]_j$, comprises a string of $[\vec{\mathbf{r}}(k)]_j$ bits successfully transmitted through the link, i.e.:

$$[\vec{\mathbf{b}}(k)]_j = [\mathcal{F}_k^l(\mathbf{b}_1, \ldots, \mathbf{b}_{\bar{r}})]_j = \left([\mathbf{b}_1]_j, \ldots, [\mathbf{b}_{[\vec{\mathbf{r}}(k)]_j}]_j\right) \tag{2.107}$$

  As such, $\mathbf{u}(k)$ can be equivalently expressed as

$$\mathbf{u}(k) = (\mathcal{F}_k^c \circ \mathcal{F}_k^l \circ \mathcal{F}_k^e)(x(0), \ldots, \mathbf{x}(k))$$

2.6.4 Construction of a stabilizing feedback scheme

The construction of a stabilizing scheme follows the same steps used in section 2.3. The following is the definition of a multidimensional upper-bound sequence.

Definition 2.6.3 *(Upper-bound Sequence) Let $\bar{z}_f^x, \bar{z}_f^u \in [0,1)$, $\bar{z}_a \in [0,1)$, $\bar{d} \geq 0$ and $\varrho \in \mathbb{N}_+ \bigcup\{\infty\}$ be given. Define the upper-bound sequence $\mathbf{v}(k)$, with $v(k) \in \mathbb{R}^n$, as:*

$$\mathbf{v}(k+1) = \mathbf{A}_{cl}(k)\mathbf{v}(k) + (\bar{z}_f^x I + \bar{z}_f^u |\mathbf{A}(k)|) \max\{\|\mathbf{v}(k-\varrho+1)\|_\infty, \ldots, \|\mathbf{v}(k)\|_\infty\}\vec{1} + \bar{d}\vec{1}, \tag{2.108}$$

where $[|A(k)|]_{ij} = |[A(k)]_{ij}|$, $v(i) = 0$ *for* $i < 0$, $[v(0)]_j = \frac{1}{2}$ *and* $A_{cl}(k)$ *is given by:*

$$A_{cl}(k) = |A(k)| \left(diag(2^{-[\vec{r}(k)]_1}, \ldots, 2^{-[\vec{r}(k)]_n}) + \bar{z}_a \vec{1}\vec{1}^T \right) \tag{2.109}$$

Adopt the feedback scheme of definition 2.3.3, mutatis mutandis, for the multi-dimensional case. By measuring the state $x(k)$ and using $[\vec{r}(k)]_j$ bits, at time k, to encode each component $[x(k)]_j$, we construct $[\hat{x}(k)]_j$ such that

$$|[x(k)]_j - [\hat{x}(k)]_j| \leq 2^{-[\vec{r}(k)]_j}[v(k)]_j \tag{2.110}$$

Accordingly, $\mathbf{u}(k)$ is defined as:

$$\mathbf{u}(k) = -\mathbf{A}(k)\hat{\mathbf{x}}(k) \tag{2.111}$$

The following lemma establishes that the stabilization of $\mathbf{v}(k)$ is sufficient for the stabilization of $\mathbf{x}(k)$.

Lemma 2.6.2 *Let* $\bar{z}_f^x, \bar{z}_f^u \in [0,1)$, $\bar{z}_a \in [0,1)$ *and* $\bar{d} \geq 0$ *be given. If* $\mathbf{x}(k)$ *is the solution of (2.98) under the feedback scheme given by (2.110)-(2.111), then the following holds:*

$$[\bar{\mathbf{x}}(k)]_j \leq [\mathbf{v}(k)]_j$$

for all $\varrho \in \mathbb{N}_+ \bigcup \{\infty\}$, $\|\mathbf{d}(k)\|_\infty \leq \bar{d}$, *every choice* $|[Z_a]_{ij}| \leq \bar{z}_a$ *and* G_f *satisfying:*

$$\begin{aligned} \|G_f(\mathbf{x},\mathbf{u})(k)\|_\infty \leq \bar{z}_f^x \max\{\|\mathbf{x}(k-\varrho+1)\|_\infty, \ldots, \|\mathbf{x}(k)\|_\infty\} \\ + \bar{z}_f^u \max\{\|\mathbf{u}(k-\varrho+1)\|_\infty, \ldots, \|\mathbf{u}(k)\|_\infty\} \end{aligned} \tag{2.112}$$

Proof: The proof follows the same steps as in lemma 2.3.1. We start by assuming that $[\bar{x}(i)]_j \leq [v(i)]_j$ for $i \in \{0, \ldots, k\}$ and proceed to prove that $[\bar{x}(k+1)]_j \leq [v(k+1)]_j$.

From (2.98) and the feedback scheme (2.110)-(2.111), we find that:

$$\begin{bmatrix} |[x(k+1)]_1| \\ \vdots \\ |[x(k+1)]_n| \end{bmatrix} \underset{element-wise}{\leq} |A(k)| \begin{bmatrix} |[v(k)]_1| 2^{-\vec{r}_1(k)} \\ \vdots \\ |[v(k)]_n| 2^{-\vec{r}_n(k)} \end{bmatrix} + |A(k)|\,|Z_a(k)| \begin{bmatrix} |[v(k)]_1| \\ \vdots \\ |[v(k)]_n| \end{bmatrix} +$$
$$\bar{z}_f^x \vec{1} \max\{\|v(k-\varrho+1)\|_\infty, \ldots, \|v(k)\|_\infty\} + \bar{z}_f^u \vec{1} \max\{\|u(k-\varrho+1)\|_\infty, \ldots, \|u(k)\|_\infty\} \tag{2.113}$$

In order to address the dependence on $\mathbf{u}$, we notice that (2.111) implies that:

$$|[\mathbf{u}(k)]_j| \leq [|A(k)|\,|\mathbf{v}(k)|]_j \tag{2.114}$$

which by substituting in (2.113) leads to the conclusion of the proof. □

2.6.5 Sufficiency for the deterministic/time-invariant case

Accordingly, the following theorem establishes the multi-dimensional analog to theorem 2.3.2.

Theorem 2.6.3 ***(Sufficiency conditions for Robust Stability)*** *Let A be the dynamic matrix of (2.98), $\varrho \in \mathbb{N}_+ \bigcup \{\infty\}$, $\bar{z}_f^x, \bar{z}_f^u \in [0,1)$, $\bar{z}_a \in [0,1)$ and $\bar{d} \geq 0$ be given and $H(k)$ be defined as*

$$H(k) = A_{cl}^k,\; k \geq 0$$

where $A_{cl} = |A| \left(diag(2^{-C_1}, \ldots, 2^{-C_n}) + \bar{z}_a \vec{1}\vec{1}^T \right)$ and $[|A|]_{ij} = |[A]_{ij}|$. Consider that $\mathbf{x}(k)$ is the solution of (2.98) under the feedback scheme of (2.110)-(2.111) as well as the following conditions:

- **(C 1)** $\max_i \lambda_i(A_{cl}) < 1$
- **(C 2)** $(\bar{z}_f^x + \|A\|_1 \bar{z}_f^u)\|H\|_1 < 1$

If conditions **(C 1)** *and* **(C 2)** *are satisfied then the following holds:*

$$\bar{x}(k) \leq \|H\|_1 \left((\bar{z}_f^x + \|A\|_1 \bar{z}_f^u) \frac{\|H\|_1 \bar{d} + \|\bar{g}\|_\infty \frac{1}{2}}{1 - (\bar{z}_f^x + \|A\|_1 \bar{z}_f^u)\|H\|_1} + \bar{d} \right) + \|\bar{g}(k)\|_\infty \frac{1}{2} \tag{2.115}$$

where $\tilde{g}(k) = A_{cl}^k \vec{1}$.

Proof: We start by noticing that the condition (C1) is necessary and sufficient to guarantee that $\|H\|_1$ is finite. Following the same steps, used in the proof of theorem 2.3.2, from definition 2.6.3, we have that:

$$\mathbf{v}(k) = A_{cl}^k v(0) + \sum_{i=0}^{k-1} A_{cl}^{k-i-1} \left((\bar{z}_f^x I + \bar{z}_f^u |\mathbf{A}(k)|) \max\{\|\mathbf{v}(i-\varrho+1)\|_\infty, \ldots, \|\mathbf{v}(i)\|_\infty\}\vec{1} + \bar{d}\vec{1} \right) \quad (2.116)$$

$$\|\pi_k \mathbf{v}\|_\infty \leq \|\tilde{g}\|_\infty \frac{1}{2} + \|H\|_\infty ((\bar{z}_f^x + \|A\|_1 \bar{z}_f^u)\|\pi_k \mathbf{v}\|_\infty + \bar{d}) \quad (2.117)$$

where we use $\|\pi_k \mathbf{v}\|_\infty = \max\{\|\mathbf{v}(0)\|_\infty, \ldots, \|\mathbf{v}(k)\|_\infty\}$ and $\tilde{g}(k) = A_{cl}^k \vec{1}$. By means of lemma 2.6.2, the formula (2.115) is obtained by isolating $\|\pi_k \mathbf{v}\|_\infty$ in (2.117) and substituting it back in (2.116).□

Interpretation for $\bar{z}_a = 0$

If $\bar{z}_a = 0$ then A_{cl}, of Theorem 2.6.3, can be written as:

$$A_{cl} = \begin{bmatrix} |\lambda_1(A)|2^{-C_1} & \cdots & \\ 0 & \ddots & \vdots \\ 0 & 0 & |\lambda_n(A)|2^{-C_n} \end{bmatrix} \quad (2.118)$$

The increase of C_i causes the decrease of $\lambda_i(A_{cl})$ and $\|H\|_1$. Accordingly, conditions (C1) and (C2), from theorem 2.6.3, lead to the conclusion that the increase in C_i gives a guarantee that the feedback loop is stable under larger uncertainty, as measured by $(\bar{z}_f^x + \|A\|_1 \bar{z}_f^u)$. In addition, if we denote $R_i = \log(|\lambda_i(A)|)$ then we can cast condition (C1) as:

$$C_i > R_i \quad (2.119)$$

2.6.6 Sufficiency for the stochastic case

We derive the multi-dimensional version of the sufficiency results in section 2.3.2. The results presented below are the direct generalizations of lemma 2.3.3 and theorem 2.3.4.

Definition 2.6.4 *(Upper-bound sequence for the stochastic case) Let* $\varrho \in \mathbb{N}_+$, $\bar{z}_f^x, \bar{z}_f^u \in [0,1)$, $\bar{z}_a \in [0,1)$ *and* $\bar{d} \geq 0$ *be given. Given* m, *consider the following sequence:*

$$v_m(k+1) = A_{cl,m}v_m(k) + \left((n\varrho)^{\frac{1}{m}}(\bar{z}_f^x + \bar{z}_f^u\|A_m\|_1)\max\{\|v_m(k-\varrho+1)\|_\infty, \ldots, \|v_m(k)\|_\infty\} + \bar{d}\right)\vec{1} \quad (2.120)$$

where $v_m(i) = 0$ *for* $i < 0$, $v_m(0) = \frac{1}{2}\vec{1}$ *and* $A_{cl,m}$ *is defined as*

$$A_{cl,m} = A_m\left(diag(2^{-C_1+\alpha_1(m)}, \ldots, 2^{-C_n+\alpha_n(m)}) + \bar{z}_a\vec{1}\vec{1}^T\right)$$

and

$$[A_m]_{ij} = \mathcal{E}[|[\mathbf{A}(k)]_{ij}|^m]^{\frac{1}{m}}$$

Lemma 2.6.4 ***(M-th moment boundedness)***

If $\mathbf{x}(k)$ *is the solution of (2.98) under the feedback scheme of (2.110)-(2.111), then the following holds*

$$\mathcal{E}[\bar{\mathbf{x}}(k)^m]^{\frac{1}{m}} \leq v_m(k)$$

Proof: We start by showing that $\mathcal{E}[[\mathbf{v}(k)]_i^m]^{\frac{1}{m}} \leq [v_m(k)]_i$. We proceed by induction, by assuming that $\mathcal{E}[[\mathbf{v}(j)]_i^m]^{\frac{1}{m}} \leq [v_m(j)]_i$ holds for $j \in \{0, \ldots, k\}$ and proving that $\mathcal{E}[[\mathbf{v}(k+1)]_i^m]^{\frac{1}{m}} \leq [v_m(k+1)]_i$.

Let $\mathbf{z}$, $\mathbf{s}$ and $\mathbf{g}$ be random variables with $\mathbf{z}$ independent of $\mathbf{s}$. By means of the Minkovsky inequality, we know that $\mathcal{E}[|\mathbf{zs}+\mathbf{g}|^m]^{\frac{1}{m}} \leq \mathcal{E}[|\mathbf{z}|^m]^{\frac{1}{m}}\mathcal{E}[|\mathbf{s}|^m]^{\frac{1}{m}} + \mathcal{E}[|\mathbf{g}|^m]^{\frac{1}{m}}$. Using such property, the following inequality is a consequence of (2.108):

$$\begin{bmatrix} \mathcal{E}[[\mathbf{v}(k+1)]_1^m]^{\frac{1}{m}} \\ \vdots \\ \mathcal{E}[[\mathbf{v}(k+1)]_n^m]^{\frac{1}{m}} \end{bmatrix} \underset{element-wise}{\leq} A_{cl,m} \begin{bmatrix} \mathcal{E}[[\mathbf{v}(k)]_1^m]^{\frac{1}{m}} \\ \vdots \\ \mathcal{E}[[\mathbf{v}(k)]_n^m]^{\frac{1}{m}} \end{bmatrix} + (\bar{z}_f^x I + \bar{z}_f^u A_m)\mathcal{E}[\max\{\|\mathbf{v}(k-\varrho+1)\|_\infty, \dots, \|\mathbf{v}(k)\|_\infty\}^m]^{\frac{1}{m}}\vec{1} + \bar{d}\vec{1} \quad (2.121)$$

But using the inductive assumption that $\mathcal{E}[[\mathbf{v}(j)]_i^m]^{\frac{1}{m}} \leq [v_m(j)]_i$ holds for $j \in \{0, \dots, k\}$ and that:

$$\mathcal{E}[\max\{\|\mathbf{v}(k-\varrho+1)\|_\infty, \dots, \|\mathbf{v}(k)\|_\infty\}^m]^{\frac{1}{m}} \leq (n\varrho)^{\frac{1}{m}} \max\{\mathcal{E}[\|[\mathbf{v}(k-\varrho+1)]\|_\infty^m]^{\frac{1}{m}}, \dots, \mathcal{E}[\|[\mathbf{v}(k)]\|_\infty^m]^{\frac{1}{m}}\} \quad (2.122)$$

we can rewrite (2.121) as:

$$\begin{bmatrix} \mathcal{E}[[\mathbf{v}(k+1)]_1^m]^{\frac{1}{m}} \\ \vdots \\ \mathcal{E}[[\mathbf{v}(k+1)]_n^m]^{\frac{1}{m}} \end{bmatrix} \underset{element-wise}{\leq} A_{cl,m} v_m(k) + \left((n\varrho)^{\frac{1}{m}}(\bar{z}_f^x + \bar{z}_f^u\|A_m\|_1)\max\{\|v_m(k-\varrho+1)\|_\infty, \dots, \|v_m(k)\|_\infty\} + \bar{d}\right)\vec{1} \quad (2.123)$$

Since the induction hypothesis is verified, we can use lemma 2.6.2 to finalize the proof.

□

Theorem 2.6.5 ***(Sufficiency conditions for Robust m-th moment Stability)*** *Let A be the dynamic matrix of (2.98), $\varrho \in \mathbb{N}_+$, $\bar{z}_f^x, \bar{z}_f^u \in [0,1)$, $\bar{z}_a \in [0,1)$ and $\bar{d} \geq 0$ be given and $H_m(k)$ be defined as*

$$H_m(k) = A_{cl,m}^k, \; k \geq 0$$

where $A_{cl,m} = A_m\left(diag(2^{-C_1+\alpha_1(m)}, \dots, 2^{-C_n+\alpha_n(m)}) + \bar{z}_a\vec{1}\vec{1}^T\right)$, $[A_m]_{ij} = \mathcal{E}\left[|[A(k)]_{ij}|^m\right]^{\frac{1}{m}}$ and $[|A|]_{ij} = |[A]_{ij}|$. Consider that $\mathbf{x}(k)$ is the solution of (2.98) under the feedback scheme of (2.110)-(2.111) as well as the following conditions:

- **(C 1)** $\max_i \lambda_i(A_{cl,m}) < 1$
- **(C 2)** $(n\varrho)^{\frac{1}{m}}(\bar{z}_f^x + \|A_m\|_1 \bar{z}_f^u)\|H_m\|_1 < 1$

If conditions **(C 1)** *and* **(C 2)** *are satisfied then the following holds:*

$$\mathcal{E}[\bar{\mathbf{x}}(k)^m]^{\frac{1}{m}} \leq \|H_m\|_1 \left((n\varrho)^{\frac{1}{m}}(\bar{z}_f^x + \|A_m\|_1 \bar{z}_f^u) \frac{\|H_m\|_1 \bar{d} + \|\tilde{g}_m\|_\infty \frac{1}{2}}{1 - (n\varrho)^{\frac{1}{m}}(\bar{z}_f^x + \|A_m\|_1 \bar{z}_f^u)\|H_m\|_1} + \bar{d} \right) + \|\tilde{g}_m(k)\|_\infty \frac{1}{2} \quad (2.124)$$

where $\tilde{g}_m(k) = A_{cl,m}^k \vec{1}$.

Proof: We start by noticing that the condition (C1) is necessary and sufficient to guarantee that $\|H_m\|_1$ is finite. From definition 2.6.4, we have that:

$$v_m(k) = A_{cl,m}^k v(0) + \sum_{i=0}^{k-1} A_{cl,m}^{k-i-1} \left((n\varrho)^{\frac{1}{m}}(\bar{z}_f^x + \bar{z}_f^u \|A_m\|_1) \max\{\|v_m(i - \varrho + 1)\|_\infty, \ldots, \|v_m(i)\|_\infty\} + \bar{d} \right) \vec{1} \quad (2.125)$$

$$\|\pi_k \mathbf{v}_m\|_\infty \leq \|\tilde{g}_m\|_\infty \frac{1}{2} + \|H_m\|_1 \left((n\varrho)^{\frac{1}{m}}(\bar{z}_f^x + \bar{z}_f^u \|A_m\|_1)\|\pi_k \mathbf{v}_m\|_\infty + \bar{d} \right) \quad (2.126)$$

where we use $\|\pi_k \mathbf{v}_m\|_\infty = \max\{\|\mathbf{v}_m(0)\|_\infty, \ldots, \|\mathbf{v}_m(k)\|_\infty\}$ and $\tilde{g}_m(k) = A_{cl,m}^k \vec{1}$. By means of lemma 2.6.4, the formula (2.124) is obtained by isolating $\|\pi_k \mathbf{v}_m\|_\infty$ in (2.126) and substituting it back in (2.125).□

2.6.7 Sufficiency for the case $\bar{z}_a = 0$

If $\bar{z}_a = 0$ then $A_{cl,m}$ of Theorem 2.6.5 can be expressed as:

$$A_{cl,m} = \begin{bmatrix} \mathcal{E}[|\lambda_1(\mathbf{A}(k))|^m]^{\frac{1}{m}} 2^{-C_1 + \alpha_1(m)} & \ldots & \\ 0 & \ddots & \vdots \\ 0 & 0 & \mathcal{E}[|\lambda_n(\mathbf{A}(k))|^m]^{\frac{1}{m}} 2^{-C_n + \alpha_n(m)} \end{bmatrix} \quad (2.127)$$

Accordingly, if we define

$$\beta_i(m) = \frac{1}{m}\log(\mathcal{E}[2^{m(\log(|\lambda_i(\mathbf{A}(k))|)-R_i)}])$$

where $R_i = \mathcal{E}[\log(|\lambda_i(\mathbf{A}(k))|)]$, then condition (C1) of Theorem 2.6.5 can be written as:

$$C_i > R_i + \alpha_i(m) + \beta_i(m) \tag{2.128}$$

Also, from condition (C2), we find that by increasing the difference $C_i - (R_i + \alpha_i(m) + \beta_i(m))$ we reduce $\|H_m\|_1$ and that improves robustness to uncertainty as measured by $(\bar{z}_f^x + \|A_m\|_1 \bar{z}_f^u)$.

