University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Zafzouf, Ghassen

Title:
Uniformly Bounded State Estimation over Multiple Access Channels

Date:
2022

Persistent Link:
https://hdl.handle.net/11343/324349

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

Uniformly Bounded State Estimation over Multiple Access Channels

Ghassen Zafzouf

ORCID: 0000-0002-9447-9598

Submitted in partial fulfillment for the degree of

Doctor of Philosophy

Department of Electrical and Electronic Engineering

THE UNIVERSITY OF MELBOURNE

June 2022

Department of Electrical and Electronic Engineering
The University of Melbourne
Title: Uniformly Bounded State Estimation over Multiple Access Channels
Author: Ghassen Zafzouf

To the best of my knowledge this work contains no material which has been previously published by any other person except where due reference or acknowledgement has been given. Further I certify that none of the work embodied in this report has been submitted for any other examination.

Melbourne, 30/06/2022

Place, Date (Ghassen Zafzouf)

Epigraph



This page intentionally left blank.

Acknowledgments

It's been a long road, but here I am at the end. Many people directly or indirectly influenced my journey, and this is the perfect opportunity to extend my thanks to them. First and foremost, I would like to express my deepest appreciation and sincerest gratitude to my advisor Prof. Girish Nair for his encouragement, his generous research spirit, and his unwavering confidence in my abilities. I am honored and feel privileged to have had the opportunity to work with you and cannot thank you enough. Not only did your ideas and insights help shape this thesis, but your passion and determination for tackling new problems taught me what good research is truly about. But of the many facets of your teaching, it is your admirable character that made the deepest impression on me and gave the most important instruction. You have been a constant source of ideas and inspiration.

Next, I would like to sincerely thank Dr. Farhad Farokhi for his continuous support, patience, enthusiasm and immense knowledge. This thesis would not have been possible without your fruitful guidance, Farhad. From you, I have learned to strive for clarity and express complex and abstract ideas in simple, yet precise statements. You constructively challenged the depths of my potential and always pushed me to de my best. I truly feel fortunate to have worked by your side.

Many thanks are also due to Prof. Jamie Evans for the numerous discussions we had together. Your invaluable input was always a source of different perspectives in the initial stages of my PhD. I have the utmost respect and gratitude for your guidance.

A big thank you goes to Prof. Margreta Kuijper, Dr. Rajitha Senanayake, Brice Shen and Dr. Brian Krongold for giving me the chance to make my first steps in the teaching experience under their excellent mentorship. Thank you sincerely for all of the continuous support and encouragement you provided.

I am also grateful to my former and current colleagues Xiangyue Meng, Seyed Salman Ahmadi, Neil Irwin Bernardo, Amir Saberi, Victor Deville, and Max Varley with whom I could discuss much more than research. Their warm friendship created a great atmosphere in the department and made my time even more enjoyable. Thank you!

To all my friends whose support and care helped me overcome the challenges of my journey, I would like to express my deepest gratitude for your sincere friendship. A special thanks go to Wassim Zaouali and Hiba Arnout for always being there all along in spite of the long distance. I am deeply indebted to you. I am also thankful to my close friends, Muhamed Minhaj, Mohammed Basheti, Sai-

fullah Shazin, Abdallah Youssef, Mohamed Abd El-Kader, Affan and Hibbaan Nawaz. You have all added so much charm to my journey.

The final word of gratitude must go to my closest loved ones, those who shared my ups and downs by long distance. My parents Mohamed Habib Zafzouf and Boutheina Hajaiej listened, advised, and supported me in all my endeavors. Thank you for patiently listening to my everyday stories, for sharing my joy when I succeed and for cheering me up when I fail. I would not have made it this far if not for you. I am deeply grateful for that.

Ghassen Zafzouf
Melbourne, June 2022

Uniformly Bounded State Estimation over Multiple Access Channels

Doctor of Philosophy

Ghassen Zafzouf

Abstract

In this doctoral thesis, a characterization of the zero-error capacity region for three different classes of *multiple access channels* (MACs) is derived. The first type of channels considered in this work is a two-user MAC with a common message that captures the correlation between transmitters. Next, this model is extended by considering an arbitrary number of users $M \geq 2$. The last class of MACs represents a further extension to a more general case where inter-user correlation is modeled by a common message seen by all users as well as pairwise shared messages. In this research, we look at the zero-error capacity, which differs from the more commonly studied small-error capacity, from a nonprobabilistic angle. In fact, the obtained characterization is based on the so-called *nonstochastic information*, denoted by I_*, and is valid not only for asymptotically large coding block-lengths but also for finite lengths. Understanding how to coordinate unambiguous communication through MACs, such that several unrelated senders can simultaneously send as much information as possible is of great interest, especially with the emergence of new paradigms such as the *Internet of Things* (IoT) and *Machine-to-Machine* (M2M) communication.

Next, using the characterization of the zero-error capacity region for the two-user MAC, we investigate the problem of distributed state estimation under the criterion of uniformly bounded estimation errors. It is shown that if there exists a coder-estimator tuple that achieves the desired criterion, namely uniformly bounded estimation error, the vector of topological entropies of the linear systems, whose state is being estimated, must lie within the zero-error capacity region of the communication channel. Additionally, we prove that if the to-be-observed plants have a topological entropy vector inside the interior of the zero-error capacity region, the existence of a coder-estimator tuple achieving uniformly bounded state estimation errors is guaranteed. This result relates the channel properties to the plant dynamics and paves the way toward understanding information flows in networked control systems with multiple transmitters.

Finally, we seek to characterize the fundamental tradeoff between the communication data rate, code-length, system dynamics and state estimation performance. To this end, a universal lower bound on the time-asymptotic estimation error is obtained using volume-based analysis. Additionally, to provide a guarantee on the estimation performance, an upper bound on the error is derived when

the measurements are quantized. When the code-length is large, we show that these lower and upper bounds converge to the same limit.

Preface

This thesis is submitted in total fulfillment of the requirements of the degree of Doctor of Philosophy at the University of Melbourne. It contains the work that I conducted during my PhD candidature in the Department of Electrical and Electronic Engineering, School of Engineering, The University of Melbourne (Australia). This thesis has not been submitted for any other qualification. All the work towards the thesis was carried out after the enrolment in the degree. No third party editorial assistance was provided in the preparation of the thesis. The research described here was supervised by Prof. Girish Nair and Dr. Farhad Farokhi. They have supervised the overall development of this research with technical guidance, fruitful discussions and constructive feedback. They have also proofread and scrutinized this thesis.

To the best of my knowledge, this work is original, except where acknowledgements and references are made to previous work. The main contributions of this thesis are presented in Chapters 3, 4, 5 and 6. For the work in these chapters and the associated papers listed below, I contributed greater than 60%, i.e., I remain the primary author and investigator. My work included the problem formulation, theoretical analysis, computer simulations and manuscript writing. The summary of publications related to this thesis is given below.

Journal Papers

- G. Zafzouf, G. N. Nair and F. Farokhi, "Uniformly Bounded State Estimation over Multiple Access Channels," under review. *Submitted to IEEE Transactions on Automatic Control.*

Conference Proceedings

- G. Zafzouf and G. N. Nair, "Zero-Error Capacity Region of a Class of Multiple Access Channels with Inter-User Correlation," *2020 IEEE Information Theory Workshop (ITW)*, Apr. 2021.
- G. Zafzouf and G. N. Nair, "Distributed State Estimation with Bounded Errors over Multiple Access Channels," *2020 IEEE International Symposium on Information Theory (ISIT)*, Los Angeles, USA, June 2020.
- G. Zafzouf, G. N. Nair and J. S. Evans, "Zero-Error Capacity of Multiple Access Channels via Nonstochastic Information," *2019 IEEE Information Theory Workshop (ITW)*, Visby, Sweden, Aug. 2019.

My research was partially supported by the Australian Research Council via my primary supervisor's grant FT140100527, which I gratefully acknowledge. My PhD candidature was also supported by the Australian Government and the University of Melbourne through the Melbourne Research

Scholarship (MRS). I acknowledge the University of Melbourne and Prof. Girish Nair for funding my attendance at the Information Theory Workshop conference in Visby, Sweden in August 2019 and IEEE International Symposium on Information Theory in June 2020.

Contents

List of Figures xiii

List of Tables 1

1 Introduction 3
- 1.1 Motivation . . . 3
- 1.2 Problem of State Estimation . . . 5
- 1.3 Literature Review . . . 7
- 1.4 Outline & Contributions . . . 8

2 Background 13
- 2.1 Preliminaries on Nonstochastic Information Theory . . . 13
  - 2.1.1 Uncertain Variables, Unrelatedness and Markovianity . . . 13
  - 2.1.2 Overlap Connectedness . . . 15
  - 2.1.3 Nonstochastic Information I_* . . . 15
  - 2.1.4 Maximal Common Variables . . . 17
  - 2.1.5 Nonstochastic Conditional Information . . . 18
- 2.2 Discrete Memoryless Channels (DMCs) . . . 19
- 2.3 Multiple Access Channels (MACs) . . . 20
- 2.4 Notion of Zero-Error Channel Capacity . . . 21

3 Zero-Error Capacity Region of Multiple Access Channels 23
- 3.1 Zero-Error Communication over Two-User MAC with Common Message . . . 23
  - 3.1.1 System Model . . . 24
  - 3.1.2 Convexity of the Zero-Error Capacity Region of Two-User MAC . . . 26
  - 3.1.3 Non-Emptiness of the Interior of the Zero-Error Capacity Region of Two-User MAC . . . 27
  - 3.1.4 Zero-Error Capacity Region of Two-User MAC with Common Message via Nonstochastic Information . . . 29
- 3.2 Zero-Error Communication over M-User MAC with Common Message . . . 32
  - 3.2.1 System Model . . . 32
  - 3.2.2 Zero-Error Capacity of M-User MAC with Common Message via Nonstochastic Information . . . 35
- 3.3 M-User MAC with Common Message & Pairwise Shared Messages . . . 38
  - 3.3.1 System Model & Notation . . . 38

3.3.2 Zero-Error Capacity Region of the M-User MAC with Common Message & Pairwise Shared Messages via Nonstochastic Information . . . 42
3.4 Summary . . . 46

4 Distributed State Estimation over Nonstochastic MACs with Bounded Noise 47
4.1 Problem Formulation . . . 48
4.2 Uniformly Bounded State Estimation over MACs . . . 50
4.2.1 Necessity Proof . . . 51
4.2.2 Sufficiency Proof . . . 56
4.3 Generality of the Proposed Setup for Noiseless Linear Systems . . . 59
4.4 Numerical Example: The Binary Adder Channel (BAC) . . . 64
4.4.1 Zero-Error Capacity Region of the BAC . . . 64
4.4.2 Application to the State Estimation Problem (Theorem 5) . . . 65
4.5 Summary . . . 70

5 Universal Lower Bounds on the State Estimation Errors 71
5.1 Problem Formulation . . . 71
5.2 Centralized Lower Bound . . . 72
5.3 Decentralized Lower Bounds . . . 76
5.4 Tightness of Decentralized Lower Bounds for Scalar Systems . . . 80
5.4.1 System Model . . . 80
5.4.2 Proof of Tightness of Decentralized Lower Bounds . . . 81
5.5 Comparison between Centralized and Decentralized Lower Bounds . . . 82
5.5.1 Numerical Results for Scenario 1 . . . 83
5.5.2 Numerical Results for Scenario 2 . . . 83
5.6 Generalization of the Centralized Lower Bound . . . 86
5.6.1 Analytical Analysis . . . 86
5.6.2 Numerical Example . . . 88
5.6.3 Analysis of the Large n Regime . . . 89
5.7 Summary . . . 92

6 Asymptotically Tight Upper Bounds on the State Estimation Error 95
6.1 System Construction . . . 96
6.2 Upper Bound on the State Estimation Error for Scalar Jordan Blocks . . . 98
6.3 Upper Bound on the State Estimation Error for Non-Scalar Jordan Blocks with Real Eigenvalues . . . 103
6.4 Upper Bound on the State Estimation Error for Non-Scalar Jordan Blocks with Complex Eigenvalues . . . 107

6.5 The Optimal Rate Allocation Policy for Scalar Jordan Blocks 111
6.5.1 Convexity of the RHS of (6.2.18) 112
6.5.2 Analysis of the Optimization Problem 113
6.5.3 Analysis of the Large n Regime 117
6.6 Numerical Example 124
6.7 Summary 127

7 Conclusion **129**
7.1 Thesis Summary 129
7.2 Future Research 131
Bibliography I

Bibliography **I**

This page intentionally left blank.

List of Figures

1.1 The state vector of a dynamic system is observed by separate sensors. Each sensor incorporates an observer and a channel encoder. In this setup, the sensors and their encoders have uncoordinated access to the channel, i.e., shared channel with *no a priori* resource allocation. 5

1.2 Illustrative model of the state estimation problem over point-to-point noiseless communication channel. 6

2.1 Notion of uncertain variable (uv). 14

2.2 Illustrative example of **(a)** Overlap disconnected points; **(b)** Overlap connected points. 16

2.3 Illustrative example of a family of conditional ranges for three realizations w_1, w_2 and w_3. 18

2.4 Illustration of a communication system transmitting over a memoryless channel defined by a fixed function $f(\cdot)$. 19

2.5 Channel transition diagram of **(a)** Binary Erasure Channel with $C_0 = 0$; **(b)** Pentagon Channel with $C_0 > 0$. 21

3.1 The two-transmitter MAC system with a common message operating at time instant k. 24

3.2 Illustrative example of the unique overlap partition, namely the family $[\![U|Y_{1:n}]\!]_*$. The horizontal lines represent the different member sets of the partition and the black filled-in circles correspond to the selected points $u(w^{(i)})$ for $i \in \{0,1,2\}$. . . . 31

3.3 The M-user MAC system with a common message $W^{(0)}$ operating at time instant k. . 34

3.4 Illustration of a communication system deploying a three-user MAC with special correlation structure. The red-colored sources produce private messages, while the outputs of the green sources are seen by two encoders. The blue block refers to the source that generates the common message accessible by all three encoders. 40

4.1 State estimation of LTI systems with disturbances over a two-input single-output MAC channel. 48

4.2 Structure of encoder $\gamma^{(l)}$ for $l \in \{1,2\}$. Note that the symbol "$\downarrow n$" denotes the downsampling block by a factor of n. 56

4.3 Structure of decoder δ. Note that the symbol "$\uparrow n$" denotes the upsampling block by a factor of n. 56

4.4 Illustrative figure of the zero-error capacity region's boundary in a two-dimensional space, i.e., $R^{(0)} = 0$. The region shadowed in blue corresponds to a part of the channel's [zoomed-in] zero-error capacity region $\mathscr{C}_0$. 58

4.5 The states of a noiseless (process and measurement) LTI system are detected by two distinct sensors. Each sensor incorporates an observer and a channel encoder. In this scenario, we consider uncoordinated access strategies with no apriori resource allocation modeled as a MAC. 59
4.6 Block diagram of the construction used for lower bound tightness proof. 67
4.7 Example of state realizations $x^{(1)}, x^{(2)}$ and their corresponding estimations $\hat{x}^{(1)}, \hat{x}^{(2)}$ for unstable systems 1 and 2 on a *logarithmic* scale. 68
4.8 Scenario 1. The topological entropy vector $h \in \mathrm{int}(\mathscr{C}_0)$, and the code rates $R^{(1)} > h^{(1)}$ and $R^{(2)} > h^{(2)}$. 69
4.9 Empirical maximum error norms on a *logarithmic* scale for Scenario 2. The topological entropy vector $h \notin \mathrm{int}(\mathscr{C}_0)$, and the zero-error code rates $R^{(1)} < h^{(1)}$ and $R^{(2)} < h^{(2)}$. 70

5.1 Decentralized lower bound $\mathbf{LB}^{\mathrm{decent}}$ for different $R^{(1)}$ and $R^{(2)}$ using the *max norm*, where $h^{(1)}$ and $h^{(2)}$ being of the same order. 84
5.2 Centralized lower bound $\mathbf{LB}^{\mathrm{cent}}$ for different $R^{(1)}$ and $R^{(2)}$ using the *max norm*, with $h^{(1)}$ and $h^{(2)}$ being of the same order. 84
5.3 Decentralized lower bound for different $R^{(1)}$ and $R^{(2)}$ using the *max norm* and $h^{(1)} = 10 \cdot h^{(2)}$. 85
5.4 Centralized lower bound for different $R^{(1)}$ and $R^{(2)}$ using the *max norm* and $h^{(1)} = 10 \cdot h^{(2)}$. 85
5.5 The function $\mathfrak{f}(\mathfrak{J})$ evaluated for different system settings. 90
5.6 Behavior of the lower bound (5.6.12) for different state dimensions $|\mathfrak{J}|$. For the purpose of this plot, the entropy sum was fixed to $\sum_{j\in\mathfrak{J}} h^{(j)} = 0.1$. 91

6.1 Behaviour of $U^{(i)}$ (6.6.7) and the lower bound (5.6.16) on the state estimation error of the i-th system for different configurations of the topological entropy vector $(h_1^{(i)}, h_2^{(i)})^T$ and pull-back factor $\delta^{(i)}$. The rate $R^{(i)}$ is set to 0.7203 bits/use. Note that the results on the LHS are plotted for $\delta^{(i)} = 0.1$, whereas the ones on the RHS are obtained for $\delta^{(i)} = 0.01$. . 128

7.1 A high-level model of the two-receiver BC operating at time instant k. 132
7.2 A high-level model of the two-receiver IC operating at time instant k. 132

List of Tables

4.1 Simulation Parameters. *The code block-length is set to* $n = 6$. 68

5.1 System parameters of Scenario 1 and 2. The system parameters are chosen such that the topological entropies of the plants are of the same order in Scenario 1, whereas, in Scenario 2, the second plant has a tenfold larger topological entropy. The code block-length is set to 6 to be consistent with the numerical examples and analysis presented in previous parts of this thesis. 83

5.2 Evaluation of RHS (5.6.10) for systems with the same total entropy $H = 1$ bit, but different state dimensions d and entropy vectors $h := (h^{(j)})_{j=1}^{d}$. 88

6.1 Summary of the problem parameters and their expressions, with $\beta \in [1 : B^{(i)}]$ and $i \in \{0,1,2\}$. 118

6.2 Domains of definition of $U^{(i)}$ and $r_{\beta}^{(i)}$, with $\beta \in [1 : B^{(i)}]$ and $i \in \{0,1,2\}$. 118

This page intentionally left blank.

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."