2.6.8 Solving the Allocation Problem for a Class of Stochastic Systems

Given $\bar{d} > 0$ and $\bar{z}_f^x, \bar{z}_f^u \in [0, 1)$, consider the following n-th order state-space representation:

$$\mathbf{x}(k+1) = \mathbf{J}(k)\mathbf{x}(k) + \mathbf{u}(k) + \mathbf{d}(k) + G_f(\mathbf{x}, \mathbf{u})(k) \tag{2.129}$$

where G_f is a causal operator satisfying $\|G_f(\mathbf{x}, \mathbf{u})\|_\infty \le \bar{z}_f^x \|\mathbf{x}\|_\infty + \bar{z}_f^u \|\mathbf{u}\|_\infty$, $\|\mathbf{d}\|_\infty \le \bar{d}$ and $\mathbf{J}(k)$ is a real Jordan block of the form:

$$\mathbf{J}(k) = \begin{cases} \begin{bmatrix} \mathbf{a}(k) & 1 & \cdots & 0 & 0 \\ 0 & \mathbf{a}(k) & 1 & \ddots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \mathbf{a}(k) \end{bmatrix} & \text{if } \mathbf{a}(k) \text{ is real} \\ \begin{bmatrix} |\mathbf{a}(k)|\mathbf{Rot}(k) & I & \cdots & 0 & 0 \\ 0 & |\mathbf{a}(k)|\mathbf{Rot}(k) & I & \ddots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & |\mathbf{a}(k)|\mathbf{Rot}(k) \end{bmatrix} & \text{otherwise} \end{cases} \tag{2.130}$$

where $\mathbf{Rot}$ is an i.i.d.sequence of random rotation matrices.

Take the scheme of section 2.2.5 as a starting point. Assume also that $\mathbf{r}(l)$ is always a multiple of n, i.e., $r(l) \in \{0, n, 2n, \ldots, \bar{r}\}$. In order to satisfy this assumption, we only need to adapt the scheme of section 2.2.5 by selecting packets whose size is a multiple of n. By doing so, we can modify the encoding/decoding scheme of section 2.2.5 and include in each packet an equal number of bits from each $[\vec{\mathbf{r}}(l)]_i$. By including the most important bits in the highest priority packets, we guarantee that each $[\vec{\mathbf{r}}(l)]_i$ corresponds to the instantaneous rate of a truncation operator. As such, we adopt the following allocation:

$$[\vec{\mathbf{r}}(l)]_i = \frac{\mathbf{r}(l)}{n} \tag{2.131}$$

where we also use C_i and define the zero mean i.i.d. random variable $[\vec{\mathbf{r}}^{\delta}(l)]_i$, satisfying:

$$[\vec{\mathbf{r}}(l)]_i = C_i - [\vec{\mathbf{r}}^{\delta}(l)]_i \tag{2.132}$$

From the definition of $\alpha_i(m)$ and $\alpha(m)$, the parameters characterizing the allocation (2.131) and $\mathbf{r}(l)$ are related through:

$$\alpha_i(m) = \frac{1}{m} \log \mathcal{E}[2^{-m[\vec{\mathbf{r}}^{\delta}(l)]_i}] = \frac{1}{n}\frac{n}{m} \log \mathcal{E}[2^{-\frac{m}{n}\mathbf{r}^{\delta}(l)}] = \frac{1}{n}\alpha(\frac{m}{n}) \tag{2.133}$$

$$C_i = \frac{1}{n} C \tag{2.134}$$

We also adopt $\beta(m)$ according to section 2.6.7:

$$\beta(m) = \frac{1}{m} \log \mathcal{E}[2^{m\mathbf{l}_a^{\delta}(k)}] \tag{2.135}$$

where $\log |\mathbf{a}(k)| = R + \mathbf{l}_a^{\delta}(k)$.

The following Proposition shows that, under the previous assumptions, the necessary condition of Theorem 2.4.1 is not conservative.

Proposition 2.6.6 *Let $\varrho \in \mathbb{N}_+$ be a given constant. If $C - \alpha(\frac{m}{n}) > n\beta(m) + nR$ then there exists constants $\bar{d} > 0$ and $\bar{z}_f^x, \bar{z}_f^u \in [0, 1)$ such that the state-space representation (2.129) can be robustly stabilized in the m-th moment sense.*

Proof: From Proposition 2.6.1, we know that (2.129) can be written in the form (2.98)-(2.99). From section 2.6.7 we know that we can use Theorem 2.6.5 to guarantee that the following is a sufficient condition for the existence of $\bar{d} > 0$ and $\bar{z}_f^x, \bar{z}_f^u \in [0, 1)$ such that (2.129) is robustly stabilizable:

$$C_i > \alpha_i(m) + \beta_i(m) + R_i \tag{2.136}$$

where, in this case, $R_i = R$ and $\beta_i(m) = \beta(m)$ is given by (2.135). On the other hand, by means of (2.133)-(2.134), the assumption $C - \alpha(\frac{m}{n}) > n\beta(m) + nR$ can be written as (2.136). □

Chapter 3

Fundamental Limitations of Disturbance Attenuation in the Presence of Finite Capacity Feedback

3.1 Introduction

The Chapter is organized in 4 sections. Besides the introduction, section 3.2 lays down the problem formulation as well as a preview and a discussion of the results; the limitations resulting from causality are derived in section 3.3 and section 3.4 develops a fundamental limitation that results from finite capacity feedback.

3.1.1 The following notation is adopted:

- Whenever it is clear from the context, we refer to a sequence $\{a(k)\}_{-\infty}^{\infty}$ of elements in $\mathbb{R}^n$ as a. A finite segment of a sequence a is indicated as $a_{k_{min}}^{k_{max}} = \{a(k)\}_{k_{min}}^{k_{max}}$. If $k_{max} < k_{min}$ then $a_{k_{min}}^{k_{max}} = \emptyset$.

- If $\mathcal{O} \subset \mathbb{R}^q$ is a Borel set then we denote its volume by $Vol(\mathcal{O})$

- If M is a matrix then the element in the i-th row and j-th column is indicated as $[M]_{i,j}$. Similarly, if $a \in \mathbb{R}^n$ then $[a]_i$ denotes the i-th component of the vector.

- Random variables are represented using boldface letters, such as $\mathbf{a}$.
- If $\mathbf{a}(k)$ is a stochastic process, then we use $a(k)$ to indicate a specific realization. Similar to the convention used for sequences, we may denote $\mathbf{a}(k)$ just as $\mathbf{a}$ and $a(k)$ as a. A finite segment of a stochastic process is indicated as $\mathbf{a}_{k_{min}}^{k_{max}}$.
- The probability density of a random variable $\mathbf{a}$, if it exists, is denoted as $p_{\mathbf{a}}$. The conditional probability, given $\mathbf{b}$, is indicated as $p_{\mathbf{a}|\mathbf{b}}$.
- The expectation operator over $\mathbf{a}$ is written as $\mathcal{E}[\mathbf{a}]$
- We write $\log_2(.)$ simply as $\log(.)$
- We adopt the convention $0 \log 0 = 0$
- The auto-covariance function of a given stochastic process $\mathbf{a}$ is given by:

$$R_{\mathbf{a}}(k,l) = \mathcal{E}\left[(\mathbf{a}(k) - \mathcal{E}[\mathbf{a}(k)])(\mathbf{a}(l) - \mathcal{E}[\mathbf{a}(l)])^T\right]$$

  If $\mathbf{a}$ is stationary then it's power spectral density is written as

$$\hat{F}_{\mathbf{a}}(\omega) = \sum_{k=-\infty}^{\infty} R_{\mathbf{a}}(k,0)e^{-i\omega k}$$

- If $\mathbf{a}$ is a stochastic process taking values in $\mathbb{R}$ then we use the following covariance matrix:

$$\left[\Sigma\left(\mathbf{a}_{k_{min}}^{k_{max}}\right)\right]_{(i-k_{min}+1),(j-k_{min}+1)} = \mathcal{E}\left[(\mathbf{a}(i) - \mathcal{E}[\mathbf{a}(i)])(\mathbf{a}(j) - \mathcal{E}[\mathbf{a}(j)])\right]$$

  where $i, j \in \{k_{min}, \ldots, k_{max}\}$.
- The Singular Value Decomposition of a matrix $M = M^H \geq 0$ is indicated as $M = V_M^T \Lambda_M V_M$, where the usual ordering of singular values is assumed $[\Lambda_M]_{i+1,i+1} \leq [\Lambda_M]_{i,i}$. The singular values of M are represented in a more streamlined form as $\lambda_i(M) = [\Lambda_M]_{i,i}$. If A is a square matrix, we also represent its eigenvalues as $\lambda_i(A)$.

- If $a \in \mathbb{R}$ then we define the negative and positive parts of a as $[a]_- = \min\{a, 0\}$ and $[a]_+ = \max\{a, 0\}$, respectively.

- The following is a shorthand notation for the log-density of the eigenvalues with magnitude smaller than 1, of a covariance matrix:

$$L_-(\mathbf{a}_{k_{min}}^{k_{max}}) = \frac{1}{k_{max} - k_{min} + 1} \sum_{i=1}^{k_{max}-k_{min}+1} [\log(\lambda_i(\Sigma(\mathbf{a}_{k_{min}}^{k_{max}}))]_-$$

Similarly, we also define the positive counterpart of L_- as:

$$L_+(\mathbf{a}_{k_{min}}^{k_{max}}) = \frac{1}{k_{max} - k_{min} + 1} \sum_{i=1}^{k_{max}-k_{min}+1} [\log(\lambda_i(\Sigma(\mathbf{a}_{k_{min}}^{k_{max}}))]_+$$

3.1.2 Basic Facts and Definitions of Information Theory

In this section, we summarize the main definitions and facts about Information Theory which are used throughout the Chapter. We adopt [52], as a primary reference, because it addresses general probabilistic spaces in a unified framework. Let $(\Omega, \mathcal{S}_\omega, \mathcal{P}_\omega)$ be a probability space along with the random variables a, b and c, taking values in the measurable spaces $(\mathcal{A}, \mathcal{S}_a)$, $(\mathcal{B}, \mathcal{S}_b)$ and $(\mathcal{C}, \mathcal{S}_c)$. We define mutual information and conditional mutual information, between any two random variables, as:

Definition 3.1.1 *(from [52] pp. 9) The mutual information, between* a *and* b, $I : (\mathbf{a}; \mathbf{b}) \to \mathbf{R}_+ \bigcup \{\infty\}$ *is given by:*

$$I(\mathbf{a}; \mathbf{b}) = \sup \sum_{ij} \mathcal{P}_{\mathbf{a},\mathbf{b}}(E_i \times F_j) \log \frac{\mathcal{P}_{\mathbf{a},\mathbf{b}}(E_i \times F_j)}{\mathcal{P}_{\mathbf{a}}(E_i)\mathcal{P}_{\mathbf{b}}(F_j)}$$

where the supremum is taken over all partitions $\{E_i\}$ of $\mathcal{A}$ and $\{F_j\}$ of $\mathcal{B}$.

Definition 3.1.2 *(from [52] pp. 37) The conditional mutual information* $I : (\mathbf{a}; \mathbf{b}|\mathbf{c}) \to \mathbf{R}_+ \bigcup \{\infty\}$, *between* a *and* b *given* c, *is defined as:*

$$I(\mathbf{a}; \mathbf{b}|\mathbf{c}) = \sup \sum_{ijk} \mathcal{P}_{\mathbf{a},\mathbf{b},\mathbf{c}}(E_i \times F_j \times N_k) \log \frac{\mathcal{P}_{\mathbf{a},\mathbf{b},\mathbf{c}}(E_i \times F_j \times N_k)}{\bar{\mathcal{P}}_{\mathbf{a},\mathbf{b}|\mathbf{c}}(E_i \times F_j \times N_k)}$$

where the supremum is taken over all partitions $E_i \in \mathcal{A}$, $F_j \in \mathcal{B}$ *and* $N_k \in \mathcal{C}$ *and* $\bar{\mathcal{P}}_{\mathbf{a},\mathbf{b}|\mathbf{c}}$ *is given by:*

$$\bar{\mathcal{P}}_{\mathbf{a},\mathbf{b}|\mathbf{c}}(E \times F \times N) = \int_N \mathcal{P}_{\mathbf{a}|\mathbf{c}}(E|\gamma)\mathcal{P}_{\mathbf{b}|\mathbf{c}}(F|\gamma)P_{\mathbf{c}}(d\gamma)$$

Notice that, in definition 3.1.1, $\mathcal{A}$ and $\mathcal{B}$ may be different.

Consistent with the usual notation [52, 20], we define entropy as:

Definition 3.1.3 *(Entropy) Let* a *and* b *be random variables. The entropy of* a *given* b *is defined as:*

$$H(\mathbf{a}|\mathbf{b}) = I(\mathbf{a};\mathbf{a}|\mathbf{b})$$

Since entropy may be infinite for random variables defined in continuous probability spaces, we also define the following quantities, denoted as differential entropy and conditional differential entropy.

Definition 3.1.4 *If* a *is a random variable, where* $\mathcal{A} = \mathbb{R}^q$*, along with a Lebesgue measurable and bounded probability density function* $p_a(\cdot)$ *then we define the differential entropy of* a *as:*

$$h(\mathbf{a}) = \int_{p_a(\gamma)\leq 1} -p_a(\gamma)\log p_a(\gamma)d\gamma - \int_{p_a(\gamma)>1} p_a(\gamma)\log p_a(\gamma)d\gamma$$

Notice that if p_a *is Lebesgue measurable and bounded then we use* $\int_{p_a(\gamma)>1} p_a(\gamma) \leq 1$ *to assert that* $\int_{p_a(\gamma)>1} p_a(\gamma)\log p_a(\gamma)d\gamma < \infty$*. This implies that* $h(a)$ *is always well defined, although not necessarily bounded. We have just shown that a bounded Lebesgue measurable* p_a *leads to an* almost-*integrable*[1] $p_a(\gamma)\log(p_a(\gamma))$*. If* b *is another random variable and* $I(\mathbf{a},\mathbf{b}) < \infty$ *then the conditional differential entropy of* a *given* b *is defined by:*

$$h(\mathbf{a}|\mathbf{b}) = h(\mathbf{a}) - I(\mathbf{a};\mathbf{b}) \tag{3.1}$$

For technical reasons, we also define the following class of random variables:

[1]According to [26], all the properties of Lebesgue integrable functions hold for *almost*-integrable functions. A Lebesgue measurable function f on (μ, X) is *almost* integrable [26] if at least one of the following holds: $\int_X [f]_+\mu(dx) < \infty$ or $\int_X [-f]_+\mu(dx) < \infty$. The integral of an *almost* integrable f is defined as $\int_X f\mu(dx) = \int_X [f]_+\mu(dx) - \int_X [-f]_+\mu(dx)$. This issue is also briefly discussed in pp. 200 of [9].

Definition 3.1.5 *(Dither Class denoted as $\mathbb{D}$) Let* b *be a random variable with alphabet* $\mathcal{B} \subset \mathrm{R}^q$*, for some* $q \in \mathbb{N}$*. We denote* b *as type 1 if it has a countable alphabet with* $\inf\{\max_i |b_i - \tilde{b}_i| : b, \tilde{b} \in \mathcal{B}, b \neq \tilde{b}\} > 0$ *and type 2 if it has a probability density* p_b *which is Lebesgue measurable. The random variable* b *is of the class Dither, denoted as* $\mathbb{D}$*, if it is type 1 or type 2. If* $\mathrm{b} \in \mathbb{D}$ *then we also define* $\breve{\mathrm{b}}$ *as:*

$$\breve{\mathrm{b}} = \begin{cases} \mathrm{b} + \Delta \mathrm{s} & \text{if } \mathrm{b} \text{ is type 1} \\ \mathrm{b} & \text{if } \mathrm{b} \text{ is type 2} \end{cases} \tag{3.2}$$

where s *is independent of* b *and uniformly distributed in* $(-1/2, 1/2)^q$ *and* Δ *is given by:*

$$\Delta = \inf\{\max_i |b_i - \tilde{b}_i| : b, \tilde{b} \in \mathcal{B}, b \neq \tilde{b}\} \tag{3.3}$$

Notice that Δ *is such that the following projections always exist:*

$$b = \pi_{\mathcal{B}}(\breve{b})$$

$$\pi_{\mathcal{S}}(\breve{b}) = \begin{cases} \mathrm{s} & \text{if } \mathrm{b} \text{ is type 1} \\ 0 & \text{if } \mathrm{b} \text{ is type 2} \end{cases}$$

The following is a list of properties used in the sections 3.3 and 3.4. The proof of such properties may be found in [52] and, in some cases, in [20]. We emphasize that, in this Chapter, we write $h(\mathrm{a}|\cdot)$ only if the assumptions stated in definition 3.1.4 are satisfied.

- **(P1):** $I(\mathrm{a};\mathrm{b}) = I(\mathrm{b};\mathrm{a}) \geq 0$ and $I(\mathrm{a};\mathrm{b}|\mathrm{c}) = I(\mathrm{b};\mathrm{a}|\mathrm{c}) \geq 0$

- **(P2) Kolmogorov's formula** [2] **(equation 3.6.6 in [52]):**

$$I((\mathrm{a},\mathrm{b});\mathrm{c}|\mathrm{d}) = I(\mathrm{b};\mathrm{c}|\mathrm{d}) + I(\mathrm{a};\mathrm{c}|(\mathrm{b},\mathrm{d}))$$

- **(P3) Theorem 3.7.1 in [52]:** If f and g are measurable functions then $I(f(\mathrm{a}); g(\mathrm{b})|\mathrm{c}) \leq$

[2] Notice that equation 3.6.3 in [52] has a typographic mistake. On the left hand side of the equality, the correct is $I(\xi, \zeta)$

$I(\mathbf{a};\mathbf{b}|\mathbf{c})$ and equality holds[3] if f and g are invertible.

- **(P3')** It follows from (P3) that if $\mathbf{b} \in \mathbb{D}$ then $I(\mathbf{a};\breve{\mathbf{b}}|\mathbf{c}) = I(\mathbf{a};(\mathbf{b},\mathbf{s})|\mathbf{c})$ which, since $\mathbf{s}$ is independent from the rest, also implies that $I(\mathbf{a};\breve{\mathbf{b}}|\mathbf{c}) = I(\mathbf{a};\mathbf{b}|\mathbf{c})$. A similar argument, using $(P2)$, also leads to $I(\mathbf{a};\mathbf{b}|\mathbf{c}) = I(\mathbf{a};\mathbf{b}|\breve{\mathbf{c}})$, provided that $\mathbf{c} \in \mathbb{D}$.

- **(P4) Corollary 2., pp. 43 in [52]:** Given a function $f : \mathcal{C} \rightarrow \mathcal{C}'$ it follows that $I(\mathbf{a}; f(\mathbf{c})|\mathbf{c}) = 0$.

- **(P5):** From property (P3), we conclude that $I(\mathbf{a};(\mathbf{b},\mathbf{c})|\mathbf{d}) = I(\mathbf{a};(\mathbf{b}-\mathbf{c},\mathbf{c})|\mathbf{d})$. Using (P2), such equality also leads to:

$$I(\mathbf{a};\mathbf{b}|(\mathbf{c},\mathbf{d})) = I(\mathbf{a};\mathbf{b}-\mathbf{c}|(\mathbf{c},\mathbf{d}))$$

- **(P6):** By means of (P1) and (3.1), we infer that $h(\mathbf{a}) \geq h(\mathbf{a}|\mathbf{b})$, where equality holds if and only if $\mathbf{a}$ and $\mathbf{b}$ are independent. Likewise, we can use properties (P1)-(P2) to state that $I(\mathbf{a};(\mathbf{b},\mathbf{c})) \geq I(\mathbf{a};\mathbf{b})$, which can be used with (3.1) to derive $h(\mathbf{a}|\mathbf{b}) \geq h(\mathbf{a}|(\mathbf{b},\mathbf{c}))$.

- **(P7) [20]:** Let $\mathbf{a}$, with $\mathcal{A} = \mathbb{R}^q$, be a random variable with p_a bounded and Lebesgue measurable and a covariance matrix denoted as Σ_a. In Proposition C.1.1, of Appendix C, we show that finite Σ_a implies finite $h(\mathbf{a})$. Under such hypothesis, $p_a(\gamma)\log(p_a(\gamma))$ is Lebesgue integrable and the following holds[20]:

$$h(\mathbf{a}) \leq \frac{1}{2}\log((2\pi e)^n \det(\Sigma_a))$$

  where equality holds if $\mathbf{a}$ is Gaussian.