— Marie Curie (1867–1934)

1 Introduction

In 1948 Claude Shannon succeeded in providing a mathematical description of an abstract concept, namely the notion of *information*. In his famous research paper *"A Mathematical Theory of Communication"* [1], Shannon revolutionized the field of communications engineering by giving a mathematically precise definition for random variables (rv) in a probability space. He was particularly interested in defining a quantity that will measure, in some sense, how much information is "produced" by a source of some generic nature, i.e., text, image, sound or even video. Shannon suggested that the logarithm of the number of elements in the message set generated by the considered source can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. By modeling noise and data sources as statistical processes with known probability distributions, concepts of *entropy* and *mutual information* were introduced. Unlike its use in thermodynamics, Shannon's entropy describes *a priori uncertainty* of a rv in a broader sense. In the early 1940s, the belief among the communication theory community was that negligible probability of error could only be guaranteed at large *signal-to-noise* (SNR) regime. However, Shannon succeeded to prove that decoding error probabilities could be made nearly zero even with low SNR. The *operational* interpretation of the concepts proposed by Shannon has made *information theory* today the basis of modern digital communications. In fact, it turns out that a stationary random process such as speech cannot be compressed below its entropy rate; and the maximal mutual information across a channel corresponds to the *channel capacity*, i.e., the highest block-coding rate.

1.1 Motivation

In spite of the decisive role of information theory in digital communications, it has not played an important role in the field of control theory. As outlined in [2], one major reason for the lack of use

of this theory to study dynamic systems is mainly the probabilistic model of *uncertainty*. In fact, most control systems are employed for safety- and/or mission-critical purposes, with any error leading to potentially devastating consequences. Hence, the system performance must be guaranteed *every* time the plant is operating. As for communication systems, we are only interested in the performance on average, i.e., usually, if some bits are erroneous the consequences are not as critical as in the case of an expensive plant. Furthermore, the disturbances occurring in control systems cannot be always modeled as probability distributions based on physical laws since they might contain mechanical and chemical components rather than just electronic or electromagnetic noise like communication systems. This motivates system designs based on the worst-case scenario [2–5]. Therefore, in order to analyze the performance of networked control systems, it is natural to investigate the possibility of constructing useful analogues of the stochastic concepts mentioned previously, *without assuming a probability space*.

Traditionally, classical control and estimation theory has assumed that communication between different components of a network occurs over point-to-point links. This standard assumption however cannot hold for recent emerging applications, where numerous subsystems are interacting with each other. In these applications, it may be impractical or costly to set up and maintain multiple physical point-to-point links between multiple subsystems. One solution is to use a shared medium, such as wireless, with frequency and/or time divided up into non-interfering slots, each for dedicated use by a single transmitter. In the case of time division, this leads to deterministic round-robin-like protocols for channel access [6]. However, such methods are generally sub-optimal, since they do not fully exploit channel resources in order to increase throughput. For real-time applications, the delays caused by a round-robin scheduling protocol may also degrade performance significantly. Event-based time scheduling is an alternative that allows sensors to transmit more quickly when their measurements exceed certain levels or increase rapidly [7]. This allows communication resources to be used where they are needed most. However, an additional network layer is typically required to make sure network access goes to the device that most needs it.

An alternative is to allow users to transmit their messages simultaneously over the shared channel and use coding and modulation to mitigate inter-user interference in addition to noise and other channel effects. Since resources such as specific time slots or frequency bands are not pre-allocated to each user, this method potentially allows faster data flows to be achieved. In real-time applications with a large number of transmitters, it could also improve latency compared to a time-division approach. A simple model of simultaneous communication is *multiple access channel* (MAC) consisting of different users who aim to each send an independent message reliably to a common receiver. This was initially introduced by Shannon in his seminal work [8].

With this in mind, the aim of this thesis is to understand the problem of remote estimation of dynamical systems over MACs with bounded noise rather than assuming a statistical distribution. In this context, we try to answer the following fundamental questions: is it possible to provide a reliable

estimation of the states of distinct plants observed by different sensors whose measurements are sent simultaneously over a multiple access channel? If so, what is the connection between the intrinsic properties of the dynamical systems and the transmission data rates?

1.2 Problem of State Estimation

A fundamental problem in the field of dynamic systems and automated control is that of *state estimation*. Simply put, the problem can be described as follows. A dynamic system of interest, e.g., an Unmanned Autonomous Vehicle (UAV), a manufacturing plant or an advanced aircraft, wishes to achieve a specified objective. To generate appropriate control inputs, the system needs to know its current *state*, such as its position and velocity. The *true* state is, however, rarely available to the system, but it rather needs to be estimated using some noisy measurements obtained by the system's sensors. A *sensor* is a device that detects and converts a physical property into a signal. Due to hardware imperfections, the conversion process of physical quantities into electric signals is far from perfect, and the uncorrectable error introduced is known as *measurement noise*. Therefore, using all of the available information about the sensors and their measurements, the system has to obtain an accurate *estimate* of its state.

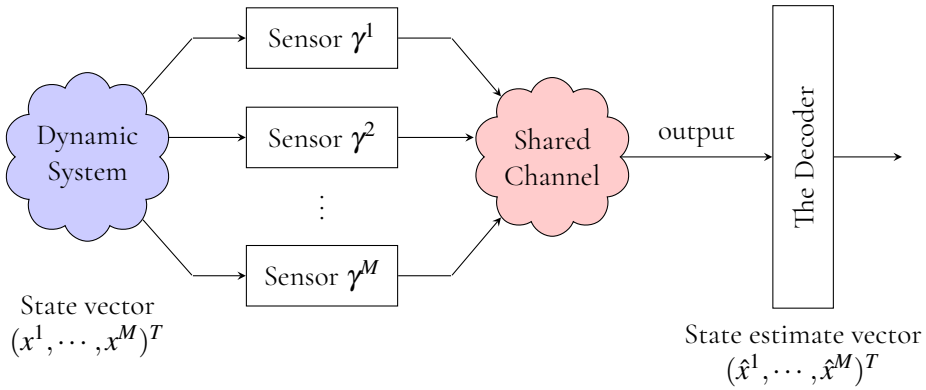


Figure 1.1: The state vector of a dynamic system is observed by separate sensors. Each sensor incorporates an observer and a channel encoder. In this setup, the sensors and their encoders have uncoordinated access to the channel, i.e., shared channel with *no a priori* resource allocation.

By virtue of the important progress in the fields of electronics, nanotechnology and computing, the world has witnessed an unprecedented development of smaller and cheaper sensors able of carrying out real-time computation and communicating over wireless channels. The availability of such sensors to a wide range of users has contributed, among other factors, to a tremendous increase in the number of connected devices and the emergence of new technological concepts such as the *Internet of Things* (IoT) and *Machine-to-Machine* (M2M) communication [9]. These technologies have enabled



Figure 1.2: Illustrative model of the state estimation problem over point-to-point noiseless communication channel.

information processing and exchange within a *network* of geographically dispersed devices that are referred to as *nodes*. A schematic of such a setup is shown in Fig. 1.1. In centralized state estimation, all nodes can send their measurements to *one* central computational unit where the system's state is estimated [10]. This technological transition requires a new set of network specifications including simultaneous communication, low latency, high device availability, and minimum maintenance and operation costs.

To appreciate the importance of understanding the impact of the communication channel limitations on the task of state estimation even under the simplest circumstances, we consider the example of a scalar unstable linear time-invariant (LTI) system whose dynamics are described by

$$x[k+1] = \lambda x[k],\ k \in \mathbb{Z}_{\geq 0} \tag{1.2.1}$$

where $|\lambda| > 1$ and the initial state is assumed to be bounded, i.e., $x[0] \leq \Upsilon < \infty$, where $\Upsilon \in \mathbb{R}$. As depicted in Fig. 1.2, the measurements of the sensor observing the plant's state are quantized and encoded before being communicated over an *errorless* q-ary channel, i.e., a channel with an input alphabet size equal to q. At the receiver, the decoder/estimator block produces an estimate $\hat{x}[k]$ of the system's dynamics using the channel output. The design criterion in this setting is to keep the estimation error bounded, i.e.,

$$\sup_{k \in \mathbb{Z}_{\geq 0}} |x[k] - \hat{x}[k]| < \infty. \tag{1.2.2}$$

Before being encoded, the observed initial state $x[0]$ undergoes first a quantization process as follows: A q-level quantizer incorporated within the sensor divides the interval $[-\Upsilon, \Upsilon]$ into q sub-intervals of the same length $\frac{2\Upsilon}{q}$. The encoder, then, maps the true state into the *closest* midpoint (using the Euclidean norm by way of example).

Under the assumption that the decoder knows the scalar λ and the interval $[-\Upsilon, \Upsilon]$, it is possible at the receiver to produce an estimate $\hat{x}[0]$ of the initial state. After a total of K channel uses, i.e., $k = \{0, 1, \cdots, K-1\}$, the estimator refines its estimate of the initial state to achieve an error of at most $\sup |x[0] - \hat{x}[0]| = (\Upsilon / q^K)$. Subsequently, at time instant $k = K$, the estimation error would

have evolved with time to reach

$$\begin{aligned}\sup|x[K]-\hat{x}[K]| &= |\lambda|^K \sup|x[0]-\hat{x}[0]| \\ &= \frac{\Upsilon}{q^K}|\lambda|^K. \end{aligned} \tag{1.2.3}$$

From (1.2.3), one could see that in order to maintain the estimation error bounded in the sense of (1.2.2), the following must hold

$$|\lambda| \leq q. \tag{1.2.4}$$

This inequality, namely (1.2.4), reflects the underlying connection between the dynamics of the observed plant and the communication channel. This result is known in literature as the *data-rate theorem* [11, §1.3]. For more details about the problem of state estimation over point-to-point memoryless channels for both undisturbed and disturbed LTI systems, please refer to [3].

1.3 Literature Review

The decisive role of information theory in digital communications arises largely from the fundamental coding bounds it provides. To approach these bounds, coding schemes with arbitrarily long block lengths are used, which result in a vanishingly small probability of decoding errors. However, long block-lengths are ill-suited for real-time control systems since they lead to long delays that degrade closed-loop performance significantly. Furthermore, in real-time control and estimation applications, where there are safety requirements or mission-critical objectives, closed-loop performance is often quantified in a worst-case sense, rather than probabilistically.

An important step towards understanding communication requirements for worst-case state estimation was taken in [12], where it was shown that, to achieve almost surely (a.s.) uniformly bounded state estimation error for a linear, disturbed dynamical system over a stochastic discrete memoryless point-to-point channel, it is necessary that the open-loop *topological entropy* h does not exceed the *channel zero-error capacity*[1] C_0. If $C_0 > h$, then there exists a coding and estimation scheme that achieves a.s. uniformly bounded state estimation error. By noting that C_0 depends on the combinatorial structure of the channel rather than its probabilistic nature, this result was rederived in [3] for surely bounded estimation error, by introducing the framework of *uncertain variables* (uv's) and *nonstochastic information* I_*. In a recent article [13], the authors introduced the notion of uncertain wiretap channel and studied the problem of secure state estimation for a noisy unstable dynamical system using the nonstochastic setup. Tools from this non-probabilistic framework were also used to

[1] In contrast to the conventional Shannon capacity C, the channel zero-error capacity C_0 is the highest rate leading to an error probability at the decoder which is *exactly* zero. This notion is further discussed in Section 2.4.

study the problem of bounded state estimation over point-to-point channels with finite memory [14]. The main focus in previous works was on studying either state estimation or control problem in the context of point-to-point channels. Stochastic stabilization of networked control systems over multiple wireless channels has been extensively studied in, e.g., [15,16]. To the best of our knowledge, the problem of state estimation over a MAC with bounded noise has not been considered yet. An exception is the work of Zaidi *et al.* [17], where classical stochastic tools were used to present sufficient conditions ensuring the mean-square stabilization of two scalar linear time-invariant (LTI) systems over a noisy two-input, single-output MAC. The paper [18] also obtains necessary and sufficient conditions for stabilizing two scalar plants across a shared Gaussian MAC. In that paper, the authors distinguish between the case where encoders are entirely independent of each other and the case where information sharing among them is allowed.
However, before proceeding with studying the problem of worst-case state estimation over MACs, the zero-error capacity region $\mathscr{C}_0$ of the communication channel must first be characterized. In fact, unlike its ordinary small-error capacity region $\mathscr{C}$ which has been extensively studied in the literature [19, 20], little is known about the MACs' $\mathscr{C}_0$, and to date, no formula for such a region has been obtained and $\mathscr{C}_0$ of MACs, in general, remains an open problem. For instance, for deterministic binary adder channel (BAC), which is a particular example of MACs[2], the best outer bound on this region has been obtained by Austrin *et al.* in [21] which presents a slight improvement on the result of Ordentlich and Shayevitz [22]. These studies rely mainly on combinatorics in order to tighten the outer bound of $\mathscr{C}_0$ and reduce the gap with its inner bound [23]. A more complex model than the MAC with two independent inputs corresponds to the generalized scenario where two or more users are allowed to transmit arbitrarily correlated messages. By considering a two-user, three-message MAC [24], Slepian and Wolf were the first to introduce the concept of MAC with correlated sources. In their model, both transmitters have access to a common message in addition to their respective private ones. In [25], Han extended this model to a class of multi-user MACs with correlated sources. More recently, Gündüz and Simeone [26] have further refined these results by reducing the number of auxiliary variables involved in the analysis by means of a special message hierarchy. Next, we discuss the contributions of this thesis and the gaps that are addressed.

1.4 Outline & Contributions

This thesis studies the problem of distributed state estimation over multiple access channels (MACs) with bounded noise. The rest of this dissertation is organized as follows. In Chapter 2, a thorough background and literature review is presented. In this part, we motivate the use of multiple access channels (MACs) as a suitable model of several transmitters sending their messages simultaneously over a shared communication medium. Moreover, we discuss the notion of zero-error capacity

[2] See Section 4.4.1 for a detailed presentation of this channel model.

of a channel as a fundamental performance limit quantifying the number of messages that may be reliably communicated with no error. Finally, basic definitions related to the proposed nonprobabilistic framework are introduced and the concept of *nonstochastic information* is described.
Next, in Chapter 3, the theory of nonstochastic information is applied to characterize the zero-error capacity region for three classes of the so-called *multiple access channels* (MACs), namely

- Two-user MAC with common message,
- $\boldsymbol{M}$-user MAC with one common message and $\boldsymbol{M} \geq 2$, and
- $\boldsymbol{M}$-user MAC with a common message and pairwise shared messages.

Unlike conventional information-theoretic proofs of channel coding theorems, where the code block-length is often assumed to be asymptotically large, the analysis presented in this dissertation is also true for *finite* block-lengths.
Using the characterization of the zero-error capacity region for the two-user MAC with common message, the problem of distributed state estimation over such channels with bounded noise is addressed in Chapter 4 under the criterion of uniformly bounded estimation errors. The setup considered here consists of three discrete noisy LTI dynamical systems connected to two sensors. More specifically, two plants are assumed to be observed in total privacy, i.e., each by a separate sensor, whereas the output of the remaining system is available to both devices. Though limited to three systems, this configuration encapsulates some of the essential elements of the problem of distributed state estimation, e.g., where each sensor observes a different subset of the overall system's dynamical modes. Under suitable assumptions, we show that if there exists a coder-estimator tuple yielding uniformly bounded estimation errors, then

$$\boldsymbol{h} \in \mathscr{C}_0, \tag{1.4.1}$$

where $\boldsymbol{h}$ is the vector of *topological entropies*[3] of the corresponding systems and $\mathscr{C}_0$ denotes the zero-error capacity region of the MAC. On the other hand, if $\boldsymbol{h} \in \mathbf{int}\,(\mathscr{C}_0)$[4], then a coder-estimator tuple that achieves uniformly bounded state estimation errors can be constructed. In other words, condition (1.4.1) is both necessary and sufficient for every point inside the interior region of $\mathscr{C}_0$ and it is shown to be necessary for all points on the boundary. This condition is considered tight as sufficiency and necessity differ only on the boundary, which is a set of measure zero. We also prove, in this chapter, the generality of the proposed setup in the case of noiseless LTI systems. In fact, it turns out that an undisturbed system observed by two sensors can be decomposed into three separate systems, namely private plants observed only by the respective sensor and a common system observed by both

[3] The notion of topological entropy of a system is defined as the sum-log of its unstable pole magnitudes.
[4] The operator $\mathbf{int}\,(\cdot)$ denotes the interior of a given region.

devices. Lastly, a numerical example of a setup where the communication occurs over a binary adder channel is presented and discussed.

Next, following the establishment of the necessary and sufficient conditions under uniformly bounded estimation error criterion over two-user MAC with a common message, Chapter 5 seeks to characterize the fundamental trade-off between the communication data rate, code-length, system dynamics and the state estimation performance. By exploiting the properties of nonstochastic information theory, universal lower bounds on the time-asymptotic max norm of the estimation errors are derived. We first start our analysis by treating the communication channel as a point-to-point channel, which allowed us to obtain a lower bound on the overall performance of the whole system. Hence, this lower bound is described as *centralized*. Afterwards, by virtue of the distributed nature of the considered setup, we refine this result and derive *decentralized* lower bounds associated with each of the estimated sub-systems. We then show that this bound is tight for one-dimensional systems by presenting a construction of an encoder-estimator tuple that achieves the lower bound. Moreover, the centralized lower bound is generalized to the case of a dynamic system with an arbitrary number of decoupled states, and it is then shown that in the large block-length regime, the estimation performance is mainly driven by the dominant eigenvalue.

At this stage of the thesis, we aim to provide a guarantee on the estimation performance of the previously discussed system. To this end, we derive an upper bound on the error when the measurements are quantized in Chapter 6. Furthermore, each state is estimated at a share of the total rate available at hand. Three configurations of the state matrix are considered in this chapter, namely

- one-dimensional Jordan blocks,
- non-scalar Jordan blocks with real eigenvalues, and
- non-scalar Jordan blocks with complex eigenvalues.

We eventually show there that, for asymptotically large code block-lengths and hypercuboidal process noise, the obtained lower and upper bounds converge to the same point.

Finally, the dissertation is concluded with a summary of the main findings followed by a discussion of possible future research directions in Chapter 7.

The novel work obtained in this thesis was published in the following conference and journal papers:

Journal Papers

- G. Zafzouf, G. N. Nair and F. Farokhi, "Uniformly Bounded State Estimation over Multiple Access Channels," under review. *Submitted to IEEE Transactions on Automatic Control.*

Conference Proceedings

- G. Zafzouf and G. N. Nair, "Zero-Error Capacity Region of a Class of Multiple Access Channels with Inter-User Correlation," *2020 IEEE Information Theory Workshop (ITW)*, Apr. 2021.
- G. Zafzouf and G. N. Nair, "Distributed State Estimation with Bounded Errors over Multiple Access Channels," *2020 IEEE International Symposium on Information Theory (ISIT)*, Los Angeles, USA, June 2020.
- G. Zafzouf, G. N. Nair and J. S. Evans, "Zero-Error Capacity of Multiple Access Channels via Nonstochastic Information," *2019 IEEE Information Theory Workshop (ITW)*, Visby, Sweden, Aug. 2019.

This page intentionally left blank.