In order to simplify our notation, we also define the following quantities:

Definition 3.1.6 *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be stochastic processes. The following are useful limit infor-*

[3]The general version of this property states that equality holds if f and g are everywhere dense [52]. Every-time we use an invertible function to claim equality in (P3) the function is everywhere dense.

mation rates:

$$\bar{I}_{\infty}(\mathbf{a};\mathbf{b}) = \limsup_{k\to\infty} \frac{I(\mathbf{a}_1^k;\mathbf{b}_1^k)}{k},\ \bar{I}_{\infty}(\mathbf{a}\to\mathbf{b}) = \limsup_{k\to\infty} \frac{I(\mathbf{a}_1^k\to\mathbf{b}_1^k)}{k}$$

where $I(\mathbf{a}_1^k \to \mathbf{b}_1^k)$ *is denoted as directed mutual information [47, 60] and is defined as:*

$$I(\mathbf{a}_1^k \to \mathbf{b}_1^k) = \sum_{i=1}^{k} I(\mathbf{a}_1^i;\mathbf{b}(i)|\mathbf{b}_1^{i-1})$$

In this Chapter, we will also refer to Channels which are stochastic operators conforming to the following definition:

Definition 3.1.7 *(Memory-less Channel) Let* $\mathcal{V}$ *and* $\mathcal{Z}$ *be given input and output alphabets, along with a white stochastic process, denoted as* c, *with alphabet* $\mathcal{C}$. *Consider also a function* $f:\mathcal{V}\times\mathcal{C}\to\mathcal{Z}$ *such that the following functions are invertible:*

$$g_1(v(k),c(k)) = (v(k), f(v(k),c(k)))$$

$$g_2(v(k),c(k)) = (f(v(k),c(k)), c(k))$$

The pair $(f,\mathbf{c})$ *defines a memory-less channel.*

The previous definition is sufficiently general to encompass the following examples:

- **Additive white Gaussian channel:** $\mathcal{V}=\mathcal{Z}=\mathcal{C}=\mathbb{R}$, c is an i.i.d. white Gaussian sequence with unit variance and $f(c,v)=c+v$.
- **Binary symmetric channel, with error probability** p_e: $\mathcal{V}=\mathcal{Z}=\mathcal{C}=\mathbb{Z}_2=\{0,1\}$, is an i.i.d sequence satisfying $\mathcal{P}(\mathbf{c}(k)=1)=p_e$ and $f(c,v)=c+_{mod2}v$

3.2 Problem Formulation and Discussion of Results

Consider the feedback interconnection depicted in Figure 1-2. In such information pattern [65], measurements of the state of the plant have to be encoded and sent over a communication channel. The transmitted information is used, at the decoder/controller, to generate

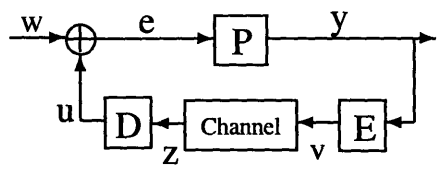

Figure 3-1: Simplified Structure of the Feedback Interconnection using $e = G^{-1}\tilde{e}$ and $u = G^{-1}\tilde{u}$, with the correspondence, relative to the blocks of Fig 1-2.

the control signal **u**. In order to simplify the presentation, we proceed with the equivalent block diagram of Fig 3-1.

3.2.1 Assumptions

Before stating our assumptions, we need the following definitions:

Definition 3.2.1 *We define the following set of probability densities:*

$$\bar{\mathbb{L}}^q = \{f : \mathbb{R}^q \to \mathbb{R}_{\geq 0} | f \text{ is Leb. meas.}, \int f(\gamma)d\gamma = 1, \int f(\gamma)\gamma^T\gamma d\gamma < \infty, \sup_\gamma f(\gamma) < \infty\} \tag{3.4}$$

In addition, we also define:

$$\mathbb{L}^q = \{f \in \bar{\mathbb{L}}^q | \exists \epsilon > 0, \text{ such that } \Gamma_f^\epsilon \text{ has limited interior}\} \tag{3.5}$$

where

$$\Gamma_f^\epsilon = \left\{\gamma \in \mathbb{R}^q : f(\gamma) > \frac{1}{((1+|\gamma_1|)\cdots(1+|\gamma_q|))^{1+\epsilon}}\right\}$$

An important property of $\mathbb{L}^q$ is that if $p_{a,b} \in \mathbb{L}^{q_a+q_b}$ then $p_a \in \bar{\mathbb{L}}^{q_a}$ and $p_b \in \bar{\mathbb{L}}^{q_b}$.

In the present formulation, which is schematically depicted in Fig 3-1, the following assumptions are made. Notice that the diagram in Fig 3-1 is derived from Fig 1-2, by means of incorporating G in the plant and G^{-1} in the decoder. The Appendix A comprises a discussion of several important aspects related to the assumptions made.

We adopt the following assumptions:

- **(A1):** w, with $w(k) \in \mathbb{R}$, is an i.i.d., zero mean, unit variance and white Gaussian process.

- **(A2):** the control signal satisfies $\mathbf{u}_1^k \in \mathbb{D}$ for every k. We denote the alphabet of $\mathbf{u}$ as $\mathcal{U} \subset \mathbf{R}$, so that $u_1^k \in \mathcal{U}^k$. According to definition 3.1.5, we indicate the dithered version of $\mathbf{u}(k)$ as $\breve{\mathbf{u}}(k)$.

- **(A3):** $G(z)$ is an all-pole stable filter of the form:

$$G(z) = \frac{\alpha}{1 - \sum_{m=1}^{p} a_m z^{-m}}$$

for some integer $p \geq 1$ and constants a_i and $\alpha > 0$.

- **(A4):** given n, P is a single input plant with state $x(k) \in \mathbb{R}^n$, which satisfies the following state-space equation:

$$\mathbf{x}(k+1) = \begin{bmatrix} x_u(k+1) \\ x_s(k+1) \end{bmatrix} = \begin{bmatrix} A_u & 0 \\ 0 & A_s \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} b_u \\ b_s \end{bmatrix} \mathrm{e}(k) \tag{3.6}$$

$$\mathbf{y}(k) = C\mathbf{x}(k), |\lambda_i(A_u)| \geq 1, |\lambda_i(A_s)| < 1$$

The state partitions $\mathbf{x}_u$ and $\mathbf{x}_s$ represent the unstable and stable open-loop dynamics, respectively. In addition, if $A \neq A_s$ then $\mathbf{x}_u(k)$ is a random variable, with a given probability density $p_{x_u(k)}(\cdot)$.

- **(A5):** the capacity [20] of a channel, specified by $(f, \mathbf{c})$, is denoted as $C_{channel}$ and is defined as:

$$C_{channel} = \sup_{P_v} I(f(\mathbf{v}(k), \mathbf{c}); \mathbf{v}(k)) < \infty$$

where the supremum is taken over all probability measures P_v, defined in $(\mathcal{V}, \mathcal{S}_v)$.

- **(A6):** the encoder and the decoder are causal operators defined in the appropriate spaces, i.e., $E : \mathcal{Y}^\infty \to \mathcal{V}^\infty$, $D : \mathcal{Z}^\infty \to \mathcal{U}^\infty$ where $\mathbf{v}(k) = f_k^e(\mathbf{y}_{-\infty}^k)$ and $\mathbf{u}(k) = f_k^d(\mathbf{z}_{-\infty}^k)$ for some functions f_k^e and f_k^d.

- **(A7):** additionally, the decoder satisfies the following finite memory condition:

$$\forall k > \alpha, \mathbf{u}_{1+\alpha}^k = \tilde{f}_k^d(\mathbf{u}_1^\alpha, \mathbf{z}_1^k) \tag{3.7}$$

for some $\alpha \in \mathbb{N}_+$ and a sequence of functions $\tilde{f}_k^d : \mathcal{U}^\alpha \times \mathcal{Z}^k \to \mathcal{U}^{k-\alpha-1}$.

- **(A8):**(Fading memory condition) For technical reasons, we assume that the following condition holds:

$$\limsup_{k\to\infty} \frac{1}{k} I(\mathbf{u}_1^\alpha; \mathbf{x}(1), \mathbf{w}_1^k | \mathbf{z}_1^k) = 0$$

where α is the smallest constant for which (A7) holds. If $\alpha = 0$ then we adopt the convention that (A8) is satisfied. Several aspects of this assumption are clarified in Appendix A. In particular, this condition is automatically satisfied if $\mathcal{U}$ is countable and $H(\mathbf{u}_1^\alpha) < \infty$ holds.

- **(A9):** We assume that, for each k, $(\mathbf{w}_1^k, \breve{\mathbf{u}}_1^k, \mathbf{x}(1))$ admits a probability distribution satisfying $p_{w_1^k, \breve{u}_1^k, x(1)} \in \mathbb{L}^{2k+1}$. In Appendix A, we explore a few special cases related to this assumption.

3.2.2 Problem Statement and Discussion of Results

We investigate the fundamental limitations of the eigenvalue distribution of $\Sigma(\mathbf{e}_1^k)$. In order to simplify the exposé, we state our results in terms of $L_-(\mathbf{e}_1^k)$ and $L_+(\mathbf{e}_1^k)$.

In section 3.3 we reach a fundamental limitation which is a consequence from causality alone. The result is presented in theorem 3.3.3, which states that if the feedback system in Fig 3-1 is stable then the following must hold:

$$\frac{1}{2} \liminf_{k\to\infty} \left(L_-(\mathbf{e}_1^k) + L_+(\mathbf{e}_1^k) \right) \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \tag{3.8}$$

The inequality in (3.8) demonstrates that not all of the eigenvalues, of $\Sigma(\mathbf{e}_1^k)$, can be made small and that the reduction of some necessarily imply the increase of others. That is comparable to the water-bed effect, associated to the classic Bode integral limitation. Such comparison is not coincidental and is explored in section 3.3.1.

In the fundamental limitation expressed in (3.8), the characteristics of the channel do not play a role. It remains the question of whether the "shaping" of the eigenvalues of $\Sigma(\mathbf{e}_1^k)$ depends on the information flow in the feedback loop. The answer is given in theorem 3.4.3

which states that:

$$\frac{1}{2}\liminf_{k\to\infty} L_-(\mathbf{e}^k_{g(k)}) + \bar{I}_\infty(\mathbf{v}\to\mathbf{z}) \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \tag{3.9}$$

where $g : \mathbb{N}_+ \to \mathbb{N}_+$ is any arbitrary function satisfying:

$$\lim_{k\to\infty} \frac{g(k)}{k} = 0 \tag{3.10}$$

As a consequence of (3.9), we find that reduction of the eigenvalues of $\Sigma(\mathbf{e}^k_{g(k)})$, for values bellow unity, must come at the expense of information flow in the channel, as quantified by $\bar{I}_\infty(\mathbf{v}\to\mathbf{z})$.

Under stationary assumptions, corollaries 3.3.4 and 3.4.4 show that the inequalities (3.8) and (3.9) can be expressed as:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} [\log(S(\omega))]_- d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi} [\log(S(\omega))]_+ d\omega \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \tag{3.11}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} [\log(S(\omega))]_- d\omega + \bar{I}_\infty(\mathbf{v}\to\mathbf{z}) \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \tag{3.12}$$

where $S(\omega) = \sqrt{\hat{F}_{\mathbf{e}}(\omega)} = \sqrt{\frac{\hat{F}_{\tilde{\mathbf{e}}}(\omega)}{|G(e^{j\omega})|^2}}$.

The inequalities (3.11) and (3.12) must be satisfied by any stable and causal loop of the form depicted in Fig 1-2 or Fig 3-1. The first inequality is the Bode integral formula, which is the basis of the disturbance attenuation/amplification water-bed effect, while the second entails a new attenuation/capacity trade-off.

3.3 Fundamental Limitations Created by Causality

In this section, we derive a fundamental limitation that arises from causality. The results are valid under the assumptions listed in section 3.2.1, with the *exception of (A7)-(A8)* which are not needed. The discussion is also used to present some of the preliminary results, which will be used in section 3.4. Our technique follows the one by [71], with the exception of the way we tackle initial conditions and unstable modes of the plant. More specifically, theorem 3.3.3 states a fundamental limitation that explicitly incorporates the eigenvalues of A. At the end of the section, we specialize the result, under stationarity assumptions, and derive the Bode-Integral formula in Corollary 3.3.4.

The following lemma shows that the difference, between the entropy rate of e and the entropy rate of w, is lower-bounded by the mutual information between the plant's state and e.

Lemma 3.3.1 *(Entropy-rate amplification) If* $\mathbf{x}(k)$ *is the solution of the state-space equation (3.6) then the following holds:*

$$\liminf_{k\to\infty} \frac{h(\mathbf{e}_1^k)}{k} \geq \liminf_{k\to\infty} \frac{I(\mathbf{e}_1^k;\mathbf{x}(1))}{k} + h(\mathbf{w}(1)) \tag{3.13}$$

Proof: We start by noticing that, since the plant is strictly proper and causal, $\mathbf{w}(k)$, with $k \geq 1$, is independent of $(\mathbf{x}(1), \mathbf{u}_1^k, \mathbf{w}_1^{k-1})$, which implies:

$$h(\mathbf{w}(k)) = h(\mathbf{w}(k)|\mathbf{x}(1), \mathbf{u}_1^k, \mathbf{w}_1^{k-1}) = h(\mathbf{e}(k)|\mathbf{x}(1), \mathbf{u}_1^k, \mathbf{e}_1^{k-1}) \leq h(\mathbf{e}(k)|\mathbf{x}(1), \mathbf{e}_1^{k-1}), k \geq 1 \tag{3.14}$$

where we used properties (P6) and lemma C.2.1 of Appendix C. Since $h(\mathbf{w}(k))$ does not depend on k, we use (3.14) and the chain rule of differential entropy to derive:

$$\sum_{i=1}^{k} h(\mathbf{e}(i)|\mathbf{x}(1), \mathbf{e}_1^{i-1}) = h(\mathbf{e}_1^k|\mathbf{x}(1)) \geq h(\mathbf{w}(1)) \tag{3.15}$$

Notice that the chain rule of differential entropy in (3.15) is valid because $p_{w_1^k, u_1^k, x(1)} \in \mathbb{L}^{2k+1}$ implies, using a change of variables and an integration argument, that the marginal densities[4] $p_{e_1^i, x(1)} \in \bar{\mathbb{L}}^{i+1}$. Consequently, the proposition C.1.1 and the lemma C.1.2, of Appendix C, guarantee that all the quantities in (3.15) are well defined. The proof is concluded once we notice, from (3.1), that $h(\mathbf{e}_1^k|\mathbf{x}(1)) = h(\mathbf{e}_1^k) - I(\mathbf{e}_1^k;\mathbf{x}(1))$. □

The following lemma, corroborates the results by [59, 60, 54, 69, 50], and unveils that

[4] Notice that Fubinni's Theorem [2] guarantees that the marginal densities are Lebesgue measurable. The co-variance matrix of $(e_1^i, x(1))$ is bounded since the covariance matrix of $(w_1^k, u_1^k, x(1))$ is bounded. The integration and change of variables are need just to show that $p_{e_1^i, x(1)}$ is bounded.

stability implies that e must carry a bit-rate, of information about the state of the plant, of at least $\sum_i \max\{0, \log(|\lambda_i(A)|)\}$.

Lemma 3.3.2 *Let* $\mathbf{x}(k)$ *be the solution of the state-space equation (3.6). If the plant is stabilized, i.e.,* $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ *holds then the following is satisfied:*

$$\liminf_{k\to\infty} \frac{I(\mathbf{e}_1^k; \mathbf{x}(1))}{k} \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \tag{3.16}$$

Proof: If $A = A_s$ then we just use $I(\mathbf{e}_1^k; \mathbf{x}(1)) \geq 0$. If $A \neq A_s$ then we consider the following homogeneous system:

$$\mathbf{x}_e(k+1) = A_u \mathbf{x}_e(k) + b_u \mathbf{e}(k),\ x_e(1) = 0 \tag{3.17}$$

and define the estimate $\hat{\mathbf{x}}(k) = A_u^{-k}\mathbf{x}_e(k)$. Since $\mathbf{x}_u(k) = \mathbf{x}_e(k) + A_u^k \mathbf{x}_u(1) = A_u^k(\hat{\mathbf{x}}(k) - \mathbf{x}_u(1))$, we know that:

$$k\log(|\det(A_u A_u^T)|) + \log(\det(R_{\mathbf{x}_{error}}(k))) = \log(\det(R_{\mathbf{x}_u}(k,k))) < \beta < \infty \tag{3.18}$$

where $\mathbf{x}_{error}(k) = \hat{\mathbf{x}}(k) - \mathbf{x}_u(1)$. Since $\hat{x}(k)$ is a function of e_1^k, we have that:

$$I(\mathbf{x}(1); \mathbf{e}_1^k) \geq I(\mathbf{x}_u(1); \mathbf{e}_1^k) \geq h(\mathbf{x}_u(1)) - h(\mathbf{x}_u(1)|\mathbf{e}_1^k) = h(\mathbf{x}_u(1)) - h(\mathbf{x}_u(1) - \hat{\mathbf{x}}(k)|\mathbf{e}_1^k) \geq h(\mathbf{x}_u(1)) - h(\hat{\mathbf{x}}(k) - \mathbf{x}_u(1)) \tag{3.19}$$

where we have used (P3), (3.1), lemma C.2.3 of Appendix C and (P6).

But, from (P7) we know that $\limsup_{k\to\infty} \frac{h(\hat{\mathbf{x}} - \mathbf{x}_u(1))}{k} \leq \limsup_{k\to\infty} \frac{\log(\det(R_{\mathbf{x}_{error}}(k)))}{2k}$. As a consequence, we can use (3.18) to get:

$$\limsup_{k\to\infty} \frac{h(\hat{\mathbf{x}}(k) - \mathbf{x}_u(1))}{k} \leq -\log(|\det(A_u)|) \tag{3.20}$$

The proof follows by direct substitution $\square$.

Using the results in the previous lemmas, we derive theorem 3.3.3. It states that causality and stability imply that the log-sum of the eigenvalues of $\Sigma(\mathbf{e}_1^k)$ are, in the limit, lower bounded by the unstable eigenvalues of the plant.

Theorem 3.3.3 ***(Causality fundamental limitation)*** *Let* $\mathbf{x}(k)$ *be the solution of the state-space equation (3.6). If the plant is stabilized, i.e.,* $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ *holds then the following is satisfied:*

$$\liminf_{k\rightarrow\infty} \left(L_-(\mathbf{e}_1^k) + L_+(\mathbf{e}_1^k)\right) \geq 2\sum_i \max\{0, \log(|\lambda_i(A)|)\} \tag{3.21}$$

Proof: From lemmas 3.3.1 and 3.3.2 we know that:

$$\liminf_{k\rightarrow\infty} \frac{h(\mathbf{e}_1^k)}{k} - h(\mathbf{w}(1)) \geq \liminf_{k\rightarrow\infty} \frac{I(\mathbf{e}_1^k; \mathbf{x}(1))}{k} \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \tag{3.22}$$

Using the fact (P7), we conclude that $h(\mathbf{e}_1^k) - kh(\mathbf{w}(1)) \leq \frac{1}{2}\log(\det(\Sigma(\mathbf{e}_1^k)))$ which, together with (3.22), leads to the final result □

3.3.1 Deriving Bode's Integral Formula

Under stationarity assumptions, theorem 3.3.3 is at the base of the Bode-integral formula. A precise description of such property is in the subsequent Corollary.

Corollary 3.3.4 *Let* $\mathbf{x}(k)$ *be the solution of the state-space equation (3.6). If the plant is stabilized, i.e.,* $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ *holds and* $\mathbf{e}$ *is a stationary process, where* $0 < m < \hat{F}_e(\omega) < M < \infty$ *is Lebesgue integrable, then the following is satisfied:*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log(S(\omega))d\omega \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \tag{3.23}$$

where $S(\omega) = \sqrt{\hat{F}_e(\omega)} = \sqrt{\frac{\hat{F}_{\tilde{\mathbf{e}}}(\omega)}{|G(e^{j\omega})|^2}}$*. The processes* $\tilde{\mathbf{e}}$ *and* $\mathbf{d}$ *are the ones depicted in Fig 1-2.*

Proof: From theorem B.2.1, of Appendix B, we have that:

$$\lim_{k\rightarrow\infty} L_-(\mathbf{e}_1^k) + L_+(\mathbf{e}_1^k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log(\hat{F}_e(\omega))d\omega \tag{3.24}$$

The proof follows by means of (3.24) and Theorem 3.3.3□

3.4 Fundamental Limitations Created by Finite Capacity Feedback

In this section, we examine the fundamental limitations, in the eigenvalues of[5] $\Sigma(\mathbf{e}^k_{g(k)})$, that originate from the constraint $\bar{I}_\infty(\mathbf{v} \to \mathbf{z}) \leq C_{channel}$. The main inequality, involving the channel directed information rate and the eigenvalues of A, is given in theorem 3.4.3.

Sub-sequentially, we provide a lemma which unveils how the information flux is allocated in the feedback loop. We identify that the directed information rate in the channel must account for two terms. The first is due to the stabilization information and is given by $I(\mathbf{x}(1); \mathbf{e}^k_1)$; while the second represents the interaction between the control signal and the disturbance and is quantified by $I(\mathbf{u}^k_1; \mathbf{w}^k_1)$.

Lemma 3.4.1 ***(Fundamental Lemma of the Flux of Information)*** *If* $\mathbf{x}(k)$ *is the solution of the state-space equation (3.6) then the following holds:*

$$\bar{I}_\infty(\mathbf{v} \to \mathbf{z}) \geq \liminf_{k\to\infty} \frac{1}{k} I(\mathbf{x}(1); \mathbf{e}^k_1) + \bar{I}_\infty(\mathbf{u}^k_1; \mathbf{w}^k_1) \tag{3.25}$$

Proof: We start by using (P2) to write $I((\mathbf{x}(1), \mathbf{w}^k_1); \mathbf{u}^k_1) = I(\mathbf{x}(1); \mathbf{u}^k_1|\mathbf{w}^k_1) + I(\mathbf{u}^k_1; \mathbf{w}^k_1)$ which can be rewritten as:

$$I((\mathbf{x}(1), \mathbf{w}^k_1); \mathbf{u}^k_1) = I(\mathbf{x}(1); \mathbf{e}^k_1|\mathbf{w}^k_1) + I(\mathbf{u}^k_1; \mathbf{w}^k_1) \tag{3.26}$$

where we used (P5) to establish that $I(\mathbf{x}(1); \mathbf{u}^k_1|\mathbf{w}^k_1) = I(\mathbf{x}(1); \mathbf{e}^k_1|\mathbf{w}^k_1)$. On the other hand, using (P2) we get

$$I(\mathbf{x}(1); \mathbf{e}^k_1|\mathbf{w}^k_1) = I(\mathbf{x}(1); \mathbf{e}^k_1) - I(\mathbf{x}(1); \mathbf{w}^k_1) + I(\mathbf{x}(1); \mathbf{w}^k_1|\mathbf{e}^k_1) \tag{3.27}$$

[5]We investigate the eigenvalues of $\Sigma(\mathbf{e}^k_{g(k)})$, for arbitrary g satisfying $\lim_{k\to\infty} \frac{g(k)}{k} = 0$. Such generalization allows the derivation of an integral formula for exponentially asymptotic stationary processes. More details are provided in section 3.5.1

Since $\mathbf{w}$ is independent from $\mathbf{x}(1)$, the second term, on the right-hand side of (3.27), vanishes and we resort to (P1) to get $I(\mathbf{x}(1);\mathbf{e}_1^k|\mathbf{w}_1^k) \geq I(\mathbf{x}(1);\mathbf{e}_1^k)$. Consequently, we substitute the aforementioned inequality in (3.26) and obtain the following:

$$I((\mathbf{x}(1),\mathbf{w}_1^k);\mathbf{u}_1^k) \geq I(\mathbf{x}(1);\mathbf{e}_1^k) + I(\mathbf{u}_1^k;\mathbf{w}_1^k) \tag{3.28}$$

The final inequality follows from (3.28) and the Theorem B.1.1 of Appendix B. □

The following lemma suggests that attenuation can happen only if the channel conveys information about the disturbance.

Lemma 3.4.2 *The following holds:*

$$\frac{1}{k-k_0+1} I(\mathbf{u}_{k_0}^k;\mathbf{w}_{k_0}^k) \geq -\frac{1}{2} L_-(\mathbf{e}_{k_0}^k) \tag{3.29}$$

Proof: Let the following be the singular value decomposition of $\Sigma(\mathbf{e}_{k_0}^k)$:

$$\Sigma(\mathbf{e}_{k_0}^k) = \begin{bmatrix} V_+ \\ V_- \end{bmatrix}^T \begin{bmatrix} \Lambda_+ & 0 \\ 0 & \Lambda_- \end{bmatrix} \begin{bmatrix} V_+ \\ V_- \end{bmatrix} \tag{3.30}$$

where $[\Lambda_-]_{ii} < 1$ and $[\Lambda_+]_{ii} \geq 1$.

As a second step, we establish the following relation[6]:

$$I(\mathbf{w}_{k_0}^k;\mathbf{u}_{k_0}^k) \geq I(V_-\mathbf{w}_{k_0}^k;V_-\mathbf{u}_{k_0}^k) = h(V_-\mathbf{w}_{k_0}^k) - h(V_-\mathbf{w}_{k_0}^k|V_-\mathbf{u}_{k_0}^k) \geq h(V_-\mathbf{w}_{k_0}^k) - h(V_-\mathbf{e}_{k_0}^k) \tag{3.31}$$

where we have used (P3), (3.1), lemma C.2.2 of Appendix C and (P6). Moreover, since $\mathbf{w}$ is i.i.d, $h(\mathbf{w}(k)) = \log(2\pi e)$ and V is unitary, we use (P7) to derive:

$$h(V_-\mathbf{w}_{k_0}^k) - h(V_-\mathbf{e}_{k_0}^k) \geq -\frac{1}{2}\log(\det(V_-\Sigma(\mathbf{e}_{k_0}^k)V_-^T)) = -\frac{k-k_0+1}{2} L_-(\mathbf{e}_{k_0}^k) \tag{3.32}$$

[6]Notice that we have used an abuse of notation in equation (3.31). We write $V_-\mathbf{e}_{k_0}^k$ to indicate the random variable whose realizations are computed as $V_- \begin{bmatrix} e(k) \\ \vdots \\ e(k_0) \end{bmatrix}$

□

Subsequently, we provide the theorem which states the main inequality in the Chapter. It reflects a trade-off between disturbance attenuation, as measured by $L_-(\mathbf{e}_{g(k)}^k)$, and the directed information rate through the channel, expressed by $\bar{I}_\infty(\mathbf{v} \to \mathbf{z})$.

Theorem 3.4.3 *(Main theorem) Let* $\mathbf{x}(k)$ *be the solution of the state-space equation (3.6) and* $g : \mathbb{N}_+ \to \mathbb{N}_+$ *be an arbitrary function satisfying*

$$\lim_{k\to\infty} \frac{g(k)}{k} = 0$$

If the plant is stabilized, i.e., $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ *then the following is satisfied:*

$$\bar{I}_\infty(\mathbf{v} \to \mathbf{z}) - \sum_i \max\{0, \log(|\lambda_i(A)|)\} \geq -\frac{1}{2} \liminf_{k\to\infty} L_-(\mathbf{e}_{g(k)}^k) \tag{3.33}$$

Proof: We begin by using (P3) to arrive at the following fact:

$$\bar{I}_\infty(\mathbf{u}; \mathbf{w}) \geq \limsup_{k\to\infty} \frac{1}{k - g(k) + 1} I(\mathbf{u}_{g(k)}^k; \mathbf{w}_{g(k)}^k) \tag{3.34}$$

The proof follows by substituting the results of lemmas 3.3.2 and 3.4.2 into lemma 3.4.1.

□

The corollary bellow is an immediate consequence of theorem 3.4.3 and shows that if $C_{channel}$ is too close to the critical stabilization rate, given by $\sum_i \max\{0, \log(|\lambda_i(A)|)\}$, then disturbance rejection is not possible.

Corollary 3.4.4 *Let* $\mathbf{x}(k)$ *be the solution of the state-space equation (3.6) and* $g(k)$ *be a function satisfying*

$$\lim_{k\to\infty} \frac{g(k)}{k} = 0$$

If the plant is stabilized, i.e., $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ *then the following is satisfied:*

$$\frac{1}{2} \liminf_{k\to\infty} L_-(\mathbf{e}_{g(k)}^k) + C_{channel} \geq \sum_i \max\{0, \log(|\lambda_i(A)|)\} \tag{3.35}$$

Proof: Follows from theorem 3.4.3 and the fact that $I(\mathbf{v}_1^k \to \mathbf{z}_1^k) \leq C_{channel}$. □

3.4.1 An Integral Formula under Stationarity Assumptions

Under stationarity assumptions, the condition in theorem 3.4.3 can be expressed by means of an integral formula.

Corollary 3.4.5 *Let* $\mathbf{x}(k)$ *be the solution of the state-space equation (3.6). If the plant is stabilized, i.e.,* $\sup_k \mathcal{E}[\mathbf{x}^T(k)\mathbf{x}(k)] < \infty$ *and* $\mathbf{e}$ *is stationary, where* $0 < m < \hat{F}_{\mathbf{e}}(\omega) < M < \infty$ *is Lebesgue integrable, then the following is satisfied:*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}[\log S(\omega)]_{-}d\omega + \bar{I}_{\infty}(\mathbf{v} \to \mathbf{z}) \geq \sum_{i} \max\{0, \log(|\lambda_i(A)|)\} \tag{3.36}$$

where $S(\omega) = \sqrt{\hat{F}_{\mathbf{e}}(\omega)} = \sqrt{\frac{\hat{F}_{\tilde{\mathbf{e}}}(\omega)}{\hat{F}_{\mathbf{d}}(\omega)}} = \sqrt{\frac{\hat{F}_{\tilde{\mathbf{e}}}(\omega)}{|G(e^{j\omega})|^2}}$*. The processes* $\tilde{\mathbf{e}}$ *and* $\mathbf{d}$ *are the ones depicted in Fig 1-2.*

Proof: By means of a direct application of the theorem B.2.1 of Appendix B, we find that:

$$\lim_{k\to\infty} L_{-}(\mathbf{e}_1^k) = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\log(\hat{F}_{\mathbf{e}}(\omega))]_{-}d\omega \tag{3.37}$$

The result follows by direct substitution of (3.37) in (3.33). □

3.5 Example

Consider the linear feedback loop of Fig 3-2 and that the blocks and signals represented satisfy:

- **(E1)** $P(z)$ is a strictly proper linear and time-invariant plant of order n_p.
- **(E2)** $K(z)$ is a proper linear and time invariant system of order n_K
- **(E3)** $\mathbf{w}$ is a zero mean Gaussian i.i.d. and unit variance random process. The process $\mathbf{c}$ is i.i.d. and Gaussian, with variance σ_c^2.
- **(E4)** $\mathbf{x}(1)$ is also zero mean Gaussian.
- **(E5)** the initial state of K is taken as $x_K(1) = 0$

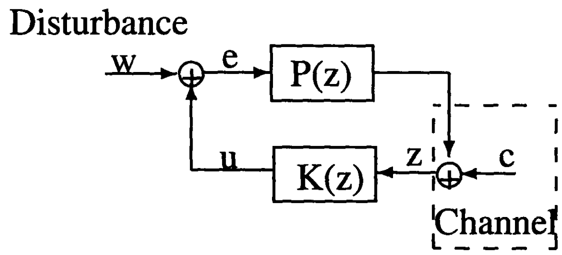

Figure 3-2: General structure for a linear feedback loop with disturbances and noisy measurements.

- **(E6)** the feedback loop is stable.

The example of Fig 3-2 is a particular instance of the scheme of Fig 3-1. All assumptions stated in section 3.2.1 are satisfied, in particular (A7) and (A8). Since we assume that $K(z)$ has zero initial state then (A7) is satisfied with $\alpha = 0$ and (A8) is also immediately true.

In the following sub-sections, we discuss a numerical computation which suggests that the following inequality (from the main Theorem 3.4.3) is not conservative:

$$\bar{I}_\infty(\mathbf{v} \to \mathbf{z}) - \sum_i \max\{0, \log(|\lambda_i(A)|)\} \geq -\frac{1}{2} \liminf_{k \to \infty} L_-(\mathbf{e}^k_{g(k)}) \tag{3.38}$$

3.5.1 Preliminary Results: Extension to the Non-Stationary Case

The aim of this subsection is to derive computable upper and lower bounds for the inequality (3.38). The subsequent Lemmas comprise integral formulas for the Gaussian asymptotic stationary case. The final inequality is presented in Corollary 3.5.4, where the upper and lower bounds are easily computable through integrals. Consequently, we can obtain the numerical results of section 3.5.2 and test for the tightness of the inequality (3.38).

We emphasize that, all the quantities in the statements and in the proofs of this subsection, refer to the example of Fig 3-2. As such, we assume that they comply with (E1)-(E6).

Lemma 3.5.1 *Let $\vec{e}$ be a stationary stochastic process with auto-covariance $R_{\vec{e}}(\tau)$. Consider that $\hat{F}_{\vec{e}}(\omega)$ is a real-valued, Lebesgue integrable function and that the following holds:*

$$\exists \beta > 0, \gamma \in (0,1) \textit{ such that } \forall k_0, \tau > 0, |R_e(k_0, k_0+\tau) - R_{\vec{e}}(\tau)| < \beta\gamma^{k_0} \tag{3.39}$$

We denote by m and M the essential lower bound and upper bound of $\hat{F}_{\vec{e}}(\omega)$, respectively, and assume that m and M are finite. The following is satisfied:

$$\liminf_{k\to\infty} L_-(\mathbf{e}_k^{k^2}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\log(\hat{F}_{\vec{e}}(\omega))]_- d\omega \tag{3.40}$$

Proof: We follow the same steps of the proof of lemma 3.4.2 and get the following decomposition of $\Sigma(\vec{\mathbf{e}}_k^{k^2})$:

$$\Sigma(\vec{\mathbf{e}}_k^{k^2}) = \begin{bmatrix} V_{k,+} \\ V_{k,-} \end{bmatrix}^T \begin{bmatrix} \Lambda_{k,+} & 0 \\ 0 & \Lambda_{k,-} \end{bmatrix} \begin{bmatrix} V_{k,+} \\ V_{k,-} \end{bmatrix} \tag{3.41}$$

where $[\Lambda_-]_{ii} < 1$ and $[\Lambda_+]_{ii} \geq 1$. Now, assumption (3.39) and Theorem B.2.1 of Appendix B, guarantee that:

$$\liminf_{k\to\infty} L_-(\mathbf{e}_k^{k^2}) = \liminf_{k\to\infty} \frac{1}{k^2-k+1}\log(det(V_{k,-}\Sigma(\mathbf{e}_k^{k^2})V_{k,-}^T)) \underset{(*)}{=}$$
$$\liminf_{k\to\infty} \frac{1}{k^2-k+1}\log(det(V_{k,-}\Sigma(\vec{\mathbf{e}}_k^{k^2})V_{k,-}^T)) = \lim_{k\to\infty} L_-(\vec{\mathbf{e}}_k^{k^2}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\log(\hat{F}_{\vec{e}}(\omega))]_- d\omega \tag{3.42}$$

The proof is complete, once we provide more detail on the validity of the equality marked with (*) in (3.42). From assumption (3.39) and the fact that $[\Lambda_{k,-}]_{ii} \geq m$, we know that:

$$\lim_{k\to\infty} \varrho(\Delta_{k,e}) = 0 \tag{3.43}$$

where $\varrho(\Delta_{k,e}) = \max_i \sum_j |[\Delta_{k,e}]_{i,j}|$ and $\Delta_{k,e}$ is a matrix satisfying:

$$V_{k,-}\Sigma(\mathbf{e}_k^{k^2})V_{k,-}^T = \underbrace{V_{k,-}\Sigma(\vec{\mathbf{e}}_k^{k^2})V_{k,-}^T}_{\Lambda_{k,-}} + V_{k,-}\left(\Sigma(\mathbf{e}_k^{k^2}) - \Sigma(\vec{\mathbf{e}}_k^{k^2})\right)V_{k,-}^T = \Lambda_{k,-}(I+\Delta_{k,e}) \tag{3.44}$$

We finalize, by using Gershgorin's circle Theorem to infer that $\lambda_i(I+\Delta_{k,e}) \in [1-\varrho(\Delta_{k,e}), 1+\varrho(\Delta_{k,e})]$. □

Lemma 3.5.2 *Let $\vec{\mathbf{z}}$ be a stationary stochastic process with auto-covariance $R_{\vec{z}}(\tau)$. Con-*

sider that $\hat{F}_{\vec{z}}(\omega)$ is a real-valued, Lebesgue integrable function and that the following holds:

$$\exists \beta > 0, \gamma \in (0,1) \text{ such that } \forall k_0, \tau > 0, |R_z(k_0, k_0+\tau) - R_{\vec{z}}(\tau)| < \beta\gamma^{k_0} \tag{3.45}$$

We denote by m and M the essential lower bound and upper bound of $\hat{F}_{\vec{z}}(\omega)$, respectively, and assume that m and M are finite. The following holds:

$$\frac{1}{4\pi}\int_{-\pi}^{\pi} \log(\frac{\hat{F}_{\vec{z}}(\omega)}{\sigma_c^2})d\omega \geq \bar{I}_\infty(\mathbf{v} \to \mathbf{z}) \tag{3.46}$$

Proof: Choose arbitrary $\nu \in \mathbb{N}_+$. We start by noticing that:

$$\limsup_{k\to\infty} h(\mathbf{z}(k)|\mathbf{z}_{k-\nu+1}^{k-1}) - h(\mathbf{c}(k)) \geq$$
$$\limsup_{k\to\infty} h(\mathbf{z}(k)|\mathbf{z}_1^{k-1}) - h(\mathbf{c}(k)) = \limsup_{k\to\infty} I(\mathbf{z}(k);\mathbf{v}_1^k|\mathbf{z}_1^{k-1}) \geq \bar{I}_\infty(\mathbf{v}\to\mathbf{z}) \tag{3.47}$$

Now, notice that assumption (3.45), guarantees that:

$$\lim_{k\to\infty} h(\mathbf{z}(k)|\mathbf{z}_{k-\nu+1}^{k-1}) = \lim_{k\to\infty} h(\vec{\mathbf{z}}(k)|\vec{\mathbf{z}}_{k-\nu+1}^{k-1}) \underset{stationarity}{=} h(\vec{\mathbf{z}}(\nu)|\vec{\mathbf{z}}_1^{\nu-1}) \tag{3.48}$$

which, from (3.47), implies:

$$h(\vec{\mathbf{z}}(\nu)|\vec{\mathbf{z}}_1^{\nu-1}) - h(\mathbf{c}(1)) \geq \bar{I}_\infty(\mathbf{v}\to\mathbf{z}) \tag{3.49}$$

Since ν was arbitrary, we can use (3.49) and Theorem B.2.1, from Appendix B, to state that:

$$\bar{I}_\infty(\mathbf{v}\to\mathbf{z}) \leq \lim_{\nu\to\infty} h(\vec{\mathbf{z}}(\nu)|\vec{\mathbf{z}}_1^{\nu-1}) - h(\mathbf{c}(1)) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\log(\frac{\hat{F}_{\vec{z}}(\omega)}{\sigma_c^2})d\omega \tag{3.50}$$

Theorem 3.5.3 *If the feedback system of Fig 3-2 is stable then the following holds:*

$$\frac{1}{4\pi}\int_{-\pi}^{\pi}\log(\frac{\hat{F}_{\vec{z}}(\omega)}{\sigma_c^2})d\omega - \sum_i \max\{0, \log(|\lambda_i(A)|)\} \geq$$
$$\bar{I}_\infty(\mathbf{v}\to\mathbf{z}) - \sum_i \max\{0, \log(|\lambda_i(A)|)\} \geq -\frac{1}{2}\liminf_{k\to\infty} L_-(\mathbf{e}_{g(k)}^k) =$$
$$-\frac{1}{2\pi}\int_{-\pi}^{\pi}[\log(\hat{F}_{\vec{e}}(\omega))]_- d\omega \quad (3.51)$$

Proof: Since the system is stable, the exponential asymptotic stationarity conditions of Lemmas 3.5.1 and 3.5.2 are satisfied. The result follows from these Lemmas and the main Theorem 3.4.3. □

The following Corollary, specializes Theorem 3.5.3 to the feed-back loop of Fig 3-2:

Corollary 3.5.4 *If the feedback system of Fig 3-2 is stable then the following holds:*

$$\frac{1}{4\pi}\int_{-\pi}^{\pi}\log\left(1+\frac{|P(e^{j\omega})|^2}{\sigma_c^2}\right)d\omega \geq$$
$$\bar{I}_\infty(\mathbf{v}\to\mathbf{z}) - \sum_i \max\{0, \log(|\lambda_i(A)|)\} \geq -\frac{1}{2}\liminf_{k\to\infty} L_-(\mathbf{e}_{g(k)}^k) =$$
$$-\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\log\left(\frac{1+\sigma_c^2|K(e^{j\omega})|^2}{|1+P(e^{j\omega})K(e^{j\omega})|^2}\right)\right]_- d\omega \quad (3.52)$$

Proof: We start by computing the power spectral density $\hat{F}_{\vec{z}}(\omega)$ to obtain:

$$\frac{\hat{F}_{\vec{z}}(\omega)}{\sigma_c^2} = \frac{1}{|1+P(e^{j\omega})K(e^{j\omega})|^2}\left(1+\frac{|P(e^{j\omega})|^2}{\sigma_c^2}\right) \quad (3.53)$$

But, from the residue theorem, we use (3.53) to show that:

$$\frac{1}{4\pi}\int_{-\pi}^{\pi}\log(\frac{\hat{F}_{\vec{z}}(\omega)}{\sigma_c^2})d\omega = \frac{1}{4\pi}\int_{-\pi}^{\pi}\log\left(1+\frac{|P(e^{j\omega})|^2}{\sigma_c^2}\right)d\omega + \sum_i \max\{0, \log(|\lambda_i(A)|)\} \quad (3.54)$$

The power spectral density of $\hat{F}_{\vec{e}}(\omega)$ leads to:

$$\frac{\hat{F}_{\vec{e}}(\omega)}{\sigma_c^2} = \frac{1+\sigma_c^2|K(e^{j\omega})|^2}{|1+P(e^{j\omega})K(e^{j\omega})|^2} \quad (3.55)$$

The proof is concluded by direct substitution of (3.54) and (3.55) in Theorem 3.5.3. □

3.5.2 Numerical results

From Corollary 3.5.4, we infer that an indication for the tightness of the inequality (3.38) is that the following lower-bound and upper-bound are close:

$$ub(\sigma_c^2) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \log\left(1 + \frac{|P(e^{j\omega})|^2}{\sigma_c^2}\right) d\omega \tag{3.56}$$

$$lb(\sigma_c^2) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[\log\left(\frac{1+\sigma_c^2|K(e^{j\omega})|^2}{|1+P(e^{j\omega})K(e^{j\omega})|^2}\right)\right]_- d\omega \tag{3.57}$$

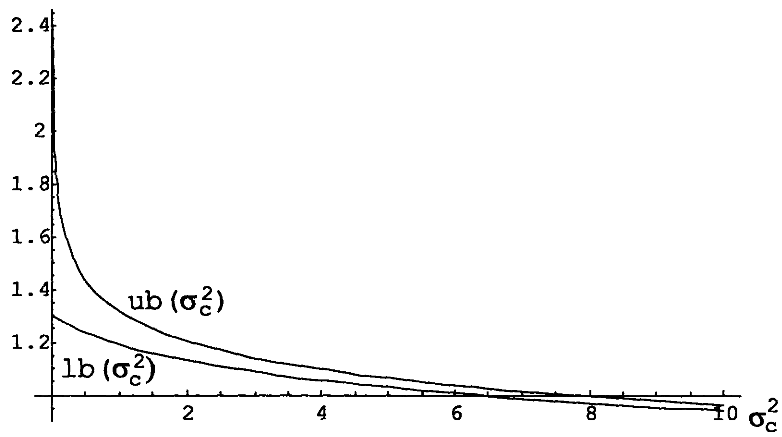

Figure 3-3: Plot of the upper-bound and lower-bound, computed as a function of σ_c^2.