"Alles Gescheite ist schon gedacht worden.
Man muss nur versuchen, es noch einmal zu denken."

"All intelligent thoughts have already been thought;
what is necessary is only to try to think them again."

— Johann Wolfgang von Goethe (1749–1832)

2
Background

We devote this chapter to presenting background concepts that are used in the rest of this thesis. To this end, we start with an overview of nonstochastic information theory in Section 2.1. Next, we explain the concept of memoryless channels in Section 2.2. Afterwards, we discuss an example of a multiple access channel (MAC), namely the Gaussian MAC, in Section 2.3. Finally, Section 2.4 aims at clarifying the notion of zero-error channel capacity using some well-known examples.

2.1 Preliminaries on Nonstochastic Information Theory

In this section, we review the *uncertain variable* (uv) framework introduced in [2–4]. The notions reviewed here will later be used to study the problem of zero-error communication over different classes of MACs.

2.1.1 Uncertain Variables, Unrelatedness and Markovianity

Consider the sample space Ω as shown in Fig. 2.1. An *uncertain variable* (uv) X consists of a mapping from Ω to a set $\mathscr{X}$ [3]. Hence, each sample $\omega \in \Omega$ induces a particular realization $X(\omega) \in \mathscr{X}$. Given a pair of uv's X and Y, the *marginal*, *joint*, and *conditional ranges* are respectively denoted by

$$[\![X]\!] := \{X(\omega) : \omega \in \Omega\} \subseteq \mathscr{X}, \tag{2.1.1}$$

$$[\![X,Y]\!] := \{(X(\omega),Y(\omega)) : \omega \in \Omega\} \subseteq \mathscr{X} \times \mathscr{Y}, \tag{2.1.2}$$

$$[\![X|y]\!] := \{X(\omega) : Y(\omega) = y, \omega \in \Omega\} \subseteq \mathscr{X}. \tag{2.1.3}$$

The marginal range and conditional ranges together fully determine the joint range as follows

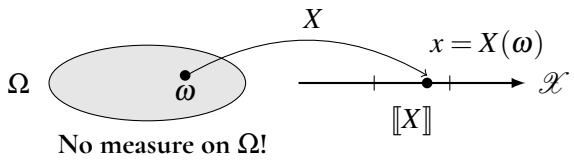


Figure 2.1: Notion of uncertain variable (uv).

$$[\![X,Y]\!] = \bigcup_{x\in[\![X]\!]} [\![Y|x]\!] \times x.$$

This is similar to the way that joint distributions are determined by the conditional and marginal ones. Note also that the family $\{[\![X|y]\!] : y \in [\![Y]\!]\}$ of conditional ranges is denoted $[\![X|Y]\!]$.

As in probability theory, the dependence on Ω will normally be hidden with most properties of interest expressed in terms of operations on these ranges. As a convention, uv's are denoted by upper-case letters while their realizations are indicated in lower-case. Furthermore, the expression $[a:b]$ represents the sequence of integers $\{a, a+1, \cdots, b-1, b\}$.

Definition 1 (Unrelatedness [3]). *A finite collection $X_1, X_2, \cdots X_n$ of uv's are said to be (mutually) unrelated if*

$$[\![X_1, X_2, \cdots X_n]\!] = [\![X_1]\!] \times [\![X_2]\!] \times \cdots [\![X_n]\!]. \tag{2.1.4}$$

Remark 1. *Unrelatedness, which is closely related to the notion of* qualitative independence *[27] between discrete sets, can be shown to be equivalent to the conditional range property*

$$[\![X_k|x_{1:k-1}]\!] = [\![X_k]\!], \quad \forall x_{1:k-1} \in [\![X_{1:k-1}]\!],\ \forall k \in [2:n]. \tag{2.1.5}$$

Definition 2 (Conditional Unrelatedness [3]). *The uv's $X_1, ..., X_n$ are said to be* conditionally unrelated given Y *if*

$$[\![X_1, \ldots, X_n|y]\!] = [\![X_1|y]\!] \times \cdots \times [\![X_n|y]\!],\ \forall y \in [\![Y]\!]. \tag{2.1.6}$$

Definition 3 (Markovianity [3]). *The uv's X_1, Y and X_2 form a* Markov uncertainty chain *denoted as $X_1 \leftrightarrow Y \leftrightarrow X_2$, if*

$$[\![X_1|y, x_2]\!] = [\![X_1|y]\!],\ \ \forall (y, x_2) \in [\![Y, X_2]\!]. \tag{2.1.7}$$

Remark 2. *It can be shown that Def. 3 is equivalent to $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ being conditionally unrelated given $\boldsymbol{Y}$, i.e.,*

$$[\![X_1, X_2|y]\!] = [\![X_1|y]\!] \times [\![X_2|y]\!], \ \forall y \in [\![Y]\!]. \tag{2.1.8}$$

By the symmetry of (2.1.8), we conclude that $\boldsymbol{X}_1 \leftrightarrow \boldsymbol{Y} \leftrightarrow \boldsymbol{X}_2$ iff $\boldsymbol{X}_2 \leftrightarrow \boldsymbol{Y} \leftrightarrow \boldsymbol{X}_1$.

Remark 3 (Markovianity for a general sequence of uv's)**.** *We consider a sequence $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ of $n \geq 3$ uv's. We say that this sequence forms a Markov uncertainty chain $\boldsymbol{X}_1 \leftrightarrow \boldsymbol{X}_2 \leftrightarrow \cdots \leftrightarrow \boldsymbol{X}_n$ if*

$$[\![X_{1:k-1}, X_{k+1}|x_k]\!] = [\![X_{1:k-1}|x_k]\!] \times [\![X_{k+1}|x_k]\!], \ \forall x_k \in [\![X_k]\!], \ \forall k \in [2:n-1]. \tag{2.1.9}$$

This means that the conditional range of $\boldsymbol{X}_{k+1}$ given past realizations depends only on the most recent one. This is also equivalent to the reversal-invariant property

$$[\![X_{1:k-1}, X_{k+1:n}|x_k]\!] = [\![X_{1:k-1}|x_k]\!] \times [\![X_{k+1:n}|x_k]\!], \ \forall x_k \in [\![X_k]\!], \ \forall k \in [2:n-1], \tag{2.1.10}$$

i.e., the future and past of the sequence are conditionally unrelated given the present.

2.1.2 Overlap Connectedness

In order to define a notion that quantifies the information that can be gained about a uv $\boldsymbol{X}$ given another uv $\boldsymbol{Y}$, namely nonstochastic information, we must first define a notion of connectedness to describe the structural properties of the family of conditional ranges $[\![X|Y]\!]$.

Definition 4 (Overlap Connectedness [3])**.** *Two points x and $x' \in [\![X]\!]$ are said to be $[\![X|Y]\!]$-overlap connected, denoted $x \leftrightsquigarrow x'$, if there exists a finite sequence $\{[\![X|y_i]\!]\}_{i=1}^m$ of conditional ranges such that $x \in [\![X|y_1]\!]$, $x' \in [\![X|y_m]\!]$ and $[\![X|y_i]\!] \cap [\![X|y_{i-1}]\!] \neq \emptyset$, for each $i \in [2:m]$.*

Obviously, the overlap connectedness is both transitive and symmetric, i.e., it is an equivalence relation. Thus, it results in equivalence classes that cover $[\![X]\!]$ and form a unique partition. We call this family of sets the $[\![X|Y]\!]$-*overlap partition* and denote by $[\![X|Y]\!]_*$. Every set $\mathscr{G} \in [\![X|Y]\!]_*$ can be expressed as follows

$$\mathscr{G} = \{x \in [\![X]\!] : x \leftrightsquigarrow \mathscr{G}\} = \bigcup_{\mathscr{P} \in [\![X|Y]\!] : \mathscr{P} \leftrightsquigarrow \mathscr{G}} \mathscr{P}. \tag{2.1.11}$$

2.1.3 Nonstochastic Information I_*

Using the notions described previously, it is possible now to measure the amount of information $I_*[X;Y]$ common to two uv's $\boldsymbol{X}$ and $\boldsymbol{Y}$. This framework allows information-theoretic tools to be used to investigate problems where statistical structure is unavailable. It will be shown in this section that



Figure 2.2: Illustrative example of **(a)** Overlap disconnected points; **(b)** Overlap connected points.

the proposed nonstochastic information allows us to quantify how much information both variables share [2].

Definition 5 (Nonstochastic Information [3]). *The* nonstochastic information *between* X *and* Y *is given by*

$$I_*[X;Y] = \log_2 |[\![X|Y]\!]_*| \,. \tag{2.1.12}$$

Remark 4. *Note that the nonstochastic information is symmetric, i.e.,* $I_*[X;Y] = I_*[Y;X]$

> **Example:** Consider uv's X and Y with conditional range family $[\![X|Y]\!] = \{[\![X|y_1]\!], [\![X|y_2]\!]\}$. Fig. 2.2(a) illustrates an example of overlap disconnected points. Observe that $[\![X|y_1]\!] \cap [\![X|y_2]\!] = \emptyset$, and hence, the unique overlap partition $[\![X|Y]\!]_*$ consists of two singleton sets. Thus, the nonstochastic information in this case is $I_*[X;Y] = \log_2 2 = 1$ bit. On the other hand, Fig. 2.2(b) shows the case where the points are connected in overlap sense. In this case, it is easy to see that $[\![X|y_1]\!] \cap [\![X|y_2]\!] \neq \emptyset$ and $[\![X|Y]\!]_*$ does no longer consist of two singleton sets. Hence, $I_*[X;Y] = \log_2 1 = 0$ bit.

Similar to mutual information, $I_*[X;Y]$ comes with the following properties:

- **Nonnegativity.** The nonnegativity of $I_*[X;Y]$ can be directly deduced from (2.1.12). Thus,

$$I_*[X;Y] \geq 0. \tag{2.1.13}$$

- **Monotonicity.** Using the maximal cv interpretation of I_*, it can be shown that, for any uv's X, Y and W,

$$I_*[X;Y,W] \geq I_*[X;Y]. \tag{2.1.14}$$

- **Data Processing Inequality.** For any Markov uncertainty chain $W \leftrightarrow X \leftrightarrow Y$,

$$I_*[W;Y] \leq I_*[X;Y]. \tag{2.1.15}$$

In other words, inner uv pairs in the chain share more information than outer ones.

2.1.4 Maximal Common Variables

Consider the basic problem in which two agents observe realizations of X and Y separately. These agents apply separate functions to the corresponding observations to produce outputs; and wish to eventually agree on a common value. In other words, the first and second agents apply functions f and g to the realizations x and y respectively so that $f(x) = g(y)$ for all $(x,y) \in [\![X,Y]\!]$. Consequently, the common variable (cv) Z can be expressed as

$$Z := f(X) = g(Y). \tag{2.1.16}$$

If the agents cannot agree on anything unambiguously, then Z is a constant. Because of the lattice form of the space of variables [28], there exists a common variable $Z_* = f_*(X) = g_*(Y)$ that is *maximal* in the sense that any other cv Z admits a function h such that $Z = h(Z_*)$. Therefore, it turns out that no cv can take more distinct values than the maximal one. We formally define the notions of common variable and maximal common variable as follows.

Definition 6 (Common Variables [28, 29])**.** *A uv Z is said to be a* common variable (cv) *for X and Y if there exist functions f and g such that $Z = f(X) = g(Y)$.*
Furthermore, a cv is called maximal *if any other cv Z' admits a function h such that $Z' = h(Z)$.*

Remark 5. *As previously discussed, no cv can take more distinct values than the maximal one. The concept of a maximal common variable was first presented by Shannon in the framework of random variables [28] to which he referred by the term* "common information element".

Remark 6. *The notion of* "common information" *has been widely studied in information theory and encapsulates different variants which are generally non-equivalent. For instance, the Gács-Körner (G-K) common information introduced in [30] captures the amount of common randomness between two rv's X and Y that can be extracted by observing them separately. The G-K common information corresponds to the log-cardinality of Shannon's common information element [28]. Another example is the Wyner's common information [31] which, on the other hand, reflects the minimum amount of common randomness required to generate the rv's X and Y separately. More recently, the α-Rényi variant of common information, with $\alpha \in [0,\infty)$, was introduced in [32, 33], where more general families of divergence are used instead of the Kullback-Leibler divergence.*

The nonstochastic information $I_*[X;Y]$ is precisely the log-cardinality of the range of a maximal cv between X and Y. This is because it can be shown that for all pairs of points $(x,y) \in [\![X,Y]\!]$, the

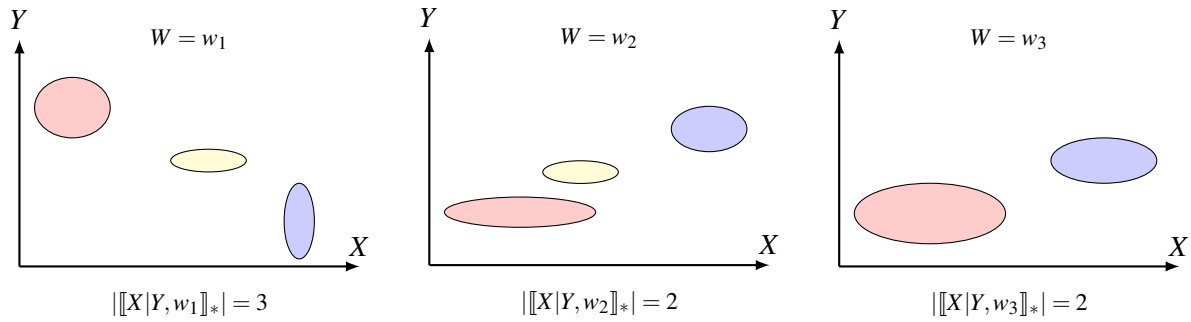


Figure 2.3: Illustrative example of a family of conditional ranges for three realizations w_1, w_2 and w_3.

partition set in $[\![X|Y]\!]_*$ that contains x also uniquely specifies the set in $[\![Y|X]\!]_*$ that contains y. Thus, these overlap partitions define a cv for X and Y, with corresponding functions f and g given by labelling. Furthermore, this cv can be proved to be maximal. See [2] for further details.

2.1.5 Nonstochastic Conditional Information

In classical information theory, *conditional mutual information* $I(X;Y|W)$ measures how much information two rv's X and Y share on average given a third rv W that is available to both [1]. To treat scenarios where joint probability distributions cannot be obtained but where ranges are known with high confidence, we present nonstochastic version of conditional information.

Definition 7 (Conditional Nonstochastic Information [2])**.** *The* conditional nonstochastic information *between X and Y given W is*

$$I_*[X;Y|W] := \min_{w \in [\![W]\!]} \log_2 |[\![X|Y,w]\!]_*|\,, \tag{2.1.17}$$

where for a given $w \in [\![W]\!]$, $[\![X|Y,w]\!]_*$ is the overlap partition of $[\![X|w]\!]$ induced by the family $[\![X|Y,w]\!]$ of conditional ranges $[\![X|y,w]\!]$, $y \in [\![Y|w]\!]$.

> **Example:** Fig. 2.3 shows an example of a family $[\![X,Y|W]\!]$ with $W = \{w_1, w_2, w_3\}$. The nonstochastic conditional information in this case is
>
> $$I_*[X;Y|W] := \log_2(2) = 1 \text{ bit}.$$

Remark 7. *It can be shown that $I_*[X;Y|W]$ also has an important interpretation in terms of cv's: it is the maximum log-cardinality of the ranges of all cv's $Z = f(X,W) = g(Y,W)$ that are unrelated with W. For more details we refer readers to [2].*

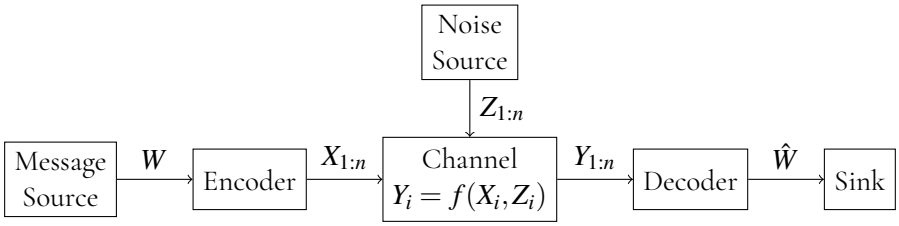


Figure 2.4: Illustration of a communication system transmitting over a memoryless channel defined by a fixed function $f(\cdot)$.

Similar to its stochastic analogue, nonstochastic conditional information has the following properties:

- **Nonnegativity:** $I_*[X;Y|W] \geq 0$.
- **Symmetry:** $I_*[X;Y|W] = I_*[Y;X|W]$.
- **Monotonicity:** $I_*[X;Y|W] \leq I_*[X;Y,Z|W]$.
- **Data Processing Inequality:** If $X \leftrightarrow Y \leftrightarrow Z|W$ is any conditional Markov uncertainty chain given W, then $I_*[W;Z|W] \leq I_*[X;Y|W]$.

2.2 Discrete Memoryless Channels (DMCs)

A basic and widely used model is the so-called *discrete memoryless channel* shown in Fig. 2.4. We distinguish between five types of variables: a source message W, channel inputs X_i, channel outputs Y_i, noise term Z_i, with time index $i \in [1:n]$, and a message estimate $\hat{W}$.
As depicted in Fig. 2.4, the *message source* generates the message $W \in \mathscr{W} := [1 : |\mathscr{W}|]$ such that $|\mathscr{W}| \in \mathbb{Z}_{\geq 1}$. An *encoder* is then employed to map the message w to a string $X_{1:n}$ of n symbols in $\mathscr{X}$. We call the sequence $X_{1:n}$ a *codeword*, and the set of all possible codewords as w ranges over $\mathscr{W}$ is called a *codebook*. The set $\mathscr{X}$ of single letters X_i, whose elements are used to form the codewords, is called an *alphabet*. The channel output is given by

$$Y_i = f(X_i, Z_i) \in \mathscr{Y}, \quad \forall i \in [1:n], \tag{2.2.1}$$

where $f(\cdot)$ is a fixed function, $Z_i \in \mathscr{Z}$ denotes the noise term, and $\mathscr{Y}$ is the output alphabet. The channel is said to be *memoryless*, if

$$p(Y_i|X_{1:i}, Y_{1:i-1}) = p(Y_i|X_i), \quad \forall i \in [1:n], \tag{2.2.2}$$

in a stochastic setting, or if

$$[\![Y_i|X_{1:i}, Y_{1:i-1}]\!] = [\![Y_i|X_i]\!], \quad \forall i \in [1:n], \tag{2.2.3}$$

in a nonstochastic setting. Furthermore, the channel is qualified as *discrete*, when the alphabets $\mathscr{X}$ and $\mathscr{Y}$ are both discrete and finite sets. Thus, we call this a *discrete memoryless channel* (DMC).