In Figure 3-3, we depict the numerical results for the following P and K:

$$P(z) = \frac{z^{-1}}{(1-1.5z^{-1})^{10}} \tag{3.58}$$

$$K(z) = z - \frac{1}{P(z)} \tag{3.59}$$

Notice that K is the dead-bit controller. We emphasize that the choice of the multiple pole of $P(z)$ was arbitrary. We have tried other values and the bounds behaved in a similar way.

By inspection, one can argue that the bounds get more accurate for increasing values

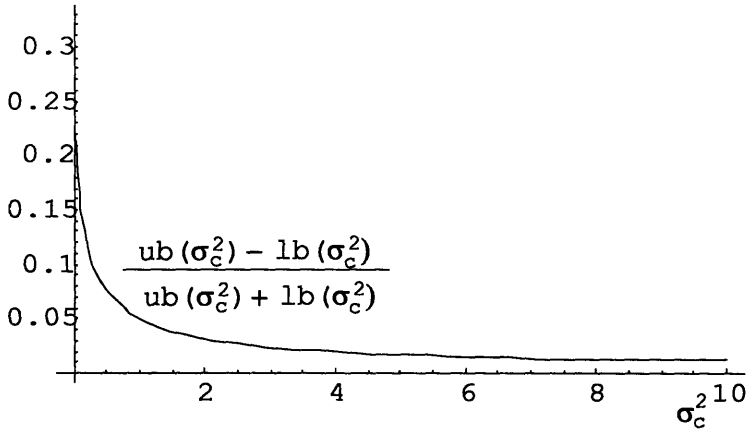

Figure 3-4: Plot of the relative difference between the upper-bound and lower-bound, computed as a function of σ_c^2.

of σ_c^2 (see Fig 3-4). Moreover, we have verified empirically that such relative accuracy can be made arbitrarily small by considering $P(z) = \frac{z-1}{(1-1.5z^{-1})^n}$, with n arbitrarily large and $K(z) = z - \frac{1}{P(z)}$.

Chapter 4

An Information Theoretic Approach to the Modal Estimation of Switching FIR Linear Systems

4.1 Introduction

The implications of modal estimation span applications in Adaptive Control and fault detection. Depending on the formulation, modal estimation may be realized as part of a hybrid state estimation paradigm[57], comprising discrete (modes) and continuous states. Although the investigation of such problems has generated a vast portfolio of algorithms and methods, [58] suggests the need to devote more attention to the modal estimation alone. By adopting the framework used in the identification of FIR-LPV systems [17], the estimation of the continuous state can be avoided. Consequently, we focus on the modal estimation problem by considering systems that switch among a set of modes which comprise finite impulse response, or moving average, filters. A discrete stochastic process q_k drives the switching, while the system is excited by a white Gaussian process (probing signal). The optimal modal MAP-estimation problem can be cast as a Bayesian Hypothesis testing. The search space grows with m^k, where m is the number of modes and k is the number of observations.

We recognize that the modal estimation problem is equivalent to a communication setup in the presence of randomly generated codes [5]. Under that framework, the input probing signal defines a constrained code and the system is perceived as an encoder of the mode sequence $\mathbf{q}_k$. Motivated by that, we adopt a decoder structure for the mode estimator. The search space is reduced to $m^{k(r^q+\epsilon)}$ elements, where $r^q \in [0, 1]$ is the entropy rate of $\mathbf{q}_k$ and $\epsilon > 0$ is a quality parameter that determines the *degree* of optimality of the estimates. Such reduction is achieved by constraining the search to the typical set [20] of mode sequences. Since such set is simultaneously a high probability set no information is lost. This reduction is a unified and asymptotically optimal way to implement the merging and pruning of hypothesis that otherwise would rely on approximations, mode transition detectors, slow variation hypothesis [30] or the use of forgetting factors. We also present an alternative low complexity estimation algorithm that converges in probability, as k tends to infinity, to the decoder proposed. It is based on a binary search that requires, at every step, an optimization using a forward dynamic program of complexity km^2.

Although using a different formulation, we refer to [12] where it is exposed the difficulty of computing measures of quality for the available modal estimation algorithms. In practice, such quality evaluation may have to resort to Monte Carlo simulations. We address that problem by using Shannon's theory to define a measure of distortion $\mathcal{D}_d$ that is suitable for a probabilistic worst-case analysis. Using this distortion, we determine the probability that a sequence of mode estimates $(\hat{q}_1, \ldots, \hat{q}_k)$ (generated by the decoder) is in a ball, of radius β, around the true sequence $(q_1, \ldots, q_k)$. An interesting feature of this framework is that there exists β such that the probability of $\mathcal{D}(q_{1,k}, \hat{q}_{1,k}) < \beta$ converges to 1 as k tends to infinity. We compute such β and show that it is an affine function of r^q, the entropy rate of the switching process. In that sense, as a theoretical result, our work relates to [70], where it is proven that the uncertainty [63] in identifying time-varying systems is directly related to the *speed* of variation. Information theoretically inspired distances, have been widely used in system identification and parameter estimation [64, 56, 3]. The distortion $\mathcal{D}_d$ can also be used as a quality measure on the design of probing signals and can be viewed as a first step to the proper study of the effect of observed inputs on the mode observability of linear hybrid systems. We also stress that our analysis and methods are applicable to other classes of stable switching systems [39]. In these cases, the moving average coefficients should be the truncation of the impulse response of such stable linear systems. Worst-case LPV-FIR system identification was also broached by [6] in the case where the coefficients have a specific functional form.

This chapter is organized as follows: Section 4.1.1 introduces the notation used throughout the text. The problem is stated and its information theoretic equivalence established in section 4.2, while section 4.2.1 provides a guide through the main results. The estimation paradigm is described in section 4.3, while the performance analysis is carried out in section 4.4. An efficient estimation algorithm is presented in section 4.4.1 and numerical examples are provided in section 4.4.2.

4.1.1 Notation

The following notation is adopted: Large caps letters are used to indicate vectors and matrices. Small caps letters are reserved for real scalars and discrete variables. In addition, p is reserved to represent probability distributions. Discrete-time sequences are indexed by time using integer subscripts, such as x_k. Finite segments of discrete-time sequences are indicated by the range of their time-indexing as in the following example ($k \geq n$):

$$q_{n,k} = (q_n, \ldots, q_k) \tag{4.1}$$

Superscripts are reserved for distinguishing different variables and functions according to their meaning. Random variables are represented using boldface letters and follow the conventions above. As an illustration, $\mathbf{x}$ is a valid representation for a scalar random variable, while a sample (or realization) is written as x. Also, a finite segment of a discrete-time sequence of random variables, would be $\mathbf{q}_{1,k}$. A realization of such process would be indicated as $q_{1,k}$. The probability of an *event* is indicated by $\mathcal{P}(event)$. We use the entropy function, of a random variable $\mathbf{Z}$, given by:

$$\mathcal{H}[\mathbf{Z}] = \mathcal{E}[-\ln p^{\mathbf{Z}}(\mathbf{Z})] \tag{4.2}$$

where $p^{\mathbf{Z}}$ is the p.d.f. of $\mathbf{Z}$ and $\mathcal{E}[\cdot]$ is the expected value, taken over $\mathbf{Z}$. Similarly, the conditional entropy is given by $\mathcal{H}[\mathbf{Z}^1|\mathbf{Z}^2] = \mathcal{H}[\mathbf{Z}^1, \mathbf{Z}^2] - \mathcal{H}[\mathbf{Z}^2]$ where, in this case, the expectation is taken with respect to $\mathbf{Z}^1$ and $\mathbf{Z}^2$. The covariance matrix of a random matrix $\mathbf{Z}$, with $Z \in \mathbb{R}^{n^1 \times n^2}$, is given by:

$$\Sigma_Z = \mathcal{E}\left[\left(\vec{\mathbf{Z}} - \mathcal{E}[\vec{\mathbf{Z}}]\right)\left(\vec{\mathbf{Z}} - \mathcal{E}[\vec{\mathbf{Z}}]\right)^T\right], \text{ where } \vec{\mathbf{Z}} = vec(\mathbf{Z}) \tag{4.3}$$

4.2 Problem statement

Consider the mode alphabet $\mathbb{A} = \{1, \ldots, m\}$, with $m \geq 2$, and the random process $\mathbf{F}_k$, with $F_k \in \mathbb{R}^{n^F}$, described by:

$$\mathbf{F}_k = \mathbf{Y}_k + \mathbf{W}_k, \; k \geq 1 \tag{4.4}$$

$$\mathbf{Y}_k = \sum_{i=0}^{\alpha} G_i(\mathbf{q}_k)\mathbf{V}_{k-i}, \; k \geq 1 \tag{4.5}$$

where $k, \alpha \in \mathbb{N}$ and $G_i : \mathbb{A} \rightarrow \mathbb{R}^{n^F \times n^V}$ are matrices that specify the switching system. The stochastic processes $\mathbf{V}_k$ (probing signal), $\mathbf{W}_k$ and $\mathbf{q}_k$ are mutually independent and satisfy:

- $\mathbf{V}_k$ and $\mathbf{W}_k$ are Gaussian zero mean i.i.d. processes, with $V_k \in \mathbb{R}^{n^V}$ and $W_k \in \mathbb{R}^{n^F}$. In order to avoid degeneracy problems, we assume $\Sigma_W > 0$. In contrast to $\mathbf{W}_k$, the probing signal $\mathbf{V}_k$ is assumed to be observed by the estimator. Examples where this is a realistic assumption are: when $\mathbf{V}_k$ is generated by the estimator; when it is an exogenous process that can be observed by the estimator or a combination of both.

- $\mathbf{q}_k$ is a discrete, stationary and ergodic Markovian stochastic process with alphabet $\mathbb{A}$. For a given $k \in \mathbb{N}$, we write the p.d.f. of $\mathbf{q}_{1,k}$ as $p^{\mathbf{q}}(q_{1,k})$, regardless of k and $p^{\mathbf{q}}(\mathbf{q}_k|\mathbf{q}_{k-1}) = \frac{p^{\mathbf{q}}(q_k, q_{k-1})}{p^{\mathbf{q}}(q_{k-1})}$. The entropy rate [20] of $\mathbf{q}_k$, designated by r^q, is computed as:

$$r^q = \mathcal{E}[-log_m p^{\mathbf{q}}(\mathbf{q}_k|\mathbf{q}_{k-1})] \tag{4.6}$$

For a given k, we wish to use the probing signal $\mathbf{V}_{1-\alpha,k}$ and a decision system that, by means of the measurement of $\mathbf{F}_{1,k}$, produces an estimate $\hat{\mathbf{q}}_{1,k}$ of $\mathbf{q}_{1,k}$. The estimation method must have an associated measure of distortion $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$ that allows the specification of a ball around $q_{1,k}$ where the estimates $\hat{q}_{1,k}$ will lie with a given probability. In particular, for a given $\beta > 0$, we are interested in computing $\mathcal{P}(\mathcal{D}_d(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \beta)$. Such probability is expected to depend on β and k. By using an information theoretic formulation, this characterization must also reflect the informativity of $\mathbf{V}_k$ and r^q, the entropy rate of the switching process. We would like to stress that, without loss of generality, we develop our analysis using the origin of time as $k = 1$. In real applications, this setup can be used in a **sliding window** scheme and k *should be viewed as the time horizon of the estimator.*

4.2.1 Main results

Among the results presented in this Chapter, the following are central to answering the questions posed above: In definition 4.4.2 a measure of distortion $\mathcal{D}_d$ is introduced. It is through this function that the informativity of $\mathbf{V}_k$ can be gauged. If $\mathbf{V}_k$, or only part of it, are generated by the estimator then that generated portion of Σ_V may be tuned to achieve a desired distortion function. Similar methods for the distance-based design of probing functions were derived in [3]. In lemma D.1.6 we show that, as k increases, $\mathcal{P}(\mathcal{D}_d(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \beta)$ decreases. Theorem 4.4.1 proves that if $\beta > r^q \ln(m) + \frac{n^F}{2}$ then $\mathcal{P}(\mathcal{D}_d(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \beta)$ can be made arbitrarily small by increasing k.

4.2.2 Posing the Problem Statement as a Coding Paradigm

As the generality of the estimation paradigms [64], the estimation of the mode of (4.4)-(4.5) can be interpreted as a problem of communication through a noisy channel. The message to be transmitted is $q_{1,k}$, the probing signal $\mathbf{V}_{1-\alpha,k}$ specifies the code while the measurement noise $\mathbf{W}_{1,k}$ completes the setup of such channel. The decoder is bound to use $\mathbf{V}_{1-\alpha,k}$ and the noisy measurements $\mathbf{F}_{1,k}$ to make a decision as to which is the best estimate $\hat{q}_{1,k}$.

4.3 Encoding and decoding

A specific feature of this communications setup is that the *encoder* does not know the message to be transmitted. This is a problem if one wants to adopt the approach of coding long words as a way to reduce the probability of error (channel coding theorem). In order to circumvent this difficulty, we follow the procedure in the proof of Shannon's channel coding theorem, i.e., the use of random coding [5]. Consequently, we consider the white Gaussian process $\mathbf{V}_{1-\alpha,k}$ as establishing a constrained random code specified by Σ_V. In the decoding process we will use the following estimate:

$$\hat{\mathbf{Y}}_k(\hat{q}_k) = \sum_{i=0}^{\alpha} G_i(\hat{q}_k)\mathbf{V}_{k-i} \tag{4.7}$$

The decoding process has a hypothesis testing structure. The likelihood of a given candidate sequence $\hat{q}_{1,k}$ is gauged by means of the estimation error $\mathbf{F}_k - \hat{\mathbf{Y}}_k(\hat{q}_k)$.

4.3.1 Description of the estimator (decoder)

By following the approach that leads to a standard decoder [20], in this section we construct a mode estimator. In the subsequent analysis we use the following result:

Remark 4.3.1 *If* $\mathbf{X}$ *is a Gaussian, zero mean random variable with covariance matrix* $\Sigma_X \in \mathbb{R}^{n^X \times n^X}$*, then:*

$$\mathcal{H}(\mathbf{X}) = \frac{1}{2}\ln\left((2\pi e)^{n^X}|\Sigma_X|\right) \tag{4.8}$$

The following is the definition of the noise entropy rate r^W. Such quantity is important in the construction of the estimator.

Definition 4.3.1 *Define the noise entropy rate* r^W *as:*

$$r^W = \mathcal{H}(\mathbf{W}_k) = \frac{\ln\left((2\pi e)^{n^F}|\Sigma_W|\right)}{2} \tag{4.9}$$

The following definitions complete the list of mathematical objects needed to describe the decoder.

Definition 4.3.2 *Consider $\hat{q}_{1,k} \in \mathbb{A}^k$ and realizations $F_{1,k}$ and $\hat{Y}_{1,k}(\hat{q}_{1,k})$. Given parameters $k \in \mathbb{N}$ and $\epsilon \in (0,1)$, define the selection indicator $s^{\epsilon,k} : \mathbb{A}^k \rightarrow \{True, False\}$ as:*

$$s^{\epsilon,k}(\hat{q}_{1,k}) = \begin{cases} True & if \, \frac{-\ln p^{\mathbf{W}_{1,k}}(F_{1,k}-\hat{Y}_{1,k}(\hat{q}_{1,k}))}{kn^F} < \frac{r^W}{n^F} + \epsilon \\ False & otherwise \end{cases} \tag{4.10}$$

The decoding process is a search in the typical set defined bellow.

Definition 4.3.3 *Given the parameters $k \in \mathbb{N}$ and $\epsilon \in (0,1)$, the set of typical sequences $\mathbb{T}^{\epsilon,k}$ is defined as:*

$$\mathbb{T}^{\epsilon,k} = \{\hat{q}_{1,k} \in \mathbb{A}^k : \frac{-\log_m p^{\mathbf{q}}(\hat{q}_{1,k})}{k} < r^q + \epsilon\} \tag{4.11}$$

It is the cardinality of $\mathbb{T}^{\epsilon,k}$ that determines the computational complexity of the estimator. According to [20] such quantity is bounded by $m^{k(r^q+\epsilon)}$. The structure of the decoder is the following:

Definition 4.3.4 ***(Decoder)*** *Given the realizations $F_{1,k}$ and $V_{1,k}$, parameters $k \in \mathbb{N}$ and $\epsilon \in (0,1)$, define the decoder as a search in $\mathbb{T}^{\epsilon,k}$ that generates $\hat{q}_{1,k}$ satisfying:*

$$\hat{q}_{1,k} \in \mathbb{T}^{\epsilon,k} \text{ and } s^{\epsilon,k}(\hat{q}_{1,k}) = True \tag{4.12}$$

If, for a given realization, there is no such $\hat{q}_{1,k}$, then the decoder generates an arbitrary $\hat{q}_{1,k} \in \mathbb{T}^{\epsilon,k}$. The decoding process defines a random variable, which we designate by $\hat{\mathbf{q}}_{1,k}$.

4.4 Performance Analysis

The following definition, describing a *conditional* test random variable, will facilitate the performance analysis of the decoder.

Definition 4.4.1 *Given $k \in \mathbb{N}$, $k \geq 1$ and $q_k, \hat{q}_k \in \mathbb{A}$, define the random variable $\mathbf{T}_k(q_k, \hat{q}_k)$ as:*

$$\mathbf{T}_k(q_k, \hat{q}_k) = \sum_{i=0}^{\alpha} (G_i(q_k) - G_i(\hat{q}_k)) \, \mathbf{V}_{k-i} + \mathbf{W}_k \tag{4.13}$$

Note that, given $\hat{q}_k \in \mathbb{A}$, the random variable $\mathbf{F}_i - \hat{\mathbf{Y}}(\hat{q}_k)$ conditioned to a fixed q_k is given by $\mathbf{T}_k(q_k, \hat{q}_k)$. If $\hat{q}_k = q_k$, then $\mathbf{T}_k(q_k, \hat{q}_k) = \mathbf{W}_k$. For any given indices k^1, k^2 and $q_{k^1,k^2}, \hat{q}_{k^1,k^2} \in \mathbb{A}^{k^2-k^1+1}$, we adopt the following abuse of notation:

$$\mathbf{T}_{k^1,k^2}(q_{k^1,k^2}, \hat{q}_{k^1,k^2}) = (\mathbf{T}_{k^1}(q_{k^1}, \hat{q}_{k^1}), \ldots, \mathbf{T}_{k^2}(q_{k^2}, \hat{q}_{k^2})) \tag{4.14}$$

The quality evaluation, of the coding/decoding process, is carried out by computing the probability that $\hat{\mathbf{q}}_{1,k}$ is in a ball around $\mathbf{q}_{1,k}$. Such *uncertainty* set is specified by means of the distortion $\mathcal{D}_d$ defined bellow. This function is related to the concept of divergence as in [3].

Definition 4.4.2 *(Measure of Distortion)The distortion $\mathcal{D}_d : \mathbb{A}^k \times \mathbb{A}^k \to \mathbb{R}$ is given by:*

$$\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k}) = \frac{\mathcal{H}(\mathbf{T}_{1,k}(q_{1,k}, \hat{q}_{1,k}))}{k} - r^W \tag{4.15}$$

The following theorem is the main result of this Chapter.

Theorem 4.4.1 *Let $\hat{\mathbf{q}}_{1,k}$ be determined according to the decoding process described in the definition 4.3.4. For any given $\delta > 0$, there exists $k \in \mathbb{N}$ and $\epsilon \in (0,1)$ such that:*

$$\mathcal{P}\left(\mathcal{D}_d(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \ln(m)r^q + \frac{n^F}{2} + \delta\right) < \delta, \text{ if } r^q \geq 0 \tag{4.16}$$

$$\mathcal{P}\left(\mathcal{D}_d(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \delta\right) < \delta, \text{ if } r^q = 0 \tag{4.17}$$

Proof:

The result for $r^q \geq 0$ follows directly from lemma D.1.6 in section D.1. □

4.4.1 An Efficient Decoding Algorithm

In this subsection, the main result is lemma 4.4.2. It shows that, for k sufficiently large, the exhaustive search of definition 4.3.4 can be replaced by a binary search over a non-negative parameter γ. The stopping condition relies on the minimal solution of the following cost functional:

Definition 4.4.3 *Given realizations $q_{1,k}$, $F_{1,k}$, $V_{1,k}$ and $W_{1,k}$, define the following cost function:*

$$\mathcal{J}_\gamma(\bar{q}_{1,k}) = \frac{-\ln\left[p^{\mathbf{W}_{1,k}}\left(F_{1,k} - \hat{Y}_{1,k}(\bar{q}_{1,k})\right)\right]}{k} + \gamma\frac{-\log_m\left(p^{\mathbf{q}}(\bar{q}_{1,k})\right)}{k} \tag{4.18}$$

where $\bar{q}_{1,k} \in \mathbb{A}^k$ is a candidate sequence and γ is a non-negative constant. The first term of the cost function is a sample divergence, while the second is designated as sample entropy rate. Note that the two terms in the cost (4.18) are identical to the ones in (4.10) and (4.11).

Lemma 4.4.2 *Let $\bar{q}^*_{1,k}$ be the optimal solution given by:*

$$\bar{q}^*_{1,k} = \underset{\bar{q}_{1,k}\in\mathbb{A}^k}{argmin}\mathcal{J}_\gamma(\bar{q}_{1,k}) \tag{4.19}$$

and $\hat{\mathbf{q}}_{1,k}$ be the estimate of definition 4.3.4. If $\mathbf{q}_k$ is i.i.d., then the following holds:

$$\lim_{k\to\infty} \mathcal{P}\left(\mathbf{E}_1 \wedge \mathbf{E}_2 \implies \exists a, b, \forall\gamma \in (a,b)\ \mathbf{E}_3 \wedge \mathbf{E}_4\right) = 1 \tag{4.20}$$

where $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$ and $\mathbf{E}_4$ are the following events:

$$\mathbf{E}_1 = \mathbf{s}^{\epsilon,k}(\hat{\mathbf{q}}_{1,k}),\ \mathbf{E}_2 = \left(\hat{\mathbf{q}}_{1,k} \in \mathbb{T}^{\epsilon,k}\right) \tag{4.21}$$

$$\mathbf{E}_3 = \mathbf{s}^{2\epsilon,k}(\bar{\mathbf{q}}^*_{1,k}),\ \mathbf{E}_4 = \left(\bar{\mathbf{q}}^*_{1,k} \in \mathbb{T}^{2\epsilon,k}\right) \tag{4.22}$$

Remark 4.4.1 ***(Binary Search)*** *The two terms in the cost (4.18) analyzed separately, in terms of the events (4.21)-(4.22), lead to the following structure for the binary search:*

- *A simple optimality argument is sufficient to show that for any $\gamma > 0$, the optimal solution satisfies $E_1 \wedge E_2 \implies E_3 \vee E_4$.*
- *If $\gamma = 0$, then the optimal solution is the maximum likelihood estimate over the unrestricted search space $\mathbb{A}^k$. It implies that, for γ sufficiently small, $E_1 \implies E_3$.*
- *As γ increases the first term (divergence) is non-decreasing and the second term (entropy rate) is non increasing at the optimal solution. In the limit, the sample*

entropy rate converges to a constant $c \leq r^q$. It shows that for γ sufficiently large E_4 is true.

- *Lemma 4.4.2 shows that, with probability arbitrarily close to 1, there exists an interval (a,b) such that for all $\gamma \in (a,b)$ the optimal solution satisfies $E_1 \wedge E_2 \implies E_3 \wedge E_4$. The aim of the binary search is to find some γ^* in this interval. Given γ at time k, if $E_3 \wedge \neg E_4$ then, at time $k+1$, γ must be increased. On the other hand, if $\neg E_3 \wedge E_4$ then γ must be decreased. If $E_1 \wedge E_2$ is satisfied, then, after a finite number of steps, γ will be in (a,b) and $E_3 \wedge E_4$ is satisfied. The optimal sequence is taken as the resulting mode sequence estimate.*

- *A minimum size for the interval (a,b) must be specified. If the binary search reaches such limit it must end, declare that the condition $E_3 \wedge E_4$ could not be satisfied and $\bar{q}^*_{1,k}$ be chosen arbitrarily. Since the smallest the minimal size of the interval the better, a safe approach is to simulate the estimator and decrease that quantity until the estimates are within the performance predicted in theory. The minimal size will impact the number of iterations and, as such, the final complexity of the algorithm. We verified empirically that, even in cases where $k > 1000$, $\alpha = 10$ and $n^F = 2$ the search is concluded within a few seconds.*

The proof of lemma 4.4.2 follows by continuity arguments. We also believe that lemma 4.4.2 extends to general Markovian $\mathbf{q}_k$. So far, we keep it as a conjecture that is empirically verifiable through simulations.

Computational complexity of minimizing $\mathcal{J}_\gamma(\bar{q}_{1,k})$

In the following analysis we show that the minimization of $\mathcal{J}_\gamma(\bar{q}_{1,k})$ has a computational complexity that grows linearly with k. We start by recalling that since $\mathbf{W}_k$ is i.i.d. and $\mathbf{q}_k$

is Markovian, we can rewrite $\mathcal{J}_\gamma(\bar{q}_{1,k})$ as:

$$\mathcal{J}_\gamma(\bar{q}_{1,k}) = \sum_{i=2}^{k} -\frac{\ln p^{\mathbf{W}_i}(F_i - \hat{Y}_i(\bar{q}_i))}{k} - \gamma\frac{\log_m p^{\mathbf{q}}(\bar{q}_i|\bar{q}_{i-1})}{k} + \\ -\frac{\ln p^{\mathbf{W}_1}(F_1 - \hat{Y}_1(\bar{q}_1))}{k} - \gamma\frac{\log_m p^{\mathbf{q}}(\bar{q}_1)}{k} \quad (4.23)$$

Now, notice that (4.23) is suitable for forward dynamic programming [8]. The optimal $\bar{q}^*_{1,k}$ is the the minimal path solution of (4.23). At every step $k^* \in \{1, \ldots, k\}$, the algorithm recursively determines the optimal path $\bar{q}^*_{1,k^*}$ for all end points. The total number of operations is $m^2 k$ and, as such, computational complexity grows linearly with k.

4.4.2 Numerical Results

In order to illustrate the previously described optimization method, we refer to figures 4-1 and 4-2 where we depict the results for the following parameters:

- $\mathbb{A} = \{1, 2\}, m = 2$
- $G_0(1) = 1, G_1(1) = 0.9, G_2(1) = 0.81$
- $G_0(2) = 1, G_1(2) = 0.3, G_2(2) = 0.09$
- $k = 200$

After inspection, we notice that the distortion measure is much smaller than the upper-bound in 4.16. In the first case, we get $\simeq 0.07$ when the upper-bound dictates $0.469 + 0.5 = 0.969$. In addition, we notice that lower r^q leads to lower distortion measures. After extensive simulations, we found that such conservatism was consistent. Moreover, it should be expected that, as r^q decreases, the distortion measure gets smaller with high probability. Such behavior cannot be captured by theorem 4.4.1 as the contribution of r^q gets masked by the $1/2$ factor. After analyzing the proof of theorem 4.4.1, we concluded that the cause was that the upper-bound expressed in lemma D.1.4 was too conservative. A more general version of theorem 4.4.1 will be reported in future publications. The aforementioned extension replaces the $1/2$ factor by a variable β which determines the probability $\mathcal{P}\left(\mathcal{D}_d(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \ln(m)r^q + \beta\right)$. If $\beta = 1/2$ then the result in theorem 4.4.1 is recovered. Lower values of β lead to a non-zero limit probability, as k tends to infinity.

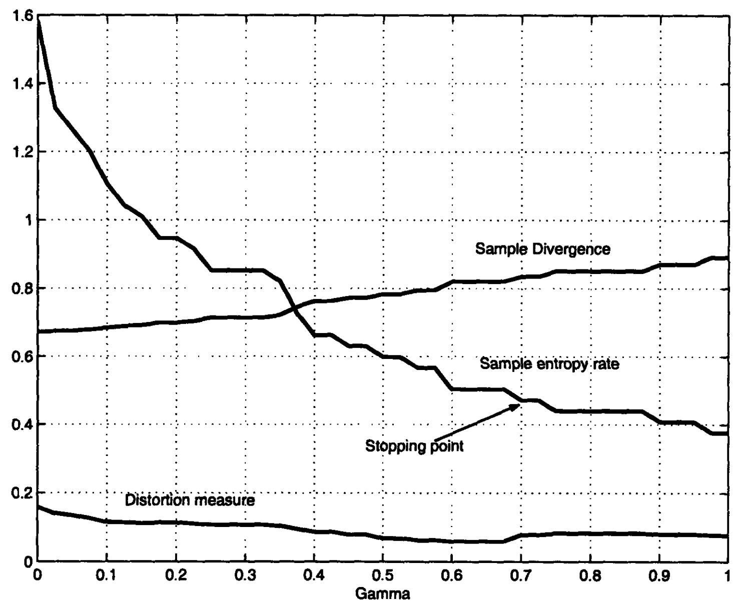

Figure 4-1: Simulation for a switching process described by $p^q(1|1) = p^q(2|2) = 0.9$ and $p^q(1|2) = p^q(2|1) = 0.1$ and entropy rate $r^q = 0.469$

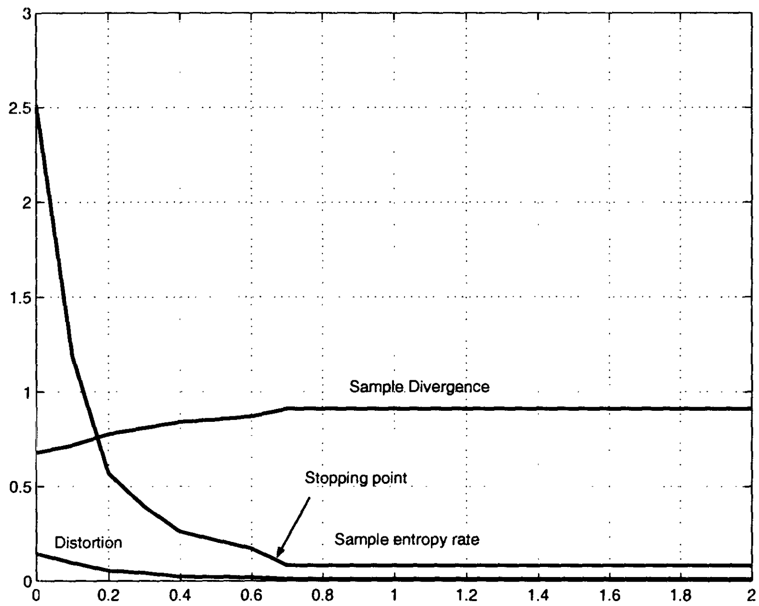

Figure 4-2: Simulation for a switching process described by $p^q(1|1) = p^q(2|2) = 0.99$ and $p^q(1|2) = p^q(2|1) = 0.01$ and entropy rate $r^q = 0.08$

Chapter 5

Final Conclusions and Future Directions

In this chapter we present the conclusions of Chapters 2, 3 and 4. We provide a separate list of conclusions along with future directions.

5.1 Conclusions About Chapter 2

The main results of Chapter 2 are the sufficiency theorems 2.3.2 and 2.3.4 proven in section 2.3 as well as the extension to the multi-state case given in section 2.6. The sufficiency conditions are proven constructively by means of the stabilizing feedback scheme of definition 2.3.3.

Our results lead to the following conclusions:

- The necessary and sufficient conditions can be expressed as inequalities involving $\mathcal{C}$ and $\mathcal{R}$ plus a few more terms that depend on the statistical behavior of the plant and the link as well as the descriptions of uncertainty. The intuition behind, the auxiliary quantities, that enable such representation is given in section 2.5.

- In order to preserve stability, the presence of randomness must be offset by an increase of the average transmission rate $\mathcal{C}$. From the necessary and sufficient conditions for stabilizability, we infer that the limitations created by randomness in the plant and the link are mathematically equivalent. In addition, we find that the higher $\mathcal{C}$ the larger the tolerance to uncertainty in the plant.

- Our results extend to a class of multi-state systems.

As a future direction, we suggest the study of stabilization of uncertain systems using output feedback, in the presence of information constraints. It is also important to investigate the stabilization of more general stochastic systems.

5.2 Conclusions About Chapter 3

From our results in Chapter 3, we conclude the following:

- Bode's integral inequality can be derived by means of Information Theoretic arguments. In this case, the fundamental limit appears as a result of causality alone.
- In Lemma 3.4.1 we show that the channel must carry information about the disturbance in addition to the information about the state of the plant.
- The ability to reject disturbances and the directed mutual information rate at the channel are strongly related. Good disturbance rejection requires a large directed mutual information rate at the channel. Theorem 3.4.3 presents the main inequality expressing the fundamental limitation.
- The numerical examples show that the inequality of Theorem 3.4.3 is not conservative.

It remains to solve the design problem, i.e., to have systematic methods of designing the encoders and the decoders that make the feedback system meet certain specifications, such as disturbance rejection or trajectory following.

5.3 Conclusions About Chapter 4

The main theorem of Chapter 4 shows that there exists an estimator for which the estimates are contained in a ball - defined by the metric $\mathcal{D}_d$ - centered around the true sequence, with probability arbitrarily close to 1. The *radius* of the ball can be made as close to $r^q + \frac{n^F}{2}$ as desired.

One of our main conclusions is that, regarding the process $\mathbf{q}_k$, the entropy rate r^q is a fundamental quantity in the study of estimation fidelity. The following is a list of reasons for that:

- If the transition probabilities are known then r^q can be computed.
- If r^q is not known then one can select any probabilistic structure for $\mathbf{q}_k$. By assuming that $\mathbf{q}_k$ is i.i.d., we get the scenario where, according to classical estimation theory (Van Trees), the probability of error can be arbitrarily close to 1.
- This is in accord with what should be expected from rate-distortion theory.

The following complete our conclusions about Chapter 4:

- The metric $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$ has several desirable properties which are explored in [43]. In particular, $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$ establishes a topology in the space of sequences. The covariance matrix of the probing signal is one of the parameters that define $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$ and, as such, it will shape the topology. In particular, scaling up the covariance matrix of $\mathbf{V}_k$ leads to a finer topology.
- The worst case analysis quantified in Theorem 4.4.1 enables the use of $\mathcal{D}_d(q_{1,k}, \hat{q}_{1,k})$ to answer the following question: **Is it possible to reliably communicate through the sequence $\mathbf{q}_k$?** From Theorem 4.4.1, we find that the answer is yes, provided that we can partition the typical sets of $\mathbf{q}_{1,k}$ with balls of size $2\ln(m)r^q + n^F$, as measured by $\mathcal{D}_d$.

In future publications, we will report an extension to the main Theorem of Chapter 4. In particular, such extension has the right limit behavior for $r^q \longrightarrow 0$. We show that, for a fixed signal to noise ratio, there exists an estimation scheme such that $\lim_{r^q \rightarrow 0} \mathcal{P}(\mathbf{q}_k \neq \hat{\mathbf{q}}_k) = 0$.

Appendix A

Aspects Related to Assumptions (A5),(A7)-(A9) of Chapter 3

A.1 Further Remarks About Assumption (A5)

The following upper-bound holds:

$$\bar{I}_\infty(\mathbf{v} \to \mathbf{z}) \leq C_{channel}$$

By means of property (P3), we know that the following is satisfied:

$$\forall k \geq i, I(\mathbf{z}(i); \mathbf{v}_1^i | \mathbf{z}_1^{i-1}) \leq I(\mathbf{z}(i); \mathbf{v}_1^k | \mathbf{z}_1^{i-1}) \implies \bar{I}_\infty(\mathbf{v} \to \mathbf{z}) \leq \bar{I}_\infty(\mathbf{v}; \mathbf{z}) \leq C_{channel} \tag{A.1}$$

where the last inequality is standard and can be found in [20].

A.2 Further Remarks About Assumption (A7)

Notice that a synchronous block decoder, with delay α, falls into this category. In addition, any dynamic system, of the form $\mathbf{u}(k) = f(\mathbf{u}_{k-\alpha}^{k-1}, \mathbf{z}_{k-\alpha}^{k})$, will satisfy (3.7). We emphasize that this representation does not pressupose a full-information system. For example, if $\mathbf{y}_c(k)$ is the output of an observable n-th order linear and time-invariant system, with input $\mathbf{z}(k)$, then it is possible to represent its input-output behavior in the form $\mathbf{y_c}(k) = f(\mathbf{y_c}_{k-n}^{k-1}, \mathbf{z}_{k-n}^{k})$.

A.3 Further Remarks About Assumption (A8)

The following is the assumption (A8), repeated here for convenience:

$$\lim_{k\to\infty} \frac{1}{k} I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k | \mathbf{z}_1^k)) = 0 \tag{A.2}$$

A.3.1 Assumption (A8) when $\mathcal{U}$ is countable

If $\mathcal{U}$ is countable then we can use (P1)-(P2) to conclude that:

$$I(\mathbf{u}_1^\alpha; \mathbf{w}_1^k | \mathbf{z}_1^k) \leq H(\mathbf{u}_1^\alpha) \tag{A.3}$$

As such, if $H(\mathbf{u}_1^\alpha) < \infty$ holds then $I(\mathbf{u}_1^\alpha; \mathbf{w}_1^k | \mathbf{z}_1^k) < \infty$ is satisfied. If $\mathcal{U}$ has $\aleph_U$ elements, such quantity is upper-bounded [20] as $H(\mathbf{u}_1^\alpha) \leq \alpha \log(\aleph_U)$. The confinement to finite control alphabets is expected if the channel, itself, is discrete or in the presence of quantizers. Finite $\mathcal{U}$ further encompasses digital controllers, as they constitute dynamic systems evolving on a finite precision algebra. The following proposition is also useful:

Proposition A.3.1 *Let $\mathcal{U}$ be countable, $\mathbf{u}_1^\alpha \in \mathbb{D}$ and $p_{\check{u}(i)} \in \bar{\mathbb{L}}^1$ for $i \in \{1, \ldots, \alpha\}$. If $Var(\mathbf{u}(i)) < \infty$ for $i \in \{1, \ldots, \alpha\}$ then the following holds:*

$$\lim_{k\to\infty} \frac{1}{k} I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k) | \mathbf{z}_1^k) = 0 \tag{A.4}$$

Proof: Since $\mathbf{u}_1^\alpha \in \mathbb{D}$, we can compute $\Delta > 0$ as:

$$\Delta = \inf\{|u - \tilde{u}| : u, \tilde{u} \in \mathcal{U}, u \neq \tilde{u}\} \tag{A.5}$$

We start by means of proposition C.1.1 of Appendix C, we can use $Var(\breve{\mathbf{u}}(i)) = Var(\mathbf{u}(i)) + \frac{\Delta^2}{4} < \infty$ and $p_{u(i)} \in \bar{\mathbb{L}}^1$ to reach the following:

$$h(\breve{\mathbf{u}}(i)) < \infty \tag{A.6}$$

On the other hand, such integral can be related to $H(\mathbf{u}(i))$ as:

$$I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^k) \leq H(\mathbf{u}_1^\alpha) \leq \sum_{i=1}^{\alpha} H(\mathbf{u}(i)) = \sum_{i=1}^{\alpha} h(\breve{\mathbf{u}}(i)) - \log(\Delta) < \infty \tag{A.7}$$

where we use the fact that $p_{\breve{u}(i)}(\breve{u}(i)) = \mathcal{P}(\mathbf{u}(i) = u(i))\frac{1}{\Delta}p_s(\frac{\breve{u}(i)-u(i)}{\Delta})$ and $p_s(s) = 1$ if $s \in (-1/2, 1/2)$ and $p_s(s) = 0$ otherwise. □

A.4 Further Remarks About Assumption (A9)

We start our comments, about (A9), by saying that we require that $p_{\mathbf{w}_1^k, \mathbf{x}(1), \breve{u}_1^k} \in \mathbb{L}^{2k+1}$, just as a way to compactly guarantee that any q dimensional marginal distribution is in $\bar{\mathbb{L}}^q$. This condition is important to ensure that the differential entropy integrals are well defined.

If $\mathcal{U}_1^k$ is not countable and $p_{\mathbf{w}_1^k, \mathbf{x}(1), \breve{u}_1^k}$ is Gaussian then $p_{\mathbf{w}_1^k, \mathbf{x}(1), \breve{u}_1^k} \in \mathbb{L}^{2k+1}$ holds if and only if the covariance matrix of $(\mathbf{w}_1^k, \mathbf{x}(1), \breve{\mathbf{u}}_1^k)$ is positive definite.

We list a few facts of relevance about assumption (A9), for the case where $\mathbf{u}(k)$ is on a countable alphabet:

- Notice that if $\mathcal{U}$ is countable then $Var(\breve{\mathbf{u}}(k)) = Var(\mathbf{u}(k)) + \frac{\Delta^2}{4}$.