At the receiver end, the *decoder* uses the channel output $Y_{1:n} = y_{1:n}$ to provide an estimate $\hat{W} = \hat{w}$ of the original message $W = w$.

2.3 Multiple Access Channels (MACs)

In many modern practical applications, information is sent by two or more users over the same communication medium to a common receiver. Eventually, all messages transmitted by any sender should be reliably recovered by the receiver. Thus, in addition to channel noise, the decoder must also take into account the interference between different users. Some real-world examples of this communication scenario include the uplink of a cellular system and wireless industrial networks.

A channel model capturing the essence of this problem is the aforementioned MAC. A natural example of such channels is the Gaussian MAC

$$Y = X^{(1)} + X^{(2)} + Z, \tag{2.3.1}$$

where $X^{(1)}$ and $X^{(2)}$ are the inputs to the channel, Z denotes zero-mean Gaussian noise independent of $X^{(1)}$ and $X^{(2)}$, and Y is the channel output. Note that $X^{(1)}$, $X^{(2)}$ and Y assume any real numbers. In [34], it was shown that a reasonably good approximation of the Gaussian MAC is a deterministic 2-user XOR MAC model carrying summation in $\mathbb{Z}_2$, i.e., a Galois field of order two. A further example of MACs is the *binary adder channel* (BAC), which unlike the XOR model, performs the addition over $\mathbb{Z}$ and will be discussed in greater detail in Chapter 4.

In communication theory, *channel capacity* is a notion that captures the maximum rate at which information can be transferred reliably in some suitable sense across the channel [35]. Unlike point-to-point channels, whose channel capacity C is a non-negative scalar, in the context of multi-user channels, including the MAC in particular, we talk about a *channel capacity region* $\mathscr{C}$ that lies in an M-dimensional space, where M is the number of users in the system. The ordinary capacity region $\mathscr{C}$ of MACs has been extensively studied in the literature, and was first found by [19, 20]. Subsequently, by means of superposition coding, the single-letter characterization of this region was obtained by Slepian and Wolf in [24]. For more details regarding this topic we refer the reader to [36].



Figure 2.5: Channel transition diagram of **(a)** Binary Erasure Channel with $C_0 = 0$; **(b)** Pentagon Channel with $C_0 > 0$.

2.4 Notion of Zero-Error Channel Capacity

In the context of worst-case state estimation, it turns out that the notion of *zero-error capacity* C_0 is a more insightful figure of merit than classical channel capacity C [2, 3]. The zero-error capacity C_0 of a point-to-point channel is defined as the highest block-coding rate which yields exactly zero decoding errors at the receiver [37]. Intuitively, the block codes with the zero-error property are those leading to an output set that consists of distinct elements, i.e., no codewords result in the same output.

Binary Erasure Channel. A well-known example of DMCs is the *binary erasure channel* (BEC) with a binary input alphabet $\mathscr{X} = \{0, 1\}$ and a ternary output alphabet $\mathscr{Y} = \{0, \varepsilon, 1\}$ as depicted in Fig. 2.5(a). The probability of correctly transmitting a message $x \in \mathscr{X}$ over the BEC is $1 - p$. If the transmitted message is lost, then the channel outputs the letter ε. This event occurs with a probability $p \in (0, 1)$. It is clear here that the input letters are indistinguishable, i.e., the corresponding output sets have a common element, as $P(\varepsilon|x) > 0$, $\forall x \in \mathscr{X}$. And indeed it can be shown, from a result of [37], that the BEC's zero-error capacity $C_0 = 0$.

Pentagon Channel. Another example of a DMC, but with strictly positive zero-error capacity, is the *pentagon channel* [37] shown in Fig. 2.5(b). Each letter of the input alphabet (also coinciding with the output alphabet) $\mathscr{X} = \mathscr{Y} = [0 : 4]$ can be mapped either to itself, i.e., $y = x$, or to the next letter, i.e., $y = x + 1 \pmod 5$. In 1979, Lovàsz [38] showed that $C_0 = (\log 5)/2$. An interesting fact here is that it is possible to achieve the zero-error capacity of the pentagon channel with a code of a block-length equal to 2. An example thereof is the following codebook $X = \{00, 12, 24, 31, 43\}$. It is possible to manually check that all of these symbols are distinguishable, i.e., do not yield overlapping output symbols.

It turns out that the zero-error capacity C_0 does not depend on the probabilities of different transitions in the channel, and hence it can be defined without referring to a stochastic framework. It

was shown in [3] that C_0 of a point-to-point channel can be expressed in terms of nonstochastic information as follows

$$C_0 = \lim_{n\to\infty} \sup_{X:[\![X]\!]\subseteq\mathscr{X}} \frac{I_*[X_{1:n};Y_{1:n}]}{n}, \tag{2.4.1}$$

where the uv's X and Y are the channel input and output, respectively.
In a similar manner to $\mathscr{C}$, in a multi-user communication setup, the *zero-error capacity region* $\mathscr{C}_0$ is characterized by an M-dimensional set when dealing with M senders rather than a single scalar like C_0. A formal definition of $\mathscr{C}_0$ is presented in Chapter 3 of this thesis.

"… when you have eliminated the impossible, whatever remains, however improbable, must be the truth."

— Sir Arthur Conan Doyle (1859–1930)

3

Zero-Error Capacity Region of Multiple Access Channels

The problem of characterizing the zero-error capacity region $\mathscr{C}_0$ for multiple access channels (MACs) has remained open for over three decades. Motivated by this challenging question, we use tools from nonstochastic information theory reviewed in Section 2.1 earlier in this thesis to derive a channel coding theorem in a zero-error communication system for three different classes of MACs. In the first section, we start by studying the case of a two-user MAC with a common message. Then, we extend this model to a system with $M \geq 2$ users and one common message. Finally, a more general channel model that encompasses pairwise inter-user correlation in addition to the common message is investigated. Unlike previous contributions, this analysis does not assume that the block-length is asymptotically large, and hence, the obtained results are applicable for finite coding lengths.

3.1 Zero-Error Communication over Two-User MAC with Common Message

We consider in this section a communication system operating over a two-user MAC. After describing the system model, we formally introduce the notion of the zero-error capacity region of the considered channel, and prove a multi-letter characterization using nonstochastic information theory.

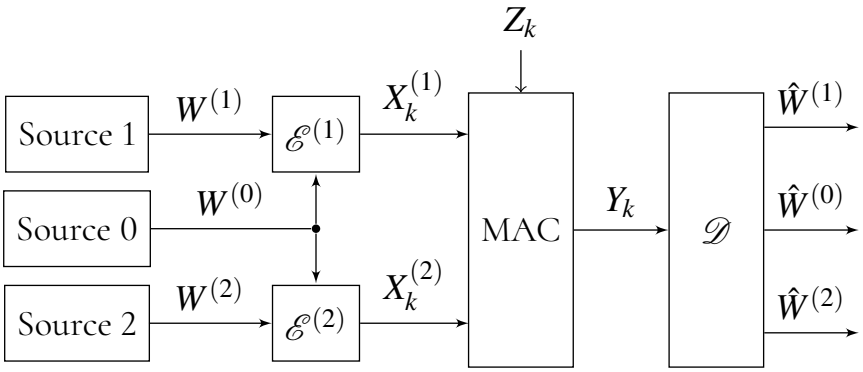


Figure 3.1: The two-transmitter MAC system with a common message operating at time instant k.

3.1.1 System Model

Consider a MAC with one receiver, two transmitters, and three sources, as illustrated in Fig. 3.1. Assume the messages $W^{(0)}$, $W^{(1)}$ and $W^{(2)}$ are mutually unrelated and finite-valued. Without loss of generality, for $i = \{0,1,2\}$, let $W^{(i)}$ take the integer values $[1 : \omega^{(i)}] =: \mathscr{W}^{(i)}$ for some integer $\omega^{(i)} \geq 1$. For a given coding block-length $n \geq 1$, the messages are encoded into channel input sequences $X_{1:n}^{(1)}$ and $X_{1:n}^{(2)}$ as

$$X_{1:n}^{(j)} = \mathscr{E}^{(j)}(W^{(0)}, W^{(j)}), \quad j \in \{1,2\}, \tag{3.1.1}$$

where $\mathscr{E}^{(1)}$ and $\mathscr{E}^{(2)}$ are the coding functions at each transmitter. Note that the *common message* $W^{(0)}$ is seen by both transmitters, while the *private messages* $W^{(1)}$ and $W^{(2)}$ are available only to the respective encoders. The *code rate* for each message is defined as

$$R^{(i)} := \left(\log_2 |\mathscr{W}^{(i)}|\right)/n, \quad i \in \{0,1,2\}. \tag{3.1.2}$$

Due to the common message, the two channel input sequences will typically be related. In the case where the common message can take only one value, so that $R^{(0)} = 0$ bits, each channel input is generated in isolation and is mutually unrelated with the other. At the other extreme, if each of the private messages can take only one value so that $R^{(1)} = R^{(2)} = 0$ bits, then the channel inputs are generated in complete cooperation.

At this point, we define a nonstochastic version of the conventional notion of *stationary memoryless two-user MAC* as follows:

Definition 8 (Stationary Memoryless Uncertain Two-User MAC)**.** *Consider the product input space $\mathscr{X} = \mathscr{X}^{(1)} \times \mathscr{X}^{(2)}$, the noise space $\mathscr{Z}$, an output space $\mathscr{Y}$, and a fixed function $f : \mathscr{X} \times \mathscr{Z} \to \mathscr{Y}$. At*

each time instant k*, the output* $Y_k \in \mathscr{Y}$ *of the* stationary memoryless uncertain two-user MAC *is*

$$Y_k = f(X_k^{(1)}, X_k^{(2)}, Z_k) \in \mathscr{Y}, \quad \forall k \in \mathbb{Z}_{\geq 1}. \tag{3.1.3}$$

The encoded data sequences are sent over the stationary memoryless MAC (Def. 8) as depicted in Fig. 3.1. The encoder $\mathscr{E}^{(1)}$ maps the realized pair $(w^{(0)}, w^{(1)})$ into the codeword $x_{1:n}^{(1)}(w^{(0)}, w^{(1)}) \in [\![X_{1:n}^{(1)}]\!]$, whilst encoder $\mathscr{E}^{(2)}$ assigns $(w^{(0)}, w^{(2)})$ to $x_{1:n}^{(2)}(w^{(0)}, w^{(2)}) \in [\![X_{1:n}^{(2)}]\!]$. Let $\mathscr{X}^{(j)}$ denote the finite range $[\![X^{(j)}]\!]$ for $j \in \{1,2\}$. Note that $[\![X^{(j)}]\!]$ is the set of single letters, unlike $[\![X_{1:n}^{(j)}]\!]$ which consists of n-long codewords that form the deployed codebook. Furthermore, it is assumed that current channel noise Z_k is mutually unrelated with the messages $W^{(0)}, W^{(1)}$ and $W^{(2)}$, as well as the past noise sequence $Z_{1:k-1}$, and has constant range $[\![Z]\!] =: \mathscr{Z}$.
At the receiver, the decoder $\mathscr{D}$ produces message estimates $\hat{W}^{(0)}$, $\hat{W}^{(1)}$ and $\hat{W}^{(2)}$ from the channel output sequence $Y_{1:n}$. Under a zero-error objective, these estimates must always be *exactly* equal to the original messages, regardless of channel noise or interference between $X_k^{(1)}$ and $X_k^{(2)}$. In other words, the conditional range $[\![W^{(i)}|y_{1:n}]\!]$ contains exactly one element for each $i \in \{0,1,2\}$ and any realization $y_{1:n} \in [\![Y_{1:n}]\!]$. We formally define the notion of an $(\lceil 2^{nR^{(0)}} \rceil, \lceil 2^{nR^{(1)}} \rceil, \lceil 2^{nR^{(2)}} \rceil, n)$ zero-error code as follows:

Definition 9 ($(\lceil 2^{nR^{(0)}} \rceil, \lceil 2^{nR^{(1)}} \rceil, \lceil 2^{nR^{(2)}} \rceil, n)$ Zero-Error Code)**.** *Let the block-length* n *be a strictly positive integer, and* $R^{(i)} \geq 0$*, for* $i \in \{0,1,2\}$*. Consider the message sets* $\mathscr{W}^{(0)}$*,* $\mathscr{W}^{(1)}$ *and* $\mathscr{W}^{(2)}$ *with respective sizes* $\lceil 2^{nR^{(0)}} \rceil$*,* $\lceil 2^{nR^{(1)}} \rceil$ *and* $\lceil 2^{nR^{(2)}} \rceil$*. Furthermore, let the function* $\mathscr{E}^{(j)} : \mathscr{W}^{(0)} \times \mathscr{W}^{(j)} \to \mathscr{X}^{(j)}$*, with* $j \in \{1,2\}$*. We say that the encoder pair* $(\mathscr{E}^{(1)}, \mathscr{E}^{(2)})$ *forms a* $(\lceil 2^{nR^{(0)}} \rceil, \lceil 2^{nR^{(1)}} \rceil, \lceil 2^{nR^{(2)}} \rceil, n)$ *zero-error code for the two-user MAC (Def. 8) if* $[\![W^{(0)}, W^{(1)}, W^{(2)}|y_{1:n}]\!]$ *is a singleton for any* $y_{1:n} \in [\![Y_{1:n}]\!]$*.*

For a given code block-length n, the *operational zero-error* n*-capacity*, denoted as $\mathscr{C}_{0,n}$, is operationally defined as the set of rate tuples $(R^{(0)}, R^{(1)}, R^{(2)})$ for which a zero-error communication via the considered channel can be achieved by an appropriate construction of encoders and a decoder. Note that this allows us to characterize the zero-error capacity region at finite code block-lengths, not just as $n \to \infty$. Additionally, if it is allowed to make the code block-length n asymptotically large, then the *zero-error capacity region* $\mathscr{C}_0$ of the channel is given by the closed union

$$\mathscr{C}_0 = \overline{\bigcup_{n \geq 1} \mathscr{C}_{0,n}}. \tag{3.1.4}$$

The notion of *achievable rates* is defined as follows:

Definition 10 (Achievable Rates)**.** *A rate triple* $R = (R^{(0)}, R^{(1)}, R^{(2)})$ *is called* achievable *if there exists a sequence of* $(\lceil 2^{nR_n^{(0)}} \rceil, \lceil 2^{nR_n^{(1)}} \rceil, \lceil 2^{nR_n^{(2)}} \rceil, n)$ *zero-error codes, with* $R_n = (R_n^{(0)}, R_n^{(1)}, R_n^{(2)})$*, approaching* R*.*

Note that the system setup considered in this section is inspired by that of [24]. The critical difference, however, is that the messages and channel here are not assumed to have any statistical structure and the aim is to recover the messages perfectly, not just with an arbitrarily small error probability.

3.1.2 Convexity of the Zero-Error Capacity Region of Two-User MAC

Our aim now is to prove the convexity of the region defined in (3.1.4). To this end, we use the *time-sharing* argument. Roughly speaking, the idea of time sharing is based on the concept of rate splitting. For instance, we divided each sequence of bits to j blocks of length n and k blocks of length m. Each type of blocks is then transmitted at a different rate, i.e., R_n for the first type and R_m for the second. Then, the total transmission rate represents a convex combination of R_n and R_m. Note that the well-known concepts of time and frequency division are special cases of time sharing. For more details on the time-sharing argument in literature, we refer the reader to, e.g., [35, 36]. A rigorous proof is presented next.

Theorem 1. *The zero-error capacity region $\mathscr{C}_0$ (3.1.4) is convex.*

Proof. Select two achievable rate triples $R^{'}$, $R^{''} \in \mathscr{C}_0$ (3.1.4). For any $\varepsilon > 0$, $\exists (n, R_n^{'})$ and $(m, R_m^{''})$ such that

$$\|R^{'} - R_n^{'}\| \leq \varepsilon, \tag{3.1.5a}$$
$$\|R^{''} - R_m^{''}\| \leq \varepsilon, \tag{3.1.5b}$$

where $R_n^{'} \in \mathscr{C}_{0,n}$ and $R_m^{''} \in \mathscr{C}_{0,m}$ for sufficiently large $n, m \in \mathbb{Z}_{\geq 1}$ that denote the block-lengths of the zero-error codes operating at $R_n^{'}$ and $R_m^{''}$. First, we show that for any $\alpha \in (0,1)$, we can construct an achievable rate triple

$$R = \alpha R^{'} + (1-\alpha) R^{''}. \tag{3.1.6}$$

We define the terms $e_n^{'}$ and $e_m^{''}$ as follows

$$e_n^{'} := R^{'} - R_n^{'}, \tag{3.1.7a}$$
$$e_m^{''} := R^{''} - R_m^{''}. \tag{3.1.7b}$$

By transmitting j blocks of length n at rate $R_n^{'}$, followed by the remaining k blocks of length m at rate $R_m^{''}$, we obtain

$$\begin{aligned} \bar{R}_{n,m} &= \frac{jn}{jn+km} R_n^{'} + \frac{km}{jn+km} R_m^{''} \\ &= \frac{jn}{jn+km} \left(R^{'} - e_n^{'} \right) + \frac{km}{jn+km} \left(R^{''} - e_m^{''} \right). \end{aligned} \tag{3.1.8}$$

For sufficiently small $\delta > 0$ and given integers $n, m \geq 1$, we seek integers j, k such that

$$\left| \alpha - \frac{jn}{jn+km} \right| \leq \delta. \tag{3.1.9}$$

Thus, for a fixed $\alpha \in (0,1)$, the ratio $(k/j) \in \mathbb{Q}$ must satisfy

$$\frac{n}{m}\left[\frac{1}{\alpha+\delta} - 1\right] \leq \frac{k}{j} \leq \frac{n}{m}\left[\frac{1}{\alpha-\delta} - 1\right]. \tag{3.1.10}$$

For arbitrarily small δ and $\alpha \in (0,1)$, there exist $k, j \in \mathbb{Z}_{\geq 1}$ such that (3.1.10) holds. Then, we have

$$\begin{aligned}
\|\bar{R}_{n,m} - R\| &= \left\| \frac{jn}{jn+km}\left(R' - e_n'\right) + \frac{km}{jn+km}\left(R'' - e_m''\right) - \alpha R' - (1-\alpha)R'' \right\| \\
&= \left\| \left(\frac{jn}{jn+km} - \alpha\right) R' + \left(\frac{km}{jn+km} - (1-\alpha)\right) R'' - \frac{jn}{jn+km} e_n' - \frac{km}{jn+km} e_m'' \right\| \\
&\leq \left| \frac{jn}{jn+km} - \alpha \right| \|R'\| + \left| -\frac{jn}{jn+km} + \alpha \right| \|R''\| + \frac{jn}{jn+km}\|e_n'\| + \frac{km}{jn+km}\|e_m''\| \qquad (3.1.11) \\
&\leq \delta(\|R'\| + \|R''\|) + \left(\frac{jn}{jn+km} + \frac{km}{jn+km}\right)\varepsilon \qquad (3.1.12) \\
&= \delta(\|R'\| + \|R''\|) + \varepsilon,
\end{aligned}$$

where (3.1.11) follows from the triangle inequality, and (3.1.12) holds by virtue of (3.1.9), (3.1.5a) and (3.1.5b). Hence, by making the block-lengths n and m sufficiently large, and ε, δ arbitrarily small, the difference $\|\bar{R}_{n,m} - R\|$ can be made arbitrarily small, i.e., the rate R is arbitrarily close to the zero-error code rate $\bar{R}_{n,m}$. We then deduce that $R \in \mathscr{C}_0$, which proves the desired property. □

3.1.3 Non-Emptiness of the Interior of the Zero-Error Capacity Region of Two-User MAC

Another property of the zero-error capacity region $\mathscr{C}_0$ of the two-user MAC is studied here, namely the required conditions to guarantee the non-emptiness of its interior region denoted by $\mathrm{int}(\mathscr{C}_0)$.

Lemma 1. *Consider the two-input, one-output MAC (Def. 8), with channel inputs $X^{(1)}$, $X^{(2)}$ and output Y, as well as its zero-error capacity region $\mathscr{C}_0$ (3.1.4). The interior region of $\mathscr{C}_0$ is non-empty, i.e., $\mathrm{int}(\mathscr{C}_0) \neq \emptyset$ if the following requirements hold:*

(i) There exists $x^{(1)} \in \mathscr{X}^{(1)}$, for which there are at least two inputs $x_1^{(2)}, x_2^{(2)} \in \mathscr{X}^{(2)}$ that are non-

confusable, i.e.,

$$[\![Y|x^{(1)},x_1^{(2)}]\!]\cap[\![Y|x^{(1)},x_2^{(2)}]\!]=\emptyset. \tag{3.1.13}$$

(ii) There exists $x^{(2)}\in\mathscr{X}^{(2)}$, for which there are at least two codewords $x_1^{(1)},x_2^{(1)}\in\mathscr{X}^{(1)}$ that are non-confusable, i.e.,

$$[\![Y|x^{(2)},x_1^{(1)}]\!]\cap[\![Y|x^{(2)},x_2^{(1)}]\!]=\emptyset. \tag{3.1.14}$$

Proof. The channel output $Y_k\in[\![Y]\!]$ at time $k\in\mathbb{Z}_{\geq 0}$ can be given in terms of a fixed function f as follows

$$Y_k=f(X_k^{(1)},X_k^{(2)},Z_k),\ \forall k\in\mathbb{Z}_{\geq 0}. \tag{3.1.15}$$

Firstly, it is clear that the trivial rate tuple $(R^{(0)},R^{(1)},R^{(2)})=(0,0,0)\in\mathscr{C}_0$. In fact, given any three distinct messages $W^{(0)}=w^{(0)}$, $W^{(1)}=w^{(1)}$ and $W^{(2)}=w^{(2)}$, the decoder will always be able to errorlessly generate the message estimates, as each source is always generating the same message. Let $\partial\mathscr{C}_0$ denote the boundary of $\mathscr{C}_0$. We now establish an inner bound $\underline{\partial\mathscr{C}_0}$ on the region $\mathscr{C}_0$.
To this end, we proceed as follows. Allow only one source, namely Source i, to operate at a rate $R^{(i)}>0$, whilst the remaining sources are kept silent, i.e., the other two messages are fixed such that $W^{(j)}=w^{(j)}$, with $j\neq i$, and hence, the corresponding rates are $R^{(j)}=0$ bits.
We now show that there exists a codebook with strictly positive rate for some $n\geq 1$. For the sake of this proof, we keep the messages $W^{(0)}$ and $W^{(1)}$ fixed, and set the codeword $X_k^{(1)}=x^{(1)}$. Hence, we write

$$\begin{aligned}Y_k&=f(x^{(1)},X_k^{(2)},Z_k)\\&\equiv\tilde{f}(X_k^{(2)},Z_k),\ \forall k\in\mathbb{Z}_{\geq 0}.\end{aligned} \tag{3.1.16}$$

In other words, for a given codeword $x^{(1)}$, the channel is now equivalent to a point-to-point (p2p) channel denoted by $\tilde{f}$. As shown in [37], a p2p channel has a strictly positive zero-error capacity C_0[1], if there exists at least two codewords $x_1^{(2)},x_2^{(2)}\in\mathscr{X}^{(2)}$ that are non-confusable, i.e.,

$$[\![Y|x^{(1)},x_1^{(2)}]\!]\cap[\![Y|x^{(1)},x_2^{(2)}]\!]=\emptyset. \tag{3.1.17}$$

As this is precisely condition (3.1.13), it is possible to transmit information over $\tilde{f}$ with exactly zero error at a strictly positive rate $R^{(2)}\geq 1$ bit/sample. By equivalence of f and $\tilde{f}$ in this setting (3.1.16), the rate tuple $(0,0,R^{(2)})$ thus belongs to the zero-error capacity region of f, i.e., $(0,0,R^{(2)})\in\mathscr{C}_0$.

[1]Note the difference here between C_0, which is a scalar, and the region $\mathscr{C}_0$. For more detail, please see Section 2.4.

In the same manner, we can show the existence of $R^{(1)} > 0$ such that $(0, R^{(1)}, 0) \in \mathscr{C}_0$ by exploiting (3.1.14).
As to the rate $R^{(0)}$ at which the common message $W^{(0)}$ is transmitted, we construct a slightly different scheme. In addition to fixing the message $W^{(1)} = w^{(1)}$, we also set the output of the second encoder $\mathscr{E}^{(2)}$ to $x^{(2)}$ regardless of the message $W^{(0)}$ that is known also to the receiver. Therefore, at time $k \in \mathbb{Z}_{\geq 1}$, the channel output Y_k can be written as

$$\begin{aligned} Y_k &= f(X_k^{(1)}, x^{(2)}, Z_k) \\ &\equiv g(X_k^{(1)}, Z_k), \ \forall k \in \mathbb{Z}_{\geq 1}. \end{aligned} \tag{3.1.18}$$

Note that by (3.1.14), there are at least two codewords $x_1^{(1)}, x_2^{(1)} \in \mathscr{X}^{(1)}$ that are non-confusable. Hence, the rate $R^{(0)} > 0$.
By virtue of the convexity of $\mathscr{C}_0$ (Theorem 1), the convex hull formed by $(0,0,0)$, $(R^{(0)},0,0)$, $(0,R^{(1)},0)$ and $(0,0,R^{(2)})$, represents an inner bound on $\mathscr{C}_0$, that we denote by $\underline{\mathscr{C}_0}$. As $\mathbf{int}(\underline{\mathscr{C}_0})$ is clearly non-empty, this completes the proof of Lemma 1. □

3.1.4 Zero-Error Capacity Region of Two-User MAC with Common Message via Nonstochastic Information

Based on the presented definition and inspired by the previous results obtained in classic information theory, we are now in a position to prove the following theorem using nonstochastic information with the aim of characterizing the zero-error capacity region $\mathscr{C}_0$ (3.1.4) for the two-user MAC (Def. 8).

Theorem 2. *For a given block-length $n \geq 1$, let $\mathscr{R}(U, X_{1:n}^{(1)}, X_{1:n}^{(2)})$ be the set of rate tuples $(R^{(0)}, R^{(1)}, R^{(2)})$ such that*

$$nR^{(0)} \leq I_* [U; Y_{1:n}] \tag{3.1.19a}$$

$$nR^{(1)} \leq I_* [X_{1:n}^{(1)}; Y_{1:n} | U] \tag{3.1.19b}$$

$$nR^{(2)} \leq I_* [X_{1:n}^{(2)}; Y_{1:n} | U] \tag{3.1.19c}$$

where $X_{1:n}^{(j)}$, for $j \in \{1,2\}$, are sequences of inputs to the multiple access channel (MAC) (3.1.3), $Y_{1:n}$ is the corresponding channel output sequence, and U is an auxiliary uncertain variable (uv). Then, the zero-error n-capacity region $\mathscr{C}_{0,n}$ *of the MAC over n uses coincides with the union of the regions $\mathscr{R}(U, X_{1:n}^{(1)}, X_{1:n}^{(2)})$ over all uv's $U, X_{1:n}^{(1)}, X_{1:n}^{(2)}$ that satisfy the Markov uncertainty chains $X_{1:n}^{(1)} \leftrightarrow U \leftrightarrow X_{1:n}^{(2)}$ and $U \leftrightarrow (X_{1:n}^{(1)}, X_{1:n}^{(2)}) \leftrightarrow Y_{1:n}$.*

This result is the zero-error analogue of the Slepian-Wolf ordinary capacity region $\mathscr{C}$ [24] in terms of nonstochastic rather than Shannon information. Although $\mathscr{C}$ is prima facie given in 'single-letter'

terms, it is operationally relevant only at large block-lengths n, to yield small probabilities of error. In contrast, the result above specifies all rates tuples that allow exactly zero errors to be achieved at a given finite n. This could potentially be of interest in safety-critical low-latency applications in distributed networked control.

Although (3.1.19a)-(3.1.19c) give a cuboidal rate region $\mathscr{R}(U, X_{1:n}^{(1)}, X_{1:n}^{(2)})$, it is not clear if the zero-error capacity regions also have geometrically simple shapes, due to the unions over U, $X_{1:n}^{(1)}$, $X_{1:n}^{(2)}$ and n.

Proof of Converse

Consider a zero-error code (3.1.1) with block-length n operating at rates $R^{(0)}, R^{(1)}$ and $R^{(2)}$ (3.1.2) over the two-user MAC (Def. 8), and set the auxiliary uv $U := W^{(0)}$. As $W^{(0)}$, $W^{(1)}$ and $W^{(2)}$ are mutually unrelated, it follows from (3.1.1) that the codewords $X_{1:n}^{(1)}$ and $X_{1:n}^{(2)}$ are conditionally unrelated given $W^{(0)}$, i.e., the first Markov uncertainty chain $X_{1:n}^{(1)} \leftrightarrow U \leftrightarrow X_{1:n}^{(2)}$ is satisfied. To prove the validity of the second Markov uncertainty chain $Y_{1:n} \leftrightarrow (X_{1:n}^{(1)}, X_{1:n}^{(2)}) \leftrightarrow U$, consider the following

$$\begin{aligned}
[\![Y|X_{1:n}^{(1)}, X_{1:n}^{(2)}, U]\!] &= [\![Y|X_{1:n}^{(1)}, X_{1:n}^{(2)}, W^{(0)}]\!] && (3.1.20)\\
&\overset{(3.1.1)}{=} [\![Y|\mathscr{E}^{(1)}(W^{(0)}, W^{(1)}), \mathscr{E}^{(2)}(W^{(0)}, W^{(2)}), W^{(0)}]\!] \\
&= [\![Y|\mathscr{E}^{(1)}(W^{(0)}, W^{(1)}), \mathscr{E}^{(2)}(W^{(0)}, W^{(2)})]\!] && (3.1.21)\\
&= [\![Y|X_{1:n}^{(1)}, X_{1:n}^{(2)}]\!], && (3.1.22)
\end{aligned}$$

where (3.1.20) follows from the fact that $U := W^{(0)}$ and (3.1.21) is valid because the encoding functions $\mathscr{E}^{(1)}$ and $\mathscr{E}^{(2)}$ are bijective in a zero-error communication setting. Since the channel noise in (8) is unrelated with the messages, and subsequently with the codewords, we also have the second Markov uncertainty chain $Y_{1:n} \leftrightarrow (X_{1:n}^{(1)}, X_{1:n}^{(2)}) \leftrightarrow U$. As the messages are all errorlessly recovered at the receiver, there exists a decoding function $\mathscr{D}^{(0)}$ such that

$$W^{(0)} = \mathscr{D}^{(0)}(Y_{1:n}). \tag{3.1.23}$$

Since $U = W^{(0)}$, we see that $W^{(0)}$ is therefore a common variable (cv) between U and $Y_{1:n}$. By the maximal cv property of I_* (see Section 2.1.4), we have

$$\begin{aligned}
nR^{(0)} &= \log_2 |[\![W^{(0)}]\!]| \\
&\leq I_*[U; Y_{1:n}], && (3.1.24)
\end{aligned}$$

proving (3.1.19a).

We next prove the remaining two inequalities (3.1.19b) and (3.1.19c). Observe that for a given realization $w^{(0)}$ of the common message, there must be a unique message $w^{(1)}$ corresponding to each

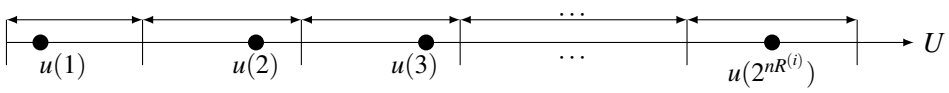


Figure 3.2: Illustrative example of the unique overlap partition, namely the family $[\![U|Y_{1:n}]\!]_*$. The horizontal lines represent the different member sets of the partition and the black filled-in circles correspond to the selected points $u(w^{(i)})$ for $i \in \{0,1,2\}$.

channel codeword $x_{1:n}^{(1)}$; otherwise, multiple values of $w^{(1)}$ would be associated with a single channel output sequence $y_{1:n}$, violating the zero-error requirement. Consequently, there must exist a mapping g such that $W^{(1)} = g(X_{1:n}^{(1)}, W^{(0)})$. Furthermore, by the zero-error property there also exists a function $\mathscr{D}^{(1)}$ such that $W^{(1)} = \mathscr{D}^{(1)}(Y_{1:n})$. Thus, $W^{(1)}$ is a cv between $(X_{1:n}^{(1)}, W^{(0)})$ and $(Y_{1:n}, W^{(0)})$. As by assumption $W^{(1)}$ is also unrelated with $U = W^{(0)}$, the interpretation of conditional I_* in terms of maximal unrelated cv's (see Remark 7) allows us to conclude that

$$\begin{aligned} nR^{(1)} &= \log_2 |[\![W^{(1)}]\!]| \\ &\leq I_*[X_{1:n}^{(1)}; Y_{1:n}|W^{(0)}] \\ &= I_*[X_{1:n}^{(1)}; Y_{1:n}|U], \end{aligned} \tag{3.1.25}$$

proving (3.1.19b). In a similar way, the bound on the rate $R^{(2)}$ stated in (3.1.19c) can be shown.

Proof of Achievability

We now prove that if we have a block-length n and uv's $U, X_{1:n}^{(1)}$ and $X_{1:n}^{(2)}$ satisfying the requirements in Theorem 2, it is possible to construct a zero-error coding scheme at rates achieving the equalities in (3.1.19a)-(3.1.19c).

Codebook Generation. Wlog set $nR^{(0)} = I_*[U; Y_{1:n}]$ and pick one point in each of the disjoint sets of the overlap partition $[\![U|Y_{1:n}]\!]_*$. With mild abuse of notation call these distinct points $u(w^{(0)})$, where $w^{(0)} \in [1 : 2^{nR^{(0)}}]$. Note that for a particular $w^{(0)} \in [1 : 2^{nR^{(0)}}]$, $u(w^{(0)})$ is a realization of the uv U. By also setting $nR^{(j)} = I_*[X_{1:n}^{(j)}; Y_{1:n}|U]$ for $j \in \{1,2\}$, (2.1.17) implies that

$$2^{nR^{(j)}} \leq \left| \left[\!\!\left[X_{1:n}^{(j)} | Y_{1:n}, U = u(w^{(0)}) \right]\!\!\right]_* \right|, \quad \forall w^{(0)} \in [1 : 2^{nR^{(0)}}]. \tag{3.1.26}$$

For any $w^{(0)}$, we may therefore pick $2^{nR^{(j)}}$ distinct codewords $x_{1:n}^{(j)}$ from $[\![X_{1:n}^{(i)}|U = u(w^{(0)})]\!]$ such that there is at most one codeword in each set of the overlap partition $[\![X_{1:n}^{(j)}|Y_{1:n}, U = u(w^{(0)})]\!]_*$. The obtained codewords are then denoted by $\mathscr{E}^{(j)}(w^{(0)}, w^{(j)})$, $w^{(j)} \in [1 : 2^{nR^{(j)}}]$. This gives us the coding laws in (3.1.1).

Zero-Error Decoding. To show that this code may be decoded with exactly zero errors, observe first that since $X_{1:n}^{(1)} \leftrightarrow U \leftrightarrow X_{1:n}^{(2)}$, the joint conditional range $[\![X_{1:n}^{(1)}, X_{1:n}^{(2)} | U = u(w^{(0)})]\!]$ is just the Cartesian product of the individual conditional ranges, i.e.,

$$\left[\!\left[X_{1:n}^{(1)} | U = u(w^{(0)})\right]\!\right] \times \left[\!\left[X_{1:n}^{(2)} | U = u(w^{(0)})\right]\!\right]. \tag{3.1.27}$$

Thus we are guaranteed that for every $w^{(0)}$, all codeword pairs $(\mathscr{E}^{(1)}(w^{(0)}, w^{(1)}), \mathscr{E}^{(2)}(w^{(0)}, w^{(2)}))$ with $w^{(j)} \in [1 : 2^{nR^{(j)}}]$ and $j \in \{1, 2\}$, lie within the conditional joint range $[\![X_{1:n}^{(1)}, X_{1:n}^{(2)} | U = u(w^{(0)})]\!]$. In other words, for every combination of $w^{(0)}, w^{(1)}$ and $w^{(2)}$, the triple $(\mathscr{E}^{(1)}(w^{(0)}, w^{(1)}), \mathscr{E}^{(2)}(w^{(0)}, w^{(2)}), u(w^{(0)}))$ is a valid point inside the joint range $[\![X_{1:n}^{(1)}, X_{1:n}^{(2)}, U = u(w^{(0)})]\!]$.
The decoding proceeds in three stages:

(a) In the first stage, the common message $w^{(0)}$ is recovered. Recall that each of the $2^{nR^{(0)}}$ points $u(w^{(0)})$ lies in a distinct set of the overlap partition $[\![U | Y_{1:n}]\!]_*$ as shown in Fig. 3.2. By the cv property of overlap partitions, this set is uniquely determined by the corresponding set of the matching overlap partition $[\![Y_{1:n} | U]\!]_*$ that contains the channel output sequence $y_{1:n}$. In this way, $w^{(0)}$ is uniquely decoded.

(b) In the second stage, having obtained $w^{(0)}$, the decoder calculates which distinct set of the conditional overlap partition $[\![Y_{1:n} | X_{1:n}^{(1)}, U = u(w^{(0)})]\!]_*$ contains $y_{1:n}$. Again by the cv property, this set uniquely determines the corresponding set of the matching conditional overlap partition $[\![X_{1:n}^{(1)} | Y_{1:n}, U = u(w^{(0)})]\!]_*$ that contains the codeword $\mathscr{E}^{(1)}(w^{(0)}, w^{(1)})$. By construction, for each $w^{(0)}$, there is at most one codeword in each set of the latter conditional overlap partition; thus $w^{(1)}$ is uniquely recovered.