- It remains to show that there are no measurability problems, provided that $p_{x(1)}$ is measurable. Assume that $\mathbf{x}(1)$ has a bounded and Lebesgue measurable $p_{x(1)}$. For every Borel set $\mathcal{O} \in \mathcal{R}^{k+1}$, we have:

$$\exists \beta > 0, \mathcal{P}((\mathbf{w}_1^k, \mathbf{x}(1)) \in \mathcal{O}, \mathbf{u}_1^k = u_1^k) \leq \int_{\mathcal{O}} p_{w_1^k}(\gamma_{w_1^k}) p_{x(1)}(\gamma_{x(1)}) \mu(d\gamma_{w_1^k} \times d\gamma_{x(1)}) \leq \beta Vol(\mathcal{O}) \tag{A.8}$$

where we used assumption (A1) to guarantee the existence of a bounded and Lebesgue measurable $p_{w_1^k}$ and causality to split $p_{w_1^k,x(1)} = p_{w_1^k}p_{x(1)}$. As such, from (A.8) and the Radon-Nikodym theorem (pp.422 [9]), we know that, for each $u_1^k \in \mathcal{U}^k$, there exists a measurable probability density function $p_{\mathbf{w}_1^k,\mathbf{x}(1)|u_1^k}(\cdot,\cdot,u_1^k) : \mathbb{R}^{k+1} \to \mathbb{R}_{\geq 0}$. On the other hand, $p_{\mathbf{w}_1^k,\mathbf{x}(1),\breve{u}_1^k}$ is given as:

$$p_{w_1^k,x(1),\breve{u}_1^k}(\gamma_{w_1^k},\gamma_{x(1)},\gamma_{\breve{u}_1^k}) = \sum_{u_1^k \in \mathcal{U}^k} p_{\mathbf{w}_1^k,\mathbf{x}(1)|u_1^k}(\gamma_{w_1^k},\gamma_{x(1)},u_1^k)p_s(\gamma_{\breve{u}_1^k} - u_1^k)\mathcal{P}(\mathbf{u}_1^k = u_1^k) \tag{A.9}$$

$$\mathbf{s} = \breve{\mathbf{u}}_1^k - \mathbf{u}_1^k \tag{A.10}$$

which allow us to infer that $p_{\mathbf{w}_1^k,\mathbf{x}(1),\breve{u}_1^k}$ is a countable linear combination of positive Lebesgue measurable functions. Clearly, besides (A1), we only need to assume bounded and Lebesgue measurable $p_{x(1)}$ to guarantee that $p_{\mathbf{w}_1^k,\mathbf{x}(1),\breve{u}_1^k}$ is bounded and Lebesgue measurable.

Appendix B

Auxiliary Results Used in Chapter 3

B.1 Extension of the Data Processing Inequality

The following theorem provides an extension of the directed data processing inequality, originally derived in [60]. Compared to the version in [60], the result presented bellow allows encoders and decoders that depend on past inputs indexed by $k < 1$. The quantities in the statement of the theorem refer to the scheme depicted in Fig B-1.

Theorem B.1.1 *(Directed Data Processing Inequality, Adaptation of Lemma 4.8.1 of [60]) Let the following assumptions, stated in section 3.2.1 and summarized bellow for convenience, hold:*

- *(A4) The plant is LTI with a state-space representation where $D = 0$ (strictly proper)*

- *(A6) The encoder and decoder are causal operators*

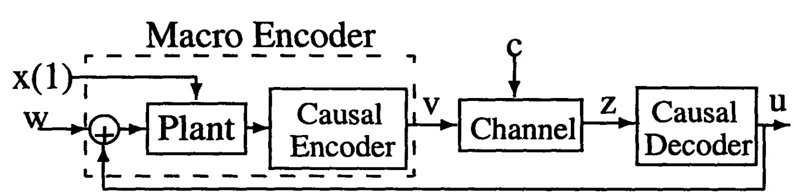

Figure B-1: Casting the feedback loop as a channel in feedback.

- *(A7) The decoder satisfies:*

$$\forall k > \alpha, \mathbf{u}_{\alpha+1}^k = \tilde{f}_k^d(\mathbf{u}_1^\alpha, \mathbf{z}_1^k) \tag{B.1}$$

for some $\alpha \in \mathbb{N}_+$ *and a sequence of functions* $\tilde{f}_k^d$.

- *(A8) The fading memory condition* $\limsup_{k\to\infty} \frac{1}{k} I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^k) = 0$ *holds.*

Under the above conditions, the following is true:

$$\limsup_{k\to\infty} \frac{1}{k} I((\mathbf{x}(1), \mathbf{w}_1^k); \mathbf{u}_1^k) \leq \bar{I}_\infty(\mathbf{v} \to \mathbf{z}) \tag{B.2}$$

Proof: We separate the proof in two parts.

As a **first step** we show that $I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \leq I(\mathbf{v}_1^k \to \mathbf{z}_1^k)$.

Using (P2) we can write the following equality, for any given $i \in \{1, \ldots, k\}$:

$$\begin{aligned} I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})|\mathbf{z}_1^{i-1}) = I(\mathbf{z}(i); \mathbf{v}_1^i|\mathbf{z}_1^{i-1}) + I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})|\mathbf{z}_1^{i-1}, \mathbf{v}_1^i) \\ - I(\mathbf{z}(i); \mathbf{v}_1^i|\mathbf{z}_1^{i-1}, \mathbf{x}(1), \mathbf{w}_1^{i-1}) \end{aligned} \tag{B.3}$$

Now notice that (P2) allows us to rewrite:

$$I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})|\mathbf{z}_1^{i-1}, \mathbf{v}_1^i) = I((\mathbf{z}_1^i, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) - I((\mathbf{z}_1^{i-1}, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) \tag{B.4}$$

But, from (P3), we know that

$$I((\mathbf{z}_1^i, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) = I((\mathbf{c}(i), \mathbf{z}_1^{i-1}, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) \tag{B.5}$$

where we used the fact that, from the definition 3.1.7 (channel), the following map is invertible:

$$(z(i), v(i)) \longmapsto (c(i), v(i))$$

Causality makes $\mathbf{c}(i)$ independent of $(\mathbf{z}_1^{i-1}, \mathbf{v}_1^i, \mathbf{x}(1), \mathbf{w}_1^{i-1})$, so that (B.5) implies the following:

$$I((\mathbf{z}_1^i, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) = I((\mathbf{z}_1^{i-1}, \mathbf{v}_1^i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})) \tag{B.6}$$

By making use of (B.6) and (B.4) we infer that $I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})|\mathbf{z}_1^{i-1}, \mathbf{v}_1^i) = 0$. Such fact, together with (P1) and (B.3), leads to:

$$I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})|\mathbf{z}_1^{i-1}) \leq I(\mathbf{z}(i); \mathbf{v}_1^i|\mathbf{z}_1^{i-1}) \tag{B.7}$$

The first part of the proof is concluded once we notice that, from causality, $\mathbf{w}_i^k$ is independent of $(\mathbf{x}(1), \mathbf{w}_1^{i-1}, \mathbf{z}_1^i)$, which implies:

$$I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^{i-1}) = I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})|\mathbf{z}_1^{i-1}) \tag{B.8}$$

so that (B.7) implies:

$$I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) = \sum_{i=1}^{k} I(\mathbf{z}(i); (\mathbf{x}(1), \mathbf{w}_1^{i-1})|\mathbf{z}_1^{i-1}) \leq \sum_{i=1}^{k} I(\mathbf{z}(i); \mathbf{v}_1^i|\mathbf{z}_1^{i-1}) = I(\mathbf{v}_1^k \rightarrow \mathbf{z}_1^k) \tag{B.9}$$

In the **second step** we prove that:

$$\limsup_{k\rightarrow\infty} \frac{1}{k} I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \leq \limsup_{k\rightarrow\infty} \frac{1}{k} I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k))$$

Once again, we use (P2) to write:

$$I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) = I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) + I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^k) - I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{u}_1^k) \tag{B.10}$$

It follows from (P2), (P4) and assumption (A7) that:

$$I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^k) = I(\mathbf{u}_{\alpha+1}^k; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^k, \mathbf{u}_1^\alpha) + I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^k) = I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^k) \tag{B.11}$$

Substitution of (B.11) in (B.10), together with property (P1), leads to:

$$I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \leq I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) + I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^k) \tag{B.12}$$

Accordingly, (B.12) and the assumption (A8), which requires that

$$\limsup_{k \to \infty} \frac{1}{k} I(\mathbf{u}_1^\alpha; (\mathbf{x}(1), \mathbf{w}_1^k)|\mathbf{z}_1^k) = 0$$

imply that:

$$\limsup_{k \to \infty} \frac{1}{k} I(\mathbf{u}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \leq \limsup_{k \to \infty} \frac{1}{k} I(\mathbf{z}_1^k; (\mathbf{x}(1), \mathbf{w}_1^k)) \tag{B.13}$$

which, together with (B.9), concludes the proof. □

B.2 A Limiting Result for Covariance Matrices

The following is the statement of the *main theorem* of Chapter 5 of [28], repeated here for convenience:

Theorem B.2.1 *(Reproduced from [28], pp.64-65) Let $\hat{F}_e(\omega)$ be a real-valued function of the class $\mathcal{L}_1$ ($|\hat{F}_e(\omega)|$ is integrable in the sense of Lebesgue). We denote by m and M the essential lower bound and upper bound of $\hat{F}_e(\omega)$, respectively, and assume that m and M are finite. If $G(\lambda)$ is any continuous function defined in the finite interval $m \leq \lambda \leq M$, we have:*

$$\lim_{k \to \infty} \frac{\sum_{i=1}^n G(\lambda_i(\Sigma(\mathbf{e}_1^k)))}{k+1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\hat{F}_e(\omega)) d\omega \tag{B.14}$$

Appendix C

Measure Theoretic Aspects of Differential Entropy

C.1 Basic Results

Proposition C.1.1 *If* a *is a random variable with a probability density* $p_a \in \bar{\mathbb{L}}^q$ *then* $h(a) < \infty$.

Proof:(By contradiction) We start by noticing that $p_a \in \bar{\mathbb{L}}^q$ implies that:

$$\int_{\mathbb{R}^q} p_a(\gamma)\gamma^T\gamma d\gamma < \infty \tag{C.1}$$

If we assume that $h(a) = \infty$ then we should have:

$$\int_{p_a(\gamma)\leq 1} -p_a(\gamma)\log p_a(\gamma)d\gamma = \infty \tag{C.2}$$

Since p_a is bounded, (C.2) also implies that:

$$\int_{\Xi} -p_a(\gamma)\log p_a(\gamma)d\gamma = \infty \tag{C.3}$$

$$\Xi = \{\gamma : p_a(\gamma) \leq 1, p_a(\gamma)\gamma^T\gamma < (\gamma^T\gamma)^{-q/2}, (\gamma^T\gamma)^{1+q/2} > e\} \tag{C.4}$$

where we used the fact that p_a is bounded and $\mathbb{R}^q \setminus \Xi$ has finite volume. On the other hand, $-p \log p$ is an increasing function of p for $p < 1/e$ and $p(\gamma) < \frac{1}{(\gamma^T\gamma)^{1+q/2}} < e$ for $\gamma \in \Xi$. These facts imply that $-p_a(\gamma) \log p_a(\gamma) \leq \frac{\log((\gamma^T\gamma)^{1+q/2})}{(\gamma^T\gamma)^{1+q/2}}$ for $\gamma \in \Xi$, but $\int_{\Xi} \frac{\log((\gamma^T\gamma)^{1+q/2})}{(\gamma^T\gamma)^{1+q/2}} d\gamma < \infty$ holds, thus reaching a contradiction.□

Lemma C.1.2 *(Mutual information expressed by means of differential entropy) Let* a *and* b *be random variables that admit* $p_a \in \bar{\mathbb{L}}^q$, $p_b \in \bar{\mathbb{L}}^{q'}$ *and* $p_{a,b} \in \bar{\mathbb{L}}^{q+q'}$, *defined in* $\mathcal{A} \times \mathcal{B} = \mathbb{R}^q \times \mathbb{R}^{q'}$.

The following holds:

$$I(\mathbf{a}; \mathbf{b}) = h(\mathbf{a}) + h(\mathbf{b}) - h(\mathbf{a}, \mathbf{b}) \tag{C.5}$$

Proof:

Fact1: We start by noticing the fact that if s is a random variable with p_s bounded then the following holds:

$$\int_S [p_s(\gamma_s) \log(p_s(\gamma_s))]_+ d\gamma_s < p_s^{sup} \log(p_s^{sup}) \tag{C.6}$$

which, together with $h(\mathrm{s}) < \infty$, implies that $p_s(\gamma_s) \log(p_s(\gamma_s))$ is Lebesgue integrable (see lemma 5 of [2]).

From fact 1, we can use proposition C.1.1, of this Appendix, to conclude that $p_{a,b} \log p_{a,b}$, $p_a \log p_a$ and $p_b \log p_b$ are Lebesgue integrable. As such, using Theorem 7 of [2], we can split the integral of Theorem 2.1.2. of [52] into a sum of three terms as in (C.5). □

C.2 Auxiliary Results for Chapter 3

Lemma C.2.1 *Let the following assumptions hold:*

- $\mathbf{u} \in \mathbb{D}$
- $p_{w_1^k, \breve{u}_1^k, x(1)} \in \mathbb{L}^{2k+1}$

Under the above assumptions, the following holds:

$$h(\mathbf{w}(k)|\mathbf{w}_1^{k-1}, \mathbf{u}_1^k, \mathbf{x}(1)) = h(\mathbf{e}(k)|\mathbf{e}_1^{k-1}, \mathbf{u}_1^k, \mathbf{x}(1)) \tag{C.7}$$

Proof: Since $\mathbf{u} \in \mathbb{D}$, we can use (P3') to write:

$$h(\mathbf{w}(k)|\mathbf{w}_1^{k-1}, \mathbf{u}_1^k, \mathbf{x}(1)) = h(\mathbf{w}(k)) - I(\mathbf{w}(k); (\mathbf{w}_1^{k-1}, \mathbf{u}_1^k, \mathbf{x}(1))) =$$
$$h(\mathbf{w}(k)) - I(\mathbf{w}(k); (\mathbf{w}_1^{k-1}, \breve{\mathbf{u}}_1^k, \mathbf{x}(1))) \tag{C.8}$$

Since $p_{w_1^k, \breve{u}_1^k, x(1)} \in \mathbb{L}^{2k+1}$, using the change of variables $e(k) = w(k)+u(k)$ and integration, we can show that $p_{w_1^{k-1}, \breve{u}_1^k, x(1)} \in \bar{\mathbb{L}}^{2k}$, $p_{e_1^k, \breve{u}_1^k, x(1)} \in \bar{\mathbb{L}}^{2k+1}$, $p_{e_1^{k-1}, \breve{u}_1^k, x(1)} \in \bar{\mathbb{L}}^{2k}$ and $p_{e(k)} \in \bar{\mathbb{L}}^1$. Accordingly, we can use lemma C.1.2, of this appendix, to express (C.8) as:

$$h(\mathbf{w}(k)|\mathbf{w}_1^{k-1}, \mathbf{u}_1^k, \mathbf{x}(1)) = h(\mathbf{w}_1^k, \breve{\mathbf{u}}_1^k, \mathbf{x}(1)) - h(\mathbf{w}_1^{k-1}, \breve{\mathbf{u}}_1^k, \mathbf{x}(1)) \tag{C.9}$$

which, by means of the change of variables[1] $e(k) = w(k) + u(k)$, leads to:

$$h(\mathbf{w}(k)|\mathbf{w}_1^{k-1}, \mathbf{u}_1^k, \mathbf{x}(1)) = h(\mathbf{e}_1^k, \breve{\mathbf{u}}_1^k, \mathbf{x}(1)) - h(\mathbf{e}_1^{k-1}, \breve{\mathbf{u}}_1^k, \mathbf{x}(1)) \tag{C.10}$$

Similarly, we can use lemma C.1.2, of this appendix, to re-express (C.10) as:

$$h(\mathbf{w}(k)|\mathbf{w}_1^{k-1}, \mathbf{u}_1^k, \mathbf{x}(1)) = h(\mathbf{e}(k)) - I(\mathbf{e}(k); (\mathbf{e}_1^{k-1}, \breve{\mathbf{u}}_1^k, \mathbf{x}(1))) \underset{(P3')}{=}$$
$$h(\mathbf{e}(k)) - I(\mathbf{e}(k); (\mathbf{e}_1^{k-1}, \mathbf{u}_1^k, \mathbf{x}(1))) \tag{C.11}$$

which, from the definition of conditional differential entropy, concludes the proof. □

Lemma C.2.2 *Let $V \in \mathbb{R}^{m \times k-k_0+1}$ be a full row-rank matrix. Assume that the following assumptions are satisfied:*

- $\mathbf{u} \in \mathbb{D}$

[1] Notice that this change of variables is legitimate because we can recover $u(k)$ from $\breve{u}(k)$. Such fact results from the existence of the inverse projection $u(k) = \pi_{\mathcal{U}} \breve{u}(k)$

- $p_{w_1^k,\breve{u}_1^k,x(1)} \in \mathbb{L}^{2k+1}$

Under the assumptions above, the following holds[2]*:*

$$h(V\mathbf{w}_{k_0}^k|V\mathbf{u}_{k_0}^k) = h(V\mathbf{e}_{k_0}^k|V\mathbf{u}_{k_0}^k) \tag{C.12}$$

Proof: The proof of this lemma is concluded by following the same steps of lemma C.2.1, mutatis-mutandis by means of the transformation $Ve_{k_0}^k = Vu_{k_0}^k + Vw_{k_0}^k$. □

Lemma C.2.3 *Let the following assumptions hold:*

- $\mathbf{u} \in \mathbb{D}$
- $p_{w_1^k,\breve{u}_1^k,x(1)} \in \mathbb{L}^{2k+1}$

Given a Lipschitz function $f : \mathbb{R}^k \to \mathbb{R}^{dim(x_u(1))}$, under the assumptions above, the following holds:

$$h(\mathbf{x}_u(1)|\mathbf{e}_1^k) = h(\mathbf{x}_u(1) - f(\mathbf{e}_1^k)|\mathbf{e}_1^k) \tag{C.13}$$

Proof: Since f is Lipschitz, we know that if $(\mathbf{w}_1^k, \mathbf{u}_1^k, \mathbf{x}(1))$ has a finite covariance matrix then $(\mathbf{w}_1^k, \mathbf{u}_1^k, \mathbf{x}_u(1) - f(\mathbf{e}_1^k))$ also has a finite covariance matrix. As such, we can use changes of variables and integration to show that $p_{w_1^k,\breve{u}_1^k,x(1)} \in \mathbb{L}^{2k+1}$ implies $p_{e_1^k,x_u(1)} \in \bar{\mathbb{L}}^{k+dim(x_u(1))}$, $p_{e_1^k,x_u(1)-f(e_1^k)} \in \bar{\mathbb{L}}^{k+dim(x_u(1))}$, $p_{x_u(1)} \in \bar{\mathbb{L}}^{k+dim(x_u(1))}$, $p_{e_1^k} \in \bar{\mathbb{L}}^k$ and $p_{x_u(1)-f(e_1^k)} \in \bar{\mathbb{L}}^{dim(x_u(1))}$. These facts allow us to use lemma C.1.2 freely to write:

$$h(\mathbf{x}_u(1)|\mathbf{e}_1^k) = h(\mathbf{x}_u(1), \mathbf{e}_1^k) - h(\mathbf{e}_1^k) \tag{C.14}$$

By applying a change of variables in (C.14), we get:

$$h(\mathbf{x}_u(1)|\mathbf{e}_1^k) = h(\mathbf{x}_u(1) - f(\mathbf{e}_1^k), \mathbf{e}_1^k) - h(\mathbf{e}_1^k) \tag{C.15}$$

We finish by recognizing that (C.15) is equal to $h(\mathbf{x}_u(1) - f(\mathbf{e}_1^k)|\mathbf{e}_1^k)$ □

[2]Here we adopt an abuse of notation by using $Ve_{k_0}^k$, $Vu_{k_0}^k$ and $Vw_{k_0}^k$ to denote matrix multiplication. For instance, $Ve_{k_0}^k = V\begin{bmatrix} e(k) & \cdots & e(k_0) \end{bmatrix}^T$.

Appendix D

Auxiliary results for Chapter 4

D.1 Auxiliary results leading to the proof of theorem 4.4.1

The main result of this section is lemma D.1.6, where explicit expressions are given to bound the probability (4.16). In order to guarantee notational simplicity, we start with the following definitions:

Definition D.1.1 *We define the following random variable:*

$$\mathbf{z}_k = -\frac{\ln p^{\mathbf{W}_{1,k}}(\mathbf{W}_{1,k})}{kn^F} \tag{D.1}$$

We also adopt $\mathcal{D} = \frac{\mathcal{D}_d}{n^F}$ as a way to make equations smaller.