(c) In the third stage, the decoder repeats Step **(b)** but with $X_{1:n}^{(2)}$ instead of $X_{1:n}^{(1)}$, and recovers $w^{(2)}$ uniquely in the same way.

3.2 Zero-Error Communication over M-User MAC with Common Message

In this section, we consider the M-user MAC with one common message and $M \geq 2$. Similarly to the previous section, we use concepts from nonstochastic information theory to obtain an exact characterization the zero-error capacity region $\mathscr{C}_0$ of this class of MACs.

3.2.1 System Model

Consider the communication setup depicted in Fig. 3.3. The system consists of M transmitters, each wishing to convey a distinct *private message* $W^{(j)}$, where $j \in [1 : M]$, and a *common message*

$W^{(0)}$ to a unique receiver over an M-user MAC. Suppose the messages $W^{(0)}, W^{(1)}, W^{(2)} \cdots, W^{(M)}$ are mutually unrelated and finite-valued. We assume without loss of generality, that, for $i \in [0:M]$, the messages $W^{(i)}$ take the integer values $[1:\omega^{(i)}] =: \mathscr{W}^{(i)}$ for some integer $\omega^{(i)} \geq 1$. For a given block-length $n \geq 1$, the messages are mapped into channel input sequences $X_{1:n}^{(1)}, X_{1:n}^{(2)}, \cdots, X_{1:n}^{(M)}$ as follows

$$X_{1:n}^{(j)} = \mathscr{E}^{(j)}(W^{(0)}, W^{(j)}), \ \ \forall j \in [1:M], \tag{3.2.1}$$

where $\{\mathscr{E}^{(j)}\}_{j=1}^{M}$ are the coding laws at each transmitter. Note that the *common message* $W^{(0)}$ is seen by all encoders, while the *private messages* $W^{(j)}$ are only available to the respective transmitters. The code rate for each message is defined as

$$R^{(i)} := \left(\log_2 |\mathscr{W}^{(i)}|\right)/n, \ \ \forall i \in [0:M]. \tag{3.2.2}$$

This general system configuration, where a common message is seen by all encoders, allows us to incorporate some form of relatedness among the channel input sequences in the model. Similar to the two-user MAC discussed in Section 3.1, in the case where the common message can take only one value, then the rate $R^{(0)} = 0$ and each channel input is generated separately and is mutually unrelated with the other. On the other hand, if the private messages can each take only one value so that $R^{(1)} = R^{(2)} = \cdots = R^{(M)} = 0$, then all of the channel inputs $X_{1:n}^{(1)}, \cdots, X_{1:n}^{(M)}$ are generated in complete cooperation.

At this stage, we define a nonstochastic version of the conventional notion of *stationary memoryless M-user MAC* by extending Def. 8 as follows:

Definition 11 (Stationary Memoryless Uncertain M-User MAC)**.** *Consider the product input space $\mathscr{X} = \mathscr{X}^{(1)} \times \mathscr{X}^{(2)} \times \cdots \times \mathscr{X}^{(M)}$, the noise space $\mathscr{Z}$, an output space $\mathscr{Y}$, and a fixed function $f : \mathscr{X} \times \mathscr{Z} \to \mathscr{Y}$. At each time instant k, the output $y_k \in \mathscr{Y}$ of the* stationary memoryless uncertain two-user MAC *is*

$$Y_k = f(X_k^{(1)}, X_k^{(2)}, \cdots, X_k^{(M)}, Z_k) \in \mathscr{Y}, \quad \forall k \in \mathbb{Z}_{\geq 1}. \tag{3.2.3}$$

The encoded data sequences are then sent over this channel (3.2.3) as depicted in Fig. 3.3. Notice that the channel noise is denoted by Z_k and is mutually unrelated with all messages and past channel noise, i.e., $Z_{1:k-1}, W^{(0)}, W^{(1)}, W^{(2)}, \cdots, W^{(M)}$. We further assume that the range of the channel noise $[\![Z_k]\!] = \mathscr{Z}$ is constant.

The receiver consists of a decoder $\mathscr{D}$ that generates estimates $\hat{W}^{(0)}$, $\hat{W}^{(1)}, \cdots, \hat{W}^{(M)}$ of the transmitted messages using the channel output sequence $Y_{1:n}$. As previously discussed, in the context of zero-error communication, these estimates must always be exactly equal to the original messages, despite the existence of channel noise or inter-user interference. This requirement means

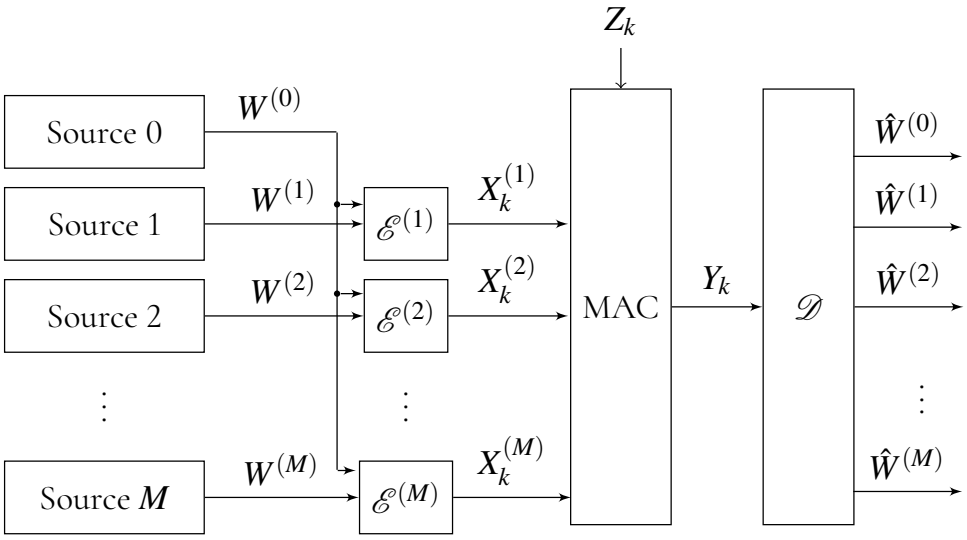


Figure 3.3: The M-user MAC system with a common message $W^{(0)}$ operating at time instant k.

that, for $i \in [0:M]$, the conditional range $[\![W^{(i)}|y_{1:n}]\!]$ consists of one element for any channel output sequence $y_{1:n} \in [\![Y_{1:n}]\!]$. In an analogous manner to Section 3.1, we define the concept of an $(\lceil 2^{nR^{(0)}} \rceil, \lceil 2^{nR^{(1)}} \rceil, \cdots, \lceil 2^{nR^{(M)}} \rceil, n)$ zero-error code as follows:

Definition 12 ($(\lceil 2^{nR^{(0)}} \rceil, \lceil 2^{nR^{(1)}} \rceil, \cdots, \lceil 2^{nR^{(M)}} \rceil, n)$ Zero-Error Code)**.** *Let the block-length n and the number of users M be strictly positive integers, and $R^{(i)} \geq 0$, with $i \in [1:M]$. Consider the message sets $\mathscr{W}^{(0)}, \mathscr{W}^{(1)}, \cdots, \mathscr{W}^{(M)}$ with respective sizes $\lceil 2^{nR^{(0)}} \rceil, \lceil 2^{nR^{(1)}} \rceil, \cdots, \lceil 2^{nR^{(M)}} \rceil$. Furthermore, let the function $\mathscr{E}^{(j)} : \mathscr{W}^{(0)} \times \mathscr{W}^{(j)} \to \mathscr{X}^{(j)}$, for all $j \in [1:M]$. We say that the M encoders $(\mathscr{E}^{(j)})_{j=1}^{M}$ form a $(\lceil 2^{nR^{(0)}} \rceil, \lceil 2^{nR^{(1)}} \rceil, \cdots, \lceil 2^{nR^{(M)}} \rceil, n)$ zero-error code for the M-user MAC (Def. 11) if there exists no two distinct codeword tuples $(a_{1:n}^{(1)}, \cdots, a_{1:n}^{(M)}) \neq (b_{1:n}^{(1)}, \cdots, b_{1:n}^{(M)}) \in \mathscr{X}^{(1)} \times \cdots \times \mathscr{X}^{(M)}$ that can result in the same output sequence $y_{1:n} \in [\![Y_{1:n}]\!]$, i.e.,*

$$\left[\!\left[Y_{1:n}|a_{1:n}^{(1)}, \cdots, a_{1:n}^{(M)}\right]\!\right] \bigcap \left[\!\left[Y_{1:n}|b_{1:n}^{(1)}, \cdots, b_{1:n}^{(M)}\right]\!\right] = \emptyset. \tag{3.2.4}$$

For a given block-length n, we define the *zero-error n-capacity region* $\mathscr{C}_{0,n}$ of the MAC as the set of rate tuples $R = (R^{(i)})_{i\in[0:M]}$ for which this is possible by suitable choice of coding functions. Note that when arbitrarily long block-lengths are permitted, the notion of an *achievable rate* becomes useful.

Definition 13 (Achievable Rate)**.** *We say that a rate tuple $R = (R^{(i)})_{i\in[0:M]}$ is* achievable *if there exists a sequence $(\lceil 2^{nR_n^{(i)}} \rceil_{i\in[0:M]}, n)_{n\in\mathbb{Z}_{\geq 1}}$ of zero-error codes with $R_n = (R_n^{(i)})_{i\in[0:M]}$ approaching R as $n \to \infty$.*

Similarly to Section 3.1, the zero-error n-capacity region $\mathscr{C}_{0,n}$ of the M-user MAC is operationally defined for a given code length n as the set of all rate tuples $(R^{(0)}, R^{(1)}, \cdots, R^{(M)})$ for which it is possible to achieve zero-error communication. Consequently, by allowing n to approach ∞, the zero-

error capacity region $\mathscr{C}_0$ can be expressed as in (3.1.4). Additionally, it is worth noting that this region, namely $\mathscr{C}_0$, is convex. The proof of this property follows the same lines of that of Theorem 1. The only difference here is that instead of defining $R^{'}$ and $R^{''}$ in a three-dimensional space, we consider them to be $M+1$-tuples. The rest of the proof remains the same.

3.2.2 Zero-Error Capacity of M-User MAC with Common Message via Nonstochastic Information

At this stage, we establish a multi-letter characterization of the zero-error n-capacity region for a give code length $n \geq 1$, denoted by $\mathscr{C}_{0,n}$ of the introduced M-user MAC.

Theorem 3. *For a given block-length $n \geq 1$, let $\mathscr{R}(U, X_{1:n}^{(1)}, X_{1:n}^{(2)}, \cdots, X_{1:n}^{(M)})$ be the set of non-negative rate tuples $(R^{(0)}, R^{(1)}, R^{(2)}, \cdots, R^{(M)})$ such that*

$$nR^{(0)} \leq I_*\left[U; Y_{1:n}\right], \tag{3.2.5a}$$

$$nR^{(j)} \leq I_*\left[X_{1:n}^{(j)}; Y_{1:n}|U\right], \quad \forall j \in [1:M], \tag{3.2.5b}$$

where $X_{1:n}^{(j)}$, with $j \in [1:M]$, are sequences of inputs to the M-user multiple access channel (MAC) (3.3.4), $Y_{1:n}$ is the corresponding channel output sequence, and U is an auxiliary uncertain variable (uv).
Then, the zero-error n-capacity region $\mathscr{C}_{0,n}$ *of the M-user MAC over n channel uses coincides with the union of the regions $\mathscr{R}(U, X_{1:n}^{(1)}, X_{1:n}^{(2)}, \cdots, X_{1:n}^{(M)})$ over all uv's $U, X_{1:n}^{(1)}, X_{1:n}^{(2)}, \cdots, X_{1:n}^{(M)}$ that satisfy:*

i) $\prod_{j=1}^{(M)} [\![X_{1:n}^{j}|U]\!] = [\![X_{1:n}^{1}, \cdots, X_{1:n}^{(M)}|U]\!]$, *and*

ii) $U \leftrightarrow \left(X_{1:n}^{(1)}, X_{1:n}^{(2)}, \cdots, X_{1:n}^{(M)}\right) \leftrightarrow Y_{1:n}$ *form a Markov uncertainty chain.*

This result is a generalization of the two-user MAC in terms of nonstochastic rather than the classical Shannon information. The zero-error n-capacity $\mathscr{C}_{0,n}$ now is a region that lies in a $M+1$-dimensional rate space. Like Theorem 2, the obtained characterization here is not only valid for asymptotically large block-lengths n, but it rather includes all rate tuples that guarantee error-free communication at a given finite n. If we are allowed to use arbitrarily long blocks, i.e., $n \to \infty$, then the relevant zero-error capacity region $\mathscr{C}_0$ has the same expression as (3.1.4).

Proof of Converse

Consider the M-user MAC model introduced in Def. 11 and let $(R^{(i)})_{i=0}^{M}$ (3.2.2) be the rates of some zero-error code (3.2.1) with block-length n. Furthermore, we set the uv $U = W^{(0)}$. By assumption, the messages $\{W^{(i)}\}_{i=0}^{M}$ are mutually unrelated and hence from (3.2.1) we conclude that

the codewords satisfy the following

$$\prod_{j=1}^{M} \left[\!\left[X_{1:n}^{(j)} | U \right]\!\right] = \left[\!\left[X_{1:n}^{(1)}, \cdots, X_{1:n}^{(M)} | U \right]\!\right]. \tag{3.2.6}$$

Additionally, the unrelatedness of the channel noise Z with the messages $W^{(i)}$, $\forall i \in [0:M]$, implies that Z is also unrelated with the codewords $\{X^{(j)}\}_{j=1}^{M}$. Thus, the Markov chain $Y_{1:n} \leftrightarrow (X_{1:n}^{(1)}, X_{1:n}^{(2)}, \cdots, X_{1:n}^{(M)}) \leftrightarrow U$ is also satisfied.

As zero-error communication is assumed, the existence of a decoding function $\mathscr{D}^{(0)}$ such that

$$W^{(0)} = \mathscr{D}^{(0)}(Y_{1:n}), \tag{3.2.7}$$

is then guaranteed. Moreover, since $U = W^{(0)}$ it can be directly deduced that $W^{(0)}$ is a cv, in the sense of Def. 6, between U and $Y_{1:n}$. The maximal cv property of nonstochastic information I_* yields the following

$$\begin{aligned} nR^{(0)} &= \log_2 |[\![W^{(0)}]\!]| \\ &\leq I_*[U;Y_{1:n}]. \end{aligned} \tag{3.2.8}$$

This proves expression (3.2.5a) of Theorem 3.

Next, we show inequality (3.2.5b) for $j \in [1:M]$. Firstly, note that given a specific realization $W^{(0)} = w^{(0)}$ of the common message, there exists a unique message $w^{(1)}$ associated with the channel codeword $x_{1:n}^{(1)}$. This observation follows also from the zero-error property of the chosen code. In fact, if different realizations $W^{(1)} = w^{(1)}$ were mapped to the same codeword, then zero-error decoding would obviously be impossible and the assumption would be contradicted. Therefore, there certainly exists a function $g^{(j)}$ such that $\forall j \in [1:M]$:

$$W^{(j)} = g^{(j)}(X_{1:n}^{(j)}, W^{(0)}). \tag{3.2.9}$$

Moreover, by the zero-error property there is indeed a decoding function $\mathscr{D}^{(i)}$ such that

$$W^{(j)} = \mathscr{D}^{(j)}(Y_{1:n}), \ \ \forall j \in [1:M]. \tag{3.2.10}$$

Hence, we conclude that the uv $W^{(j)}$ is a cv between $(X_{1:n}^{(j)}, W^{(0)})$ and $(Y_{1:n}, W^{(0)})$. Recall that in this MAC model the private messages $W^{(j)}$ are unrelated with $U = W^{(0)}$ for all $j \in [1:M]$. Therefore,

the interpretation of conditional I_* in terms of maximal cv results in

$$\begin{aligned} nR^{(j)} &= \log_2 |[\![W^{(j)}]\!]| \\ &\leq I_*[X_{1:n}^{(j)}; Y_{1:n}|W^{(0)}] \\ &= I_*[X_{1:n}^{(j)}; Y_{1:n}|U], \quad \forall j \in [1:M], \end{aligned} \tag{3.2.11}$$

proving (3.2.5b).

Proof of Achievability

The achievability proof consists in showing that if we have a set of uv's U and $\{X_{1:n}^{(j)}\}_{j=1}^{M}$ for some block-length $n \geq 1$ such that the outlined requirements in Theorem 3 are fulfilled, then it is possible to construct a zero-error coding scheme at rates achieving (3.2.5a) and (3.2.5b) with equality.

Codebook Generation. We firstly fix the rate of the common message such that

$$R^{(0)} = \frac{I_*[U; Y_{1:n}]}{n}. \tag{3.2.12}$$

Next, select one point from each set of the family $[\![U|Y_{1:n}]\!]_*$. We then denote the chosen points $u(w^{(0)})$ with $w^{(0)} \in [1 : 2^{nR^{(0)}}]$.
Since $nR^{(j)} = I_*[X_{1:n}^{(j)}; Y_{1:n}|U]$, for $j \in [1:M]$, then (2.1.17) means that the following inequality holds

$$2^{nR^{(j)}} \leq \left| [\![X_{1:n}^{(j)}|Y_{1:n}, U = u(w^{(0)})]\!]_* \right|, \tag{3.2.13}$$

for all $j \in [1:M]$ and $w^{(0)} \in [1 : 2^{nR^{(0)}}]$. It is therefore possible to select $2^{nR^{(j)}}$ distinct codewords $x_{1:n}^{(j)}$ from $[\![X_{1:n}^{(j)}|U = u(w^{(0)})]\!]$ for any realization $w^{(0)}$ such that each nonempty set of the overlap partition $[\![X_{1:n}^{(j)}|Y_{1:n}, U = u(w^{(0)})]\!]_*$ contains exactly one element. Subsequently, these codewords denoted as $\mathscr{E}^{(j)}(w^{(0)}, w^{(j)})$, where $w^{(j)} \in [1 : 2^{nR^{(j)}}]$ and $j \in [1:M]$, are obtained by means of the encoding functions (3.2.1).

Zero-Error Decoding. In this part of the proof, we construct a decoding scheme that achieves exactly zero decoding errors using the previously introduced code.
Firstly, recall that the uv's U and $\{X_{1:n}^{(j)}\}_{j=1}^{M}$ satisfy

$$\prod_{j=1}^{M} \left[\!\!\left[X_{1:n}^{(j)}|U \right]\!\!\right] = \left[\!\!\left[X_{1:n}^{(1)}, \cdots, X_{1:n}^{(M)}|U \right]\!\!\right]. \tag{3.2.14}$$

Then, the codeword tuples $(\mathscr{E}^{(1)}(w^{(0)}, w^{(1)}), \cdots, \mathscr{E}^{(M)}(w^{(0)}, w^{(M)}))$ certainly belong to the conditional joint range $[\![X_{1:n}^{(1)}, \cdots, X_{1:n}^{(M)}|U = u(w^{(0)})]\!]$. This means that any combination of messages

$\{(w^{(0)}, w^{(j)})\}_{j=1}^{M}$ is mapped to a valid tuple of codewords lying within $[\![X_{1:n}^{(1)}, \cdots, X_{1:n}^{(M)}, U]\!]$. Eventually, the receiver performs the following $M+1$-stage decoding procedure:

(a) First of all, the decoder determines the transmitted common message $w^{(0)}$. By construction, each of the $2^{nR^{(0)}}$ points $u(w^{(0)})$ is inside a disjoint set of the family $[\![U|Y_{1:n}]\!]_*$. Furthermore, recall that the cv property of the overlap partition implies that each set in $[\![U|Y_{1:n}]\!]_*$ containing u also uniquely specifies the matching set in $[\![Y_{1:n}|U]\!]_*$ that contains $y_{1:n}$. Hence, the common message $w^{(0)}$ is decoded with zero error.

(b) Next, after having determined $w^{(0)}$, it is now possible to determine which set of the conditional overlap partition $[\![Y_{1:n}|X_{1:n}^{(1)}, U = u(w^{(0)})]\!]_*$ contains the sequence $y_{1:n}$. In a similar manner as Step **(a)**, this set uniquely determines the corresponding set of the family $[\![X_{1:n}^{(1)}|Y_{1:n}, U = u(w^{(0)})]\!]_*$ where the codeword $\mathscr{E}^{(1)}(w^{(0)}, w^{(1)})$ lies. Since at most one codeword has been selected from each set of this family for each realization $w^{(0)}$, then the private message of user 1, namely $w^{(1)}$, is uniquely decoded.

(c) In the subsequent $M-1$ stages, the decoder repeats Step **(b)** with $x_{1:n}^{(j)}$ for $j \in [2:M]$ and similarly recovers $w^{(j)}$ with zero error.

3.3 Zero-Error Communication over M-User MAC with Common Message & Pairwise Shared Messages

In the previous couple of sections, the inter-user relatedness was modeled by means of one common message seen by all encoders. Nonetheless, this structure is not always true, and it is also important to investigate channel models that incorporate some degree of flexibility in terms of the relatedness among the users. To this end, we study in this section the M-user MAC with both one common message as well as pairwise shared messages. After formally introducing the channel model, an intrinsic characterization of its zero-error capacity region $\mathscr{C}_0$ is derived in this section using tools from nonstochastic information theory.

3.3.1 System Model & Notation

Before introducing the system model, we start by explaining the notation used throughout this section. Firstly, let $M \in \mathbb{Z}_{\geq 2}$ be the number of users in the system. Furthermore, consider the index $k \in \{1, 2, M\}$ which corresponds to the number of encoders seeing the message, i.e.,

- $k = 1$ refers to a private message (only seen by one encoder);
- $k = 2$ refers to a pairwise shared message (seen by exactly two encoders); and
- $k = M$ refers to the common message to all users.

Now, let $\mathscr{I}_k$ denote the set of all k-combinations of message indices from $[1:M]$, and $\mathscr{I}_k(m)$ be the collection of subsets from $\mathscr{I}_k$ containing the index $m \in [1:M]$. The set $\mathscr{I}$ is defined as

$$\mathscr{I} := \bigcup_{k\in\{1,2,M\}} \mathscr{I}_k. \tag{3.3.1}$$

By way of example, if $M = 3,\ k = 2$ and $m = 2$, then the set of all pairwise shared message indices is $\mathscr{I}_2 = \{\{1,2\};\{1,3\};\{2,3\}\}$. Note that each member set of consists of the users' number to whom the corresponding message is common. Additionally, the set of indices of the pairwise shared messages available to the second user, i.e., $m = 2$, is $\mathscr{I}_2(2) = \{\{1,2\};\{2,3\}\}$.
We consider the multiple access communication system consisting of $\frac{M(M+1)+2}{2}$ sources[2], M users and one receiver. The m-th user, with $m \in [1:M]$, has access to the set of messages $\mathscr{W}^{(m)} = (W^{(i)})_{i\in\mathscr{I}(m)}$ where $\mathscr{I}(m) := \mathscr{I}_1(m) \cup \mathscr{I}_2(m) \cup \mathscr{I}_M(m)$. Note that the index i is an element-set from $\mathscr{I}$, and hence

- $\mathscr{I}_1(m)$ is a singleton containing the index of the private message $W^{(m)}$, which is only seen by the respective m-th user;
- $\mathscr{I}_2(m)$ is the set of the indices of all pairwise common messages $W^{(m,\ell)}$ available to the m-th user and shared with the ℓ-th user, for all $\ell \neq m \in [1:M]$. Notice that $W^{(m,\ell)} = W^{(\ell,m)}$;
- $\mathscr{I}_M(m)$ denotes the singleton containing the index of the common message $W^{(1,2,\cdots,M)}$, which is available to all M users. Note that $\mathscr{I}_M(m) = \mathscr{I}_M$ for any $m \in [1:M]$, i.e., the common message's index does not depend on m.

Moreover, as previously outlined, we denote by

- $\mathscr{I}_1$ the set of indices of all private messages $W^{(m)}$, $\forall m \in [1:M]$, i.e., $\mathscr{I}_1 := \cup_{m\in[1:M]}\mathscr{I}_1(m)$;
- $\mathscr{I}_2$ the set of indices of all pairwise shared messages $W^{(m,\ell)}$ for all $\ell \neq m \in [1:M]$ users in the communication system, i.e., $\mathscr{I}_2 := \cup_{m\in[1:M]}\mathscr{I}_2(m)$;
- $\mathscr{I}$ the set of indices of all transmitted messages in the setup as defined in (3.3.1).

Additionally, Let $\mathscr{I}_{\text{shared}}(m) := \mathscr{I}_2(m) \cup \mathscr{I}_M(m)$ and $\mathscr{I}_{\text{shared}} := \mathscr{I}_2 \cup \mathscr{I}_M$ denote the set of indices of all shared messages associated with the m-th encoder and that of all shared messages in the entire system, respectively. We assume that the messages are mutually unrelated and finite-valued. The range of the message $W^{(i)}$ is denoted by $[\![W^{(i)}]\!]$, $\forall i \in \mathscr{I}$. The structure of such a system for 3 users , i.e., $M = 3$, is shown in Fig. 3.4.

[2]The number of sources is equal to $\binom{M}{1} + \binom{M}{2} + \binom{M}{M}$, and hence the result above.

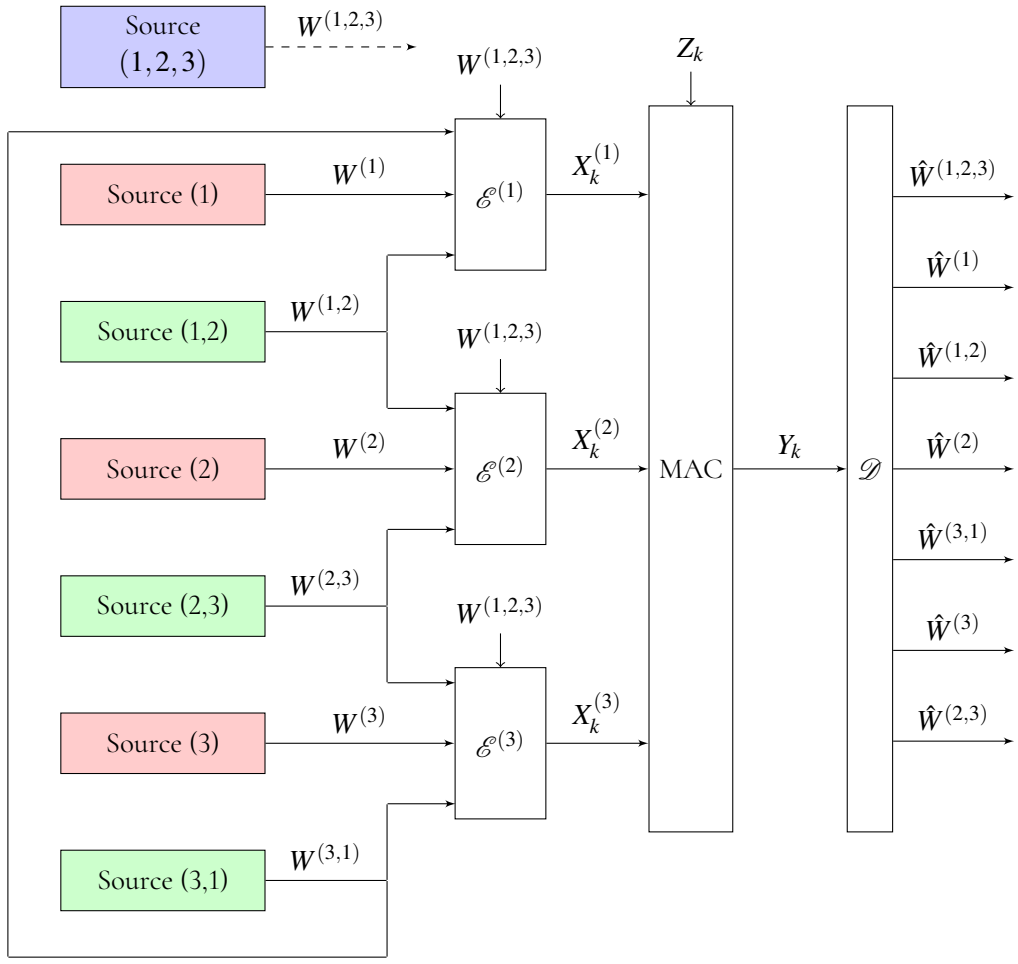


Figure 3.4: Illustration of a communication system deploying a three-user MAC with special correlation structure. The red-colored sources produce private messages, while the outputs of the green sources are seen by two encoders. The blue block refers to the source that generates the common message accessible by all three encoders.

Given the code block-length $n \in \mathbb{Z}_{\geq 1}$, the m-th user maps its respective set of messages, namely $\mathscr{W}^{(m)}$, into the channel input $X_{1:n}^{(m)}$ by means of the encoding law $\mathscr{E}^{(m)}$, i.e.,

$$X_{1:n}^{(m)} = \mathscr{E}^{(m)}\left(\mathscr{W}^{(m)}\right), \ \forall m \in [1:M]. \tag{3.3.2}$$

For each message $W^{(i)}$, $\forall i \in \mathscr{I}$, we define the code rate $R^{(i)}$ such that

$$R^{(i)} := \left(\log_2 |[\![W^{(i)}]\!]|\right)/n. \tag{3.3.3}$$

For a given block-length $n \geq 1$, the n-long sequences $\{X_{1:n}^{(m)}\}_{m=1}^{M}$ are then transmitted through the MAC, to generate the channel output $Y_{1:n}$ according to the memoryless MAC law (3.2.3) introduced in Def. 11. We present in Remark 8 an equivalent definition of the M-user MAC using the concept of set-valued channel transition functions. It turns out that, for the analysis of M-user MAC with common message and pairwise shared messages, this equivalent definition is more useful.

Remark 8 (Equivalent Definition of Stationary Memoryless Uncertain M-User MAC)**.** *Consider the product input space $\mathscr{X} = \mathscr{X}^{(1)} \times \cdots \times \mathscr{X}^{(M)}$, an output space $\mathscr{Y}$, and a set-valued transition function $\mathbf{T} : \mathscr{X} \mapsto 2^{\mathscr{Y}}$. A* stationary memoryless uncertain MAC *maps any tuple of input uncertain variables $X := (X^{(j)})_{j\in[1:M]}$ for any $(x_{1:k}, y_{1:k-1}) \in [\![X_{1:k}, Y_{1:k-1}]\!]$ to an output uncertain variable Y such that*

$$[\![Y_k | x_{1:k}^{(1)}, \cdots, x_{1:k}^{(M)}, y_{1:k-1}]\!] = \mathbf{T}(x_k^{(1)}, \cdots, x_k^{(M)}), \tag{3.3.4}$$

for all $k \in \mathbb{Z}_{\geq 1}$.

Upon reception of the full sequence $Y_{1:n}$, the decoder $\mathscr{D}$ produces estimates of the transmitted messages denoted by $(\hat{W}^{(i)})_{i\in\mathscr{I}}$. In a zero-error communication setup, these estimates match *exactly* the original messages, i.e., $\hat{W}^{(i)} = W^{(i)}$, $\forall i \in \mathscr{I}$. Hence, the decoder should be able to perfectly recover $(W^{(i)})_{i\in\mathscr{I}}$ in spite of the noise affecting the channel and the occurring inter-user interference. Consequently, given any output realization $y_{1:n} \in [\![Y_{1:n}]\!]$, the conditional range $[\![W^{(i)}|y_{1:n}]\!]$ must contain exactly *one* element $\forall i \in \mathscr{I}$.

For the code block-length n, the *operational zero-error n-capacity region* $\mathscr{C}_{0,n}$ of the MAC at hand is then defined as the set of rate vectors $R := (R^{(i)})_{i\in\mathscr{I}}$ for which it is possible to find encoding and decoding rules achieving exactly zero error. Note that when the code block-length is allowed to be arbitrarily large, then the notion of *achievable rate* becomes useful. We define achievable rates in a similar manner to Def. 13 as follows:

Definition 14. *We say that a rate tuple $R := (R^{(i)})_{i\in\mathscr{I}}$ is* achievable *if there exists a sequence $(\lceil 2^{nR_n^{(i)}} \rceil_{i\in\mathscr{I}}, n)_{n\in\mathbb{Z}_{\geq 1}}$ of zero-error codes with $R_n = (R_n^{(i)})_{i\in\mathscr{I}}$ approaching R as $n \to \infty$.*

3.3.2 Zero-Error Capacity Region of the M-User MAC with Common Message & Pairwise Shared Messages via Nonstochastic Information

We are now in a position to establish the main result of this section, namely to characterize the zero-error capacity region $\mathscr{C}_0$ of the presented class of M-user MAC.

Theorem 4. *Given the code block-length $n \in \mathbb{Z}_{\geq 1}$, consider the set of rate tuples $(R^{(i)})_{i\in\mathscr{I}}$ denoted by $\mathscr{R}((U^{(j)})_{j\in\mathscr{I}_{\text{shared}}}, (X_{1:n}^{(m)})_{m\in[1:M]})$ such that*

$$nR^{(j)} \leq I_*[U^{(j)}; Y_{1:n}], \tag{3.3.5a}$$

$$nR^{(m)} \leq I_*[X_{1:n}^{(m)}; Y_{1:n}|U^{\mathscr{I}_{\text{shared}}(m)}], \tag{3.3.5b}$$

where the indices $j \in \mathscr{I}_{\text{shared}}$ and $m \in [1:M]$. Furthermore, $2^{nR^{(j)}}$ and $2^{nR^{(m)}}$ are positive integers, $X_{1:n}^{(m)}$ denote the input sequences transmitted through the M-user MAC (3.3.4), $Y_{1:n}$ is the channel output, $U^{(j)}$ and $U^{\mathscr{I}_{\text{shared}}(m)}$ are auxiliary uncertain variables (uv's).

Then, the zero-error n-capacity region $\mathscr{C}_{0,n}$ of the channel at hand over n uses is given by the union of the regions $\mathscr{R}((U^{(j)})_{j\in\mathscr{I}_{\text{shared}}}, (X_{1:n}^{(m)})_{m\in[1:M]})$ over all uv's $(U^{(j)})_{j\in\mathscr{I}_{\text{shared}}}, (X_{1:n}^{(m)})_{m\in[1:M]}$ satisfying

i) $X_{1:n}^{(1)}, \cdots, X_{1:n}^{(m)}$ *are conditionally unrelated given* $(U^{(j)})_{j\in\mathscr{I}_{\text{shared}}}$,

ii) $X_{1:n}^{(m)} \leftrightarrow U^{\mathscr{I}_{\text{shared}}(m)} \leftrightarrow U^{\mathscr{I}_{\text{shared}}\setminus\mathscr{I}_{\text{shared}}(m)}$, $\forall m \in [1:M]$,

iii) $[\![Y_k|x_{1:k}^{(1)}, \cdots, x_{1:k}^{(M)}, y_{1:k-1}, u]\!] = \mathbf{T}(x_k^{(1)}, \cdots, x_k^{(M)})$ *for all* $(x_{1:k}, y_{1:k-1}, u) \in [\![X_{1:k}, Y_{1:k-1}, U^{\mathscr{I}_{\text{shared}}}]\!]$ *and all $k \in [1:n]$, with $\mathbf{T}$ is the set-valued channel transition function defined in (3.3.4).*

In contrast to the model studied in Section 3.2, where only one common message is seen by all encoders, in this part we considered a more general form of relatedness among the channel inputs. As a result of this generalization, (3.3.5b) involves all shared messages present in the setup, unlike (3.2.5b) obtained in Theorem 3 which only involves the common message.

It is also worth noting that the channel model investigated here can be extended to MACs with messages shared between three or more users. This represents a preliminary step towards the characterization of zero-error capacity regions in general networks with arbitrary correlations between multiple users.

This result obtained in Theorem 4 treats the users symmetrically, with the result that the number of auxiliary variables match the total number of shared messages. It is possible that an alternative analysis could yield fewer auxiliary uv's.

Proof of Converse

Consider a setup where M transmitters are communicating via the previously discussed multi-user MAC (3.3.4) and using a zero-error code (3.3.2) with block-length n. The code rates are given by

the tuple denoted as $(R^{(i)})_{i\in\mathscr{I}}$. Set the auxiliary uv's $U^{(j)}$ for all $j \in \mathscr{I}_{\text{shared}}$ as follows

$$U^{(j)} := W^{(j)}. \tag{3.3.6}$$

In other words, for each of the common messages in the communication setup, whether it is the common message to all users or the pairwise shared ones, we set an auxiliary uv. Notice that since all messages $W^{(i)}$ for all $i \in \mathscr{I}$ are mutually unrelated by assumption, it follows from the encoding rule (3.3.2) that each two codewords $X_{1:n}^{(m)}$ and $X_{1:n}^{(\ell)}$, with $m \neq \ell \in [1:M]$, are conditionally unrelated given the tuple $(U^{(1,2,\cdots,M)}, U^{(m,\ell)})$. Thus, the Markov chain given in item **(i)** of Theorem 4 is satisfied. Therefore, we obtain also the following Markov uncertainty chain $(U^{(i)})_{i\in\mathscr{I}_{\text{shared}}} \leftrightarrow (X_{1:n}^{(m)})_{m\in[1:M]} \leftrightarrow Y_{1:n}$.