Lemma D.1.1 *For any given positive definite $\Sigma_X \in \mathbb{R}^{n^X \times n^X}$, designate by $\mathbf{X}$ the zero-mean Gaussian random variable with covariance matrix Σ_X. The following holds:*

$$\sup_{\Sigma_X > 0} \left(\mathcal{E}[\left(\frac{\ln p^{\mathbf{X}}(\mathbf{X})}{n^X}\right)^2] - \left(\mathcal{E}[\frac{\ln p^{\mathbf{X}}(\mathbf{X})}{n^X}]\right)^2 \right) < \frac{a}{n^X} \tag{D.2}$$

where $p^{\mathbf{X}}(X) = \frac{e^{-\frac{1}{2}X^T \Sigma_X^{-1} X}}{\left((2\pi)^{n^X} |\Sigma_X|\right)^{1/2}}$ and $a = \frac{1}{2}$.

Proof: The result is obtained by evaluating the expected values in (D.2) and applying the change of variables $\tilde{X} = \Sigma_X^{-1/2} X$. □

Lemma D.1.2 *Given $\epsilon > 0$ and $k \geq 1$ the following holds:*

$$\mathcal{P}\left(\left|\mathbf{z}_k - \frac{r^W}{n^F}\right| > \epsilon\right) < \frac{1}{2k\epsilon^2 n^F} \tag{D.3}$$

Proof: Since $\mathbf{W}_k$ is i.i.d. and $\Sigma_W > 0$, lemma D.1.1 and the fact that $W_k \in \mathbb{R}^{kn^F}$ lead to:

$$Var(\mathbf{z}_k) \leq \frac{1}{2kn^F} \tag{D.4}$$

Now notice that $\mathcal{E}[\mathbf{z}_k] = \frac{r^W}{n^F}$, so that the final result is a direct application of the Chebyshev-Bienaymé inequality [9]. □

Definition D.1.2 *Given $\epsilon \in (0,1)$ and $n \in \mathbb{N}$, define Γ_k as the event that the search space of the decoder is empty:*

$$\Gamma_k = \begin{cases} \textit{True} & \textit{if } \{\bar{q}_{1,k} \in \mathbb{T}^{\epsilon,k} : s^{\epsilon,k}(\bar{q}_{1,k}) = True\} = \emptyset \\ \textit{False} & \textit{otherwise} \end{cases} \tag{D.5}$$

Lemma D.1.3 *For any given $\epsilon \in (0,1)$, if $k \geq 1$, then the following holds:*

$$\lim_{k \to \infty} P(\Gamma_k) = 0 \tag{D.6}$$

Proof:

Notice that, from definition 4.3.4, if a given realization $q_{1,k}$ satisfies $q_{1,k} \in \mathbb{T}^{\epsilon,k}$ and $s^{\epsilon,k}(q_{1,k}) = True$, then Γ_k is false because $q_{1,k}$ is itself a valid choice for $\bar{q}_{1,k}$. That leads to the following inequality:

$$\mathcal{P}(\Gamma_k) \leq \mathcal{P}(|\mathbf{z}_k - \frac{r^W}{n^F}| > \epsilon) + \mathcal{P}(\mathbf{q}_{1,k} \notin \mathbb{T}^{\epsilon,k}) \tag{D.7}$$

The first term in the RHS of (D.7) can be bounded by means of lemma D.1.2. For the second term it suffices to show that $\mathcal{P}(|\frac{-\log_m p^{\mathbf{q}}(\mathbf{q}_{1,k})}{k} - r^q| > \epsilon) \underset{k \to \infty}{\longrightarrow} 0$. Our analysis starts with the expansion:

$$\log_m p^{\mathbf{q}}(\mathbf{q}_{1,k}) = \sum_{i=2}^{k} \log_m p^{\mathbf{q}}(\mathbf{q}_i|\mathbf{q}_{i-1}) + \log_m p^{q}(\mathbf{q}_1) \tag{D.8}$$

Since $\mathcal{P}\left(|\frac{\log_m p^q(\mathbf{q}_1)}{k}| > \frac{\epsilon}{2}\right) \underset{k\to\infty}{\longrightarrow} 0$, we only have to show that:

$$\mathcal{P}\left(|\frac{-\sum_{i=2}^{k} \log_m p^{\mathbf{q}}(\mathbf{q}_i|\mathbf{q}_{i-1})}{k} - r^q| > \frac{\epsilon}{2}\right) \underset{k\to\infty}{\longrightarrow} 0 \tag{D.9}$$

If $r^q = 0$, then $\mathcal{P}(\log_m p^{\mathbf{q}}(\mathbf{q}_i|\mathbf{q}_{i-1}) = 0) = 1$ and the result is proved. For $r^q > 0$, recall that $\mathbf{q}_i$ is ergodic. Since $\mathbf{q}_k$ has a finite alphabet, there are no unboundedness problems and we conclude that $\log_m p^{\mathbf{q}}(\mathbf{q}_i|\mathbf{q}_{i-1})$ is ergodic, which implies (D.9). □

Lemma D.1.4 *For any given $q_{1,k}, \hat{q}_{1,k} \in \mathbb{A}^k$ and $k \geq 1$ the following holds:*

$$\max_{X \in \mathbb{R}^{kn^F}} p^{\mathbf{T}_{1,k}(q_{1,k},\hat{q}_{1,k})}(X) = m^{k\frac{n^F}{\ln(m)}(-\mathcal{D}(q_{1,k},\hat{q}_{1,k}) - \frac{r^W}{n^F} + \frac{1}{2})} \tag{D.10}$$

Proof: From the definition of the Gaussian distribution:

$$\max_{X \in \mathbb{R}^{kn^F}} p^{\mathbf{T}_{1,k}(q_{1,k},\hat{q}_{1,k})}(X) = \frac{\left|\Sigma_{\mathbf{T}_{1,k}(q_{1,k},\hat{q}_{1,k})}\right|^{-1/2}}{(2\pi)^{\frac{kn^F}{2}}} \tag{D.11}$$

or equivalently, using (4.8), it can also be written as:

$$\max_{X \in \mathbb{R}^{kn^F}} p^{\mathbf{T}_{1,k}(q_{1,k},\hat{q}_{1,k})}(X) = m^{-\frac{\mathcal{H}\left(\mathbf{T}_{1,k}(q_{1,k},\hat{q}_{1,k})\right) + \frac{kn^F}{2}}{\ln(m)}} \tag{D.12}$$

The final result is achieved once we recall that $\mathcal{D}(q_{1,k}, \hat{q}_{1,k}) = \frac{\mathcal{H}\left(\mathbf{T}_{1,k}(q_{1,k},\hat{q}_{1,k})\right)}{kn^F} - \frac{r^W}{n^F}$. □

Lemma D.1.5 *Let $k \geq 1$, $\epsilon \in (0,1)$, $q_{1,k} \in \mathcal{A}^k$ and $\hat{q}_{1,k} \in \mathbb{T}^{\epsilon,k}$. The following is an upper-bound for the conditional probability that $\hat{q}_{1,k}$ satisfies the decoding condition $s^{\epsilon,k}(\hat{q}_{1,k}) = True$.*

$$P(\mathbf{s}^{\epsilon,k}(\hat{q}_{1,k}) = True|q_{1,k}) < m^{k\frac{n^F}{\ln(m)}(-\mathcal{D}(q_{1,k},\hat{q}_{1,k}) + \frac{1}{2} + \epsilon)} \tag{D.13}$$

Proof: The structure of the decoder leads to :

$$P(\mathbf{s}^{\epsilon,k}(\hat{q}_{1,k}) = True|q_{1,k}) = \int_{\mathbb{S}} p^{\mathbf{T}_{1,k}(q_{1,k},\hat{q}_{1,k})}(X)dX \tag{D.14}$$

where $\mathbb{S}$, the set of realizations leading to (4.10), is given by:

$$\mathbb{S} = \{X \in \mathbb{R}^{kn^F} : \frac{-\ln p^{\mathbf{W}_{1,k}}(X)}{kn^F} < \frac{r^W}{n^F} + \epsilon\} \tag{D.15}$$

We can use lemma D.1.4 to infer that:

$$P(\mathbf{s}^{\epsilon,k}(\hat{q}_{1,k}) = True|q_{1,k}) \le \int_{\mathbb{S}} m^{k\frac{n^F}{\ln(m)}\left(-\mathcal{D}(q_{1,k},\hat{q}_{1,k}) - \frac{r^W}{n^F} + \frac{1}{2}\right)} dX \le$$

$$\le Vol(\mathbb{S}) m^{n^F \frac{k}{\ln(m)}\left(-\mathcal{D}(q_{1,k},\hat{q}_{1,k}) - \frac{r^W}{n^F} + \frac{1}{2}\right)} \tag{D.16}$$

The proof is completed once we notice that the volume of $\mathbb{S}$ is upper-bounded by $Vol(\mathbb{S}) < m^{k\frac{n^F}{\ln(m)}(\frac{r^W}{n^F}+\epsilon)}$. □

Lemma D.1.6 (Main lemma) *Given $\beta \in \mathbb{R}_+$ and $\epsilon \in (0,1)$, the following holds:*

$$\mathcal{P}(\mathcal{D}(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \beta) < \mathcal{P}(\mathbf{\Gamma}_k) + m^{k\frac{n^F}{\ln(m)}\left(\ln(m)\frac{r^q}{n^F} - \beta + \frac{1}{2} + \frac{\ln(m)+n^F}{n^F}\epsilon\right)} \tag{D.17}$$

where $\mathcal{P}(\mathbf{\Gamma}_k)$ is given in lemma D.1.3.

Proof:

We separate the two events that may generate $\mathcal{D}(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \beta$, and write the following bound:

$$\mathcal{P}(\mathcal{D}(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \beta) \le \mathcal{P}(\mathbf{\Gamma}_k) + \sum_{q_{1,k}\in\mathbb{A}^k} \sum_{\hat{q}_{1,k}\in\mathbb{D}(q_{1,k})} P(\mathbf{s}^{\epsilon,k}(\hat{q}_{1,k}) = True|q_{1,k}) p^q(q_{1,k}) \tag{D.18}$$

where $\mathbb{D}(q_{1,k}) = \{\hat{q}_{1,k} \in \mathbb{T}^{\epsilon,k} : \mathcal{D}(q_{1,k}, \hat{q}_{1,k}) > \beta\}$. Using lemma D.1.5 and the inequality above, we get:

$$\mathcal{P}(\mathcal{D}(\mathbf{q}_{1,k}, \hat{\mathbf{q}}_{1,k}) > \beta) \le \left(\sharp\mathbb{T}^{\epsilon,k}\right) m^{k\frac{n^F}{\ln(m)}(-\beta+\epsilon+\frac{1}{2})} + \mathcal{P}(\mathbf{\Gamma}_k) \tag{D.19}$$

The fact that $\sharp\mathbb{T}^{\epsilon,k} \leq m^{k(r^q+\epsilon)}$ concludes the proof. □

Bibliography

[1] G. A. Ackerson and K. S. Fu. On state estimation in switching environments. *IEEE Trans. on Automatic Control*, 15(1):10–17, 1970.

[2] M. Adams and V. Guillemin. *Measure Theory and Probability*. Birkhauser, 1996.

[3] S. Arimoto and H. Kimura. Optimum input test signals for system identification - an information-theoretical approach. *Int. J. Syst. Sci.*, 1(3):279–90, 1971.

[4] Y. Bar-Shalom and X. R. Li. *Estimation and Tracking: Principle, Techinques and Software*. Boston, MA:Artech House, 1993.

[5] G. Battail. On random-like codes. *Information Theory and Applications II, Lecture notes in Computer Science 1133*, pages 76–94, 1995.

[6] G. Belforte and P. Gay. Optimal worst case estimation for lpv-fir models with bounded errors. In IEEE, editor, *Proceedings of the IEEE Conference on Decision and Control*, pages 4573–4577, 2000.

[7] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Eglewood Cliffs, NJ:Prentice Hall, 1971.

[8] D. P. Bertsekas. *Dynamic Programming and Optimal Control.* Volume One, Athena Scientific, 1995.

[9] P. Billingsley. *Probability and measure*. Wiley Inter-5Science, 3rd Ed., 1995.

[10] S. Bittanti, M.C. Campi, and L. Guo. Persistence of excitation properties for the identification of time-varying systems. *IEEE Conf. on Decision and Control*, pages 680–684, 2000.

[11] R. E. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Trans. on Inf. Theory*, 18:460–473, 1972.

[12] H. A. P. Blom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with markovian switching coefficients. *IEEE Trans. A. C.*, (8):780–783, 1988.

[13] H. Bode. *Network Analysis and Feedback Amplifier Design*. D. Van Nostrand, Princeton, 1945.

[14] R. W. Brockett and D. Liberzon. Quantized feedback stabilization of linear systems. *IEEE Trans. Automat. Control*, 45:1279–1289, 2000.

[15] Edited by Richard M. Murray. Control in an information rich world: Report of the panel on future directions in control, dynamics, and systems. SIAM, 2002.

[16] J. J. Campbell and L. M. Viceira. *Strategic Asset Allocation*. Oxford University Press, 2002.

[17] M. C. Campi. Exponentially weighted least squares identification of time-varying systems with white disturbances. *IEEE Tr. Sig. Proc.*, 42:2906–14, 1994.

[18] C.B. Chang and M. Athans. State estimation for discrete systems with switching parameters. *IEEE Trans. on Aerospace and Electronic Systems*, 14(3):418–425, 1978.

[19] European Comission. European commission workshop on future and emerging control systems, June 2000.

[20] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[21] N. Elia. Stabilization of systems with erasure actuation and sensory channels. In *Proc. 40th Allerton Conference in Communication, Control and Computing*, 2002.

[22] N. Elia. Control-oriented feedback communication schemes. In IEEE, editor, *Proceedings of the IEEE Conference on Decision and Control*, pages 3161–3166, 2003.

[23] N. Elia. Stabilization in the presence of fading channels. In IEEE, editor, *Proc. of American Control Conference*, pages 4438 – 4443, 2003.

[24] Robert Elliot. *Hidden Markov Models, Estimation and Control.* Springer-Verlag, 1995.

[25] M. Feder and N. Merhav. Relations between entropy and error probability. *IEEE Trans. on Inf. Theory*, 40:259–266, 1994.

[26] P. J. Fernandez. *Medida e Integração (In Portuguese).* Instituto de Matemática Pura e Aplicada, Projeto Euclides, Segunda Edição, 1996.

[27] J. S. Freudenberg and D. P. Looze. *Frequency Domain Properties of Scalar and Multivariable Systems.* Springer-Berlin, 1988.

[28] U. Grenander and G. Szego. *Toeplitz Forms and Their Applications.* University of California Press, 1958.

[29] S. Gunarsson and L. Ljung. Adaptation and tracking in system identification - a survey. *Automatica*, 26:7–21, 1990.

[30] S. Gunnarson. On the quality of recursively identified fir models. *IEEE Tr. Sig. Proc.*, 40:679–82, 1994.

[31] P. R. Halmos. *Measure Theory.* Springer Verlag, 1974.

[32] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and G. Duda. Performance anomaly of 802.11b. In IEEE, editor, *Proceedings of the IEEE Infocom*, 2003.

[33] M. W. Hofbaur and B. C. Williams. Mode estimation of probabilistic hybrid systems. *HSCC 2002, LNCS 2289*, pages 253–266, 2002.

[34] R. Horn and C. Johnson. *Matrix Analysis.* New York, NY: Cambridge University Press, 1985.

[35] R. Jain, T. Simsek, and P. Varayia. Control under communication constraints. In IEEE, editor, *Proceedings of the IEEE Conference on Decision and Control*, pages 3209–15, 2002.

[36] Hanseok Ko and R.H. Baran. Entropy and information rates for hidden markov parameters. In IEEE, editor, *Proc.IEEE ISIT*, page 374, 1998.

[37] K. Li and J. Baillieul. Robust quantization for digital finite communication bandwidth (dfcb) control. In IEEE, editor, *Proceedings of the IEEE Conference on Decision and Control*, pages 3130 – 3135, 2003.

[38] X. Li and W.S. Wong. State estimation with communication constraints. *Systems and Control Letters*, 28:49–54, 1996.

[39] D. Liberzon. On stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 48:304–7, 2003.

[40] D. Liberzon. On stabilization of non-linear systems with limited information feedback. *IEEE Conf. on Decision and Control*, pages 182–186, 2003.

[41] N. C. Martins and M. A. Dahleh. An information theoretic approach to the modal estimation of fir linear systems. In IEEE, editor, *Proceedings of the IEEE Conference on Decision and Control*, 2003.

[42] N. C. Martins and M. A. Dahleh. Fundamental limitations of distrubance attenuation in the presence of finite capacity feedback. In University of Illinois, editor, *Allerton Conference on Communication, Control and Computing*, 2004.

[43] N. C. Martins and M. A. Dahleh. Rate distortion trade-off in the modal estimation of switching fir linear systems. *Proceedings of the IEEE Conf. on Decision and Control*, 2004.

[44] N. C. Martins, M. A. Dahleh, and N. Elia. Stabilization of uncertain systems in the presence of a stochastic digital link. *Proceedings of the IEEE Conf. on Decision and Control*, 2004.

[45] N. C. Martins, M. A. Dahleh, and N. Elia. Stabilization of uncertain systems in the presence of finite data-rate feedback. In IFAC, editor, *Proceedings of the IFAC Nolcos, Germany*, 2004.

[46] N. C. Martins, S. Venkatesh, and M. A. Dahleh. Controller design and implementation for large scale systems, a block decoupling approach. *Proceedings of the IEEE Amreican Control Conference*, 2001.

[47] J. Massey. Causality, feedback and directed information. In ISITA-90, editor, *Proc. of the 1990 Int. Symp. on Information Theory and its Applications (ISITA-90)*, pages 303–305, 1990.

[48] E. Mosca. Probing signal design for linear channel identification. *IEEE Trans. on Information Theory*, 18(4):481–487, 1972.

[49] G. N. Nair, S. Dey, and R. J. Evans. Infimum data rates for stabilizing markov jump linear systems. In IEEE, editor, *Proceedings of the IEEE Conference on Decision and Control*, pages 1176–81, 2003.

[50] G. N. Nair and R. J. Evans. Stabilization with data-rate-limited feedback: Tightest attainable bounds. *Systems and Control Letters*, 41:49–76, 2000.

[51] G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran. Topological entropy and nonlinear stabilization. *to appear in the Special Issue on Networked Control Systems, IEEE Transactions on Automatic Control*, 2004.

[52] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden Day, 1964.

[53] L. R. Rabiner. A tutorial on hidden markov model and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 2003.

[54] A. Sahai. Evaluating channels for control: Capacity reconsidered. In IEEE, editor, *Proc. of American Control Conference*, pages 2358 – 2362, 2000.

[55] A. Sahai. *Anytime Information Theory.* PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2001.

[56] B. R. Upadhyaya H. W. Sorenson. Bayesian discriminant approach to input signal selection in parameter estimation problems. *Information Sciences*, 1(12):61–91, 1977.

[57] D. D. Sworder and J. E. Boyd. *Estimation Problems in Hybrid Systems.* Cambridge University Press, 1999.

[58] D. D. Sworder, J. E. Boyd, and R. J. Elliott. Modal estimation in hybrid systems. *Journal of Math. Analis. and Ap.*, (245):225–247, 2000.

[59] S. Tatikonda. Control under communication constraints: Parts i and ii. *submitted to the IEEE Transactions on Automatic Control.*

[60] S. Tatikonda. *Control under Communication Constraints.* PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2000.

[61] S. Tatikonda, A. Sahai, and S. Mitter. Control of lqg systems under communication constraints. In IEEE, editor, *Proc.Conference on Decision and Control*, 1998.

[62] Jitenda K. Tugnait. Detection and estimation for abruptly changing systems. *IFAC - Automatica*, 18(5):607–615, 1982.

[63] S. R. Venkatesh and M. A. Dahleh. On system identification of complex systems from finite data. *IEEE Trans. A. C.*, 46(2):235–257, 2001.

[64] H. L. Weidemann. Entropy analysis of parameter estimation. *Inf. and Control*, (14):493–506, 1969.

[65] H. Witsenhausen. Separation of estimation and control for discrete-time systems. *Proceedings of the IEEE*, 59, 1971.

[66] W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints -i: State estimation problems. *IEEE Trans. Automat. Control*, 42:1294–1298, 1997.

[67] W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints -ii: Stabilization with limited information feedback. *IEEE Trans. Automat. Control*, 44:1049–1053, 1999.

[68] Y. Dong Y. Rate distortion analysis of pose estimation via hidden markov model. In IEEE, editor, *Proc.IEEE ICASP*, pages 2985–2988, 2002.

[69] S. Yuksel and T. Basar. Quantization and coding for decentralized lti systems. In IEEE, editor, *Proc.Conference on Decision and Control*, 2003.

[70] G. Zames, L. Lin, and L. Y. Wang. Fast identification n-widths and uncertainty principles for lti and slowly varying systems. *IEEE Trans. A. C.*, 39(9):1827–1838, 1994.

[71] G. Zang and P. A. Iglesias. Nonlinear extension of bode's integral based on an information theoretic interpretation. *Systems and Control Letters*, 50:11–19, 2003.