We firstly start by proving (3.3.5a). By construction, there exists a decoding function $\mathscr{D}^{(i)}$ that generates an estimate $\hat{W}^{(i)}$ using the received sequence $Y_{1:n}$ for all $i \in \mathscr{I}$, i.e.,

$$\hat{W}^{(i)} := \mathscr{D}^{(i)}(Y_{1:n}). \tag{3.3.7}$$

Furthermore, in a zero-error communication setup the messages are perfectly recovered at the receiver end, and thus it holds

$$\begin{aligned} W^{(i)} &= \hat{W}^{(i)} \\ &= \mathscr{D}^{(i)}\left(Y_{1:n}\right), \ \forall i \in \mathscr{I}. \end{aligned} \tag{3.3.8}$$

As $j \in \mathscr{I}_{\text{shared}} \subset \mathscr{I}$, (3.3.8) holds in particular for all $j \in \mathscr{I}_{\text{shared}}$. Hence, from (3.3.6) and (3.3.8), we conclude that $W^{(j)}$ is a cv between $U^{(j)}$ and $Y_{1:n}$ for all $j \in \mathscr{I}_{\text{shared}}$. Hence, by cv property, we obtain

$$\log_2 |[\![W^{(j)}]\!]| \leq I_*[U^{(j)}; Y_{1:n}]. \tag{3.3.9}$$

Along with the definition (3.3.3), this result completes the necessity proof of (3.3.5a).

Now we derive the bound (3.3.5b) on the private messages' rates, namely $R^{(m)}$, $\forall m \in \mathscr{I}_1 = [1:M]$. Consider the vector of auxiliary uv's $(U^{\mathscr{I}_{\text{shared}}(m)})$. Furthermore, we denote by $(W^{\mathscr{I}_{\text{shared}}(m)})$ the collection of all common messages fed into the user $\mathscr{E}^{(m)}$. Together with the private message $W^{(m)}$, $W^{\mathscr{I}_{\text{shared}}(m)}$ form the complete message tuple $\mathscr{W}^{(m)} = (W^{(m)}, W^{\mathscr{I}_{\text{shared}}(m)})$ encoded by $\mathscr{E}^{(m)}$ (3.3.2).

The zero-error property implies that for a given tuple of messages $\mathscr{W}^{(m)}$ there exists a *unique* codeword $X_{1:n}^{(m)} = x_{1:n}^{(m)}(w^{(m)}, w^{\mathscr{I}_{\text{shared}}(m)})$. Subsequently, if the tuple $(X_{1:n}^{(m)}, W^{\mathscr{I}_{\text{shared}}(m)})$ was given, there certainly exists a bijective function $g^{(m)}$ such that

$$W^{(m)} = g^{(m)}\left(X_{1:n}^{(m)}, W^{\mathscr{I}_{\text{shared}}(m)}\right), \tag{3.3.10}$$

with $W^{(m)}$ being unrelated with all messages $W^{\mathscr{I}_{\text{shared}}(m)}$. Additionally, we have by construction

$$\hat{W}^{(m)} = \mathscr{D}^{(m)}(Y_{1:n}) \tag{3.3.11}$$

$$\equiv \mathscr{D}^{(m)}\left(Y_{1:n}, W^{\mathscr{I}_{\text{shared}}(m)}\right). \tag{3.3.12}$$

Under zero-error objective, $W^{(m)} = \hat{W}^{(m)}$. Hence, from (3.3.10) and (3.3.12), it can be concluded that $W^{(m)}$ is a cv between $(X_{1:n}^{(m)}, W^{\mathscr{I}_{\text{shared}}(m)})$ and $(Y_{1:n}, W^{\mathscr{I}_{\text{shared}}(m)})$. By (2.1.17) of the conditional nonstochastic information, we obtain

$$\begin{aligned}\log_2 |[\![W^{(m)}]\!]| &\leq I_*[X_{1:n}^{(m)}; Y_{1:n} | W^{\mathscr{I}_{\text{shared}}(m)}] \\ &= I_*[X_{1:n}^{(m)}; Y_{1:n} | U^{\mathscr{I}_{\text{shared}}(m)}].\end{aligned} \tag{3.3.13}$$

This concludes the converse proof.

Proof of Achievability

Given a block-length $n \in \mathbb{Z}_{\geq 1}$ and a set of uv's $U^{(j)}$ and $X_{1:n}^{(m)}$, with $j \in \mathscr{I}_{\text{shared}}$ and $m \in [1:M]$, such that requirements *i)*, *ii)* and *iii)* in Theorem 4 are fulfilled, we construct a zero-error coding scheme operating at rates that achieve equality in (3.3.5a)-(3.3.5b).

Codebook Generation. Initially, we consider the rates of all common messages (both pairwise and to all users), i.e., $R^{(j)}$ with $j \in \mathscr{I}_{\text{shared}}$. Set each of these rates such that it satisfies

$$R^{(j)} = \left(I_*\left[U^{(j)}; Y_{1:n}\right]\right)/n, \ \forall j \in \mathscr{I}_{\text{shared}}. \tag{3.3.14}$$

From the disjoint sets of the overlap partition $[\![U^{(j)}|Y_{1:n}]\!]_*$, we select exactly one point that we call $u^{(j)}(w^{(j)})$ with $w^{(j)} \in [1:2^{nR^{(j)}}]$.

Next, after having fixed the common messages we construct the privately transmitted ones, i.e., $W^{(m)}$, $\forall m \in [1:M]$. Recall that the rates of $W^{(m)}$ are given as $nR^{(m)} = I_*[X_{1:n}^{(m)}; Y_{1:n} | U^{\mathscr{I}_{\text{shared}}(m)}]$. By (2.1.17), we know that

$$2^{nR^{(m)}} \leq \left| \left[\!\!\left[X_{1:n}^{(m)} | Y_{1:n}; u^{\mathscr{I}_{\text{shared}}(m)}(w^{\mathscr{I}_{\text{shared}}(m)}) \right]\!\!\right]_* \right| \tag{3.3.15}$$

This implies that for any realization of common message tuple $w^{\mathscr{I}_{\text{shared}}(m)}$, it is possible to find $2^{nR^{(m)}}$ distinct codewords $x_{1:n}^{(m)}$ for all $m \in [1:M]$ such that no more than one is drawn from the same member-set of the family $[\![X_{1:n}^{(m)}|Y_{1:n}; u^{\mathscr{I}_{\text{shared}}(m)}(w^{\mathscr{I}_{\text{shared}}(m)})]\!]_*$. We call these points $\mathscr{E}^{(m)}(w^{(m)}, w^{\mathscr{I}_{\text{shared}}(m)})$. This establishes the encoding scheme (3.3.2).

Zero-Error Decoding. This part of the proof shows that using the constructed encoding procedure, it is possible to achieve an error probability exactly equal to zero. Consider the conditional range $[\![X_{1:n}^{(1)},\cdots,X_{1:n}^{(M)}|U^{(j)}]\!]$ for all $j\in\mathscr{I}_{\text{shared}}$. It can be written as follows

$$\left[\!\left[X_{1:n}^{(1)},\cdots,X_{1:n}^{(M)}|U^{(j)}\right]\!\right]=\prod_{m=1}^{M}\left[\!\left[X_{1:n}^{(m)}|U^{(j)}\right]\!\right],\ \forall j\in\mathscr{I}_{\text{shared}} \tag{3.3.16}$$

$$=\prod_{m=1}^{M}\left[\!\left[X_{1:n}^{(m)}|U^{\mathscr{I}_{\text{shared}}(m)}\right]\!\right], \tag{3.3.17}$$

where (3.3.16) and (3.3.17) result from items **(i)** and **(ii)** in Theorem 4, respectively. Hence, as the auxiliary uv's are set to the common messages, it is guaranteed that for all configurations of common messages $w^{\mathscr{I}_{\text{shared}}(m)}$, all M-dimensional codeword tuples

$$\left(\mathscr{E}^{(1)}(w^{(1)},w^{\mathscr{I}_{\text{shared}}(1)}),\cdots,\mathscr{E}^{(M)}(w^{(M)},w^{\mathscr{I}_{\text{shared}}(M)})\right)$$

belong to the conditional joint range $[\![X_{1:n}^{(1)},\cdots,X_{1:n}^{(M)}|U^{\mathscr{I}_{\text{shared}}}]\!]$. This means that for any combination of message realizations, the $\frac{M(M+1)+2}{2}$-dimensional tuple

$$\left(\mathscr{E}^{(1)}(w^{(1)},w^{\mathscr{I}_{\text{shared}}(1)}),\cdots,\mathscr{E}^{(M)}(w^{(M)},w^{\mathscr{I}_{\text{shared}}(M)}),u^{(j)}(w^{(j)})\right),$$

with $j\in\mathscr{I}_{\text{shared}}$, is well defined in the joint range $[\![X_{1:n}^{(1)},\cdots,X_{1:n}^{(M)},U^{\mathscr{I}_{\text{shared}}}]\!]$.

Upon reception of the sequence $Y_{1:n}$, the receiver recovers the $\frac{M(M+1)+2}{2}$ transmitted messages errorlessly by proceeding as follows

(a) It starts by reconstructing all shared messages $w^{(j)},\ \forall j\in\mathscr{I}_{\text{shared}}$. Since each of the $2^{nR^{(j)}}$ selected points $u^{(j)}(w^{(j)})$ is drawn from a *distinct* member-set of the family $[\![U^{(j)}|Y_{1:n}]\!]_*$, it is possible to determine $[\![U^{(j)}|Y_{1:n}=y_{1:n}]\!]$ by using the cv property of overlap partitions to find the set of the corresponding overlap partition $[\![Y_{1:n}|U^{(j)}]\!]_*$ in which lies the received sequence $y_{1:n}$. Consequently, the message w^j is retrieved with exactly zero error for all $j\in\mathscr{I}_{\text{shared}}$.

(b) After having determined the shared messages, the member-set of the family $[\![Y_{1:n}|X_{1:n}^{(m)},u^{\mathscr{I}_{\text{shared}}(m)}(w^{\mathscr{I}_{\text{shared}}(m)})]\!]_*$ containing the sequence $y_{1:n}$ can be computed. In a similar manner to Step **(a)**, the corresponding set of the conditional overlap partition $[\![X_{1:n}^{(m)}|Y_{1:n},u^{\mathscr{I}_{\text{shared}}(m)}(w^{\mathscr{I}_{\text{shared}}(m)})]\!]_*$ where the codeword $\mathscr{E}^{(m)}(w^{(m)},w^{\mathscr{I}_{\text{shared}}(m)})$ lies can be determined, $\forall m\in[1:M]$. By construction, it is known that for each message tuple $(w^{(m)},w^{\mathscr{I}_{\text{shared}}(m)})$, each set of this family contains no more than one codeword. Hence, the private message $w^{(m)}$ for all $m\in[1:M]$ is errorlessly reconstructed. This operation is performed for all M encoders.

3.4 Summary

In this chapter, we derived a multi-letter characterization of the zero-error capacity region for three different classes of MACs. Unlike classical information theory, the tools from its nonstochastic analogue allowed us to obtain a result that is not only valid for asymptotically large coding block-lengths, but also true for finite ones. Firstly, we started with the two-user MAC with one common message that models existing correlation among both users. This setup is inspired from the work of Slepian and Wolf [24]. Next, this model is extended to accommodate M users instead of two, with $M \geq 2$. In this case, we have M private sources, whose messages are seen only by the respective transmitter, in addition to one common message among all users. Finally, we consider a system where the inter-user correlation is not only modeled by means of a single common message, but includes also pairwise shared messages between the M transmitters. Both converse and achievability proofs are established for the channel coding theorems of the different MACs.

In the upcoming chapter, the result of Section 3.1, namely Theorem 2, is used to find tight conditions for achieving bounded state estimation errors over a two-user MAC.

"Science is beautiful when it makes simple explanations of phenomena or connections between different observations."

— Stephen Hawking (1942–2018)

4

Distributed State Estimation over Nonstochastic MACs with Bounded Noise

Following the characterization in Chapter 3 of the zero-error capacity region $\mathscr{C}_0$ of different MAC classes, and in particular the two-user MAC with one common message, we are now in a suitable position to study the problem of distributed state estimation over such channels. The setup that is investigated throughout this chapter consists of two sensors measuring the output of three mutually unrelated dynamical systems. The output signals of two of these plants are privately measured by each of the sensors, whereas the output of the remaining system is common to both sensors. The aim of such formulation is to capture some of the essential elements of distributed state estimation problem, e.g., where each sensor observes a different subset of the overall system's dynamical modes and any possible overlap between these subsets is modeled by the common plant. Following a rigorous and detailed problem formulation, we prove a theorem (Theorem 5) linking the intrinsic properties of the dynamical systems and the zero-error capacity of the communication channel. To this end, we first present a necessity proof using the notions of nonstochastic information as well as nonstochastic conditional information, followed by showing the sufficient condition by constructing a suitable scheme. This setup is then shown to be general for noiseless linear systems. Finally, an example of state estimation over *binary adder channel* (BAC) is discussed. Please note that, unless otherwise stated, the ℓ_∞-norm, denoted by the symbol $\|\cdot\|$, is used in the remainder of this thesis.

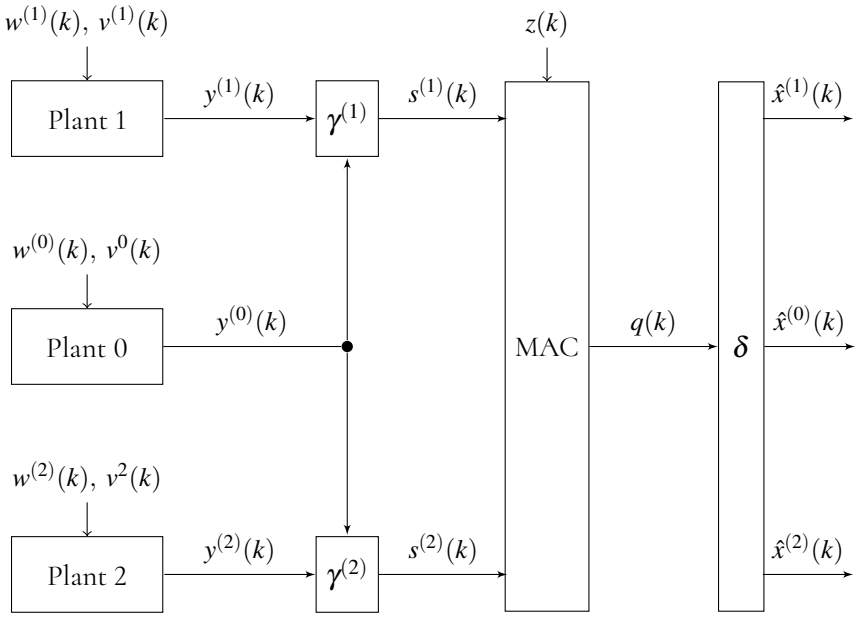


Figure 4.1: State estimation of LTI systems with disturbances over a two-input single-output MAC channel.

4.1 Problem Formulation

We consider three discrete LTI dynamical systems characterized by the following system equations for $i \in \{0,1,2\}$:

$$X^{(i)}(k+1) = A^{(i)}X^{(i)}(k) + V^{(i)}(k) \in \mathbb{R}^{d_i}, \tag{4.1.1a}$$

$$Y^{(i)}(k) = C^{(i)}X^{(i)}(k) + W^{(i)}(k) \in \mathbb{R}^{b_i}, \tag{4.1.1b}$$

where $A^{(i)} \in \mathbb{R}^{d_i \times d_i}$ refers to the state matrix, $C^{(i)} \in \mathbb{R}^{b_i \times d_i}$ is the output matrix and the uncertain variables (uv's) $V^{(i)}(k)$ and $W^{(i)}(k)$ denote process and measurement noise at time instant $k \in \mathbb{Z}_{\geq 0}$. For the remainder of this thesis, time is indexed as an argument in parentheses rather than as a subscript.

Before being transmitted, the plant output sequences are first encoded into channel input signals $S^{(1)}(k), S^{(2)}(k)$ via the coding functions $\gamma^{(j)}$ as

$$S^{(j)}(k) = \gamma^{(j)}(k, Y^{(0)}(0:k), Y^{(j)}(0:k)),\ j \in \{1,2\}. \tag{4.1.2}$$

Note that the outputs of System 0 are available to both encoders, whereas Systems 1 and 2 are observed only by their respective users. The encoded data sequences are then sent through a stationary memoryless two-user MAC (Def. 8) as shown in Fig. 4.1. The received symbol $Q(k)$ is the output of

a fixed function $f : \mathscr{S}^1 \times \mathscr{S}^2 \times \mathscr{Z} \to \mathscr{Q}$,

$$Q(k) = f(S^{(1)}(k), S^{(2)}(k), Z(k)) \in \mathscr{Q}, \quad k \in \mathbb{Z}_{\geq 0}, \tag{4.1.3}$$

where $Z(k)$ is the channel noise at time k. At the receiver side, the symbols are used to generate an estimation $\hat{X}(k) := \left(\hat{X}^{(0)}(k), \hat{X}^{(1)}(k), \hat{X}^{(2)}(k)\right)^T$ of the original plant states $X(k) := \left(X^{(0)}(k), X^{(1)}(k), X^{(2)}(k)\right)^T$ by means of decoder δ, i.e.,

$$\hat{X}(k) = \delta(k, Q(0:k)) = \begin{pmatrix} \delta^{(0)}(k, Q(0:k)) \\ \delta^{(1)}(k, Q(0:k)) \\ \delta^{(2)}(k, Q(0:k)) \end{pmatrix} \in \mathbb{R}^d, \tag{4.1.4}$$

where $d = \sum_{i=0}^{2} d_i$ and, for $k = 0$, the initial estimate $\hat{X}(0) = 0$. The prediction error is denoted by the uv

$$E(k) := \begin{pmatrix} E^{(0)}(k) \\ E^{(1)}(k) \\ E^{(2)}(k) \end{pmatrix} = \begin{pmatrix} X^{(0)}(k) - \hat{X}^{(0)}(k) \\ X^{(1)}(k) - \hat{X}^{(1)}(k) \\ X^{(2)}(k) - \hat{X}^{(2)}(k) \end{pmatrix} \in \mathbb{R}^d. \tag{4.1.5}$$

Consider now the following definition.

Definition 15 (Uniformly Bounded Errors)**.** *The state estimation errors are said to be* uniformly bounded *if* $\exists l > 0$ *such that for any initial condition with range* $[\![X(0)]\!]$*, that lies within the closed ball* $\mathbf{B}_l \subseteq \mathbb{R}^d$ *centered at the origin with radius* l*, i.e.,* $[\![X(0)]\!] \subseteq \mathbf{B}_l$*, it holds*

$$\sup_{k\geq 0} [\![\|E(k)\|]\!] = \sup_{k\geq 0} [\![\|X(k) - \hat{X}(k)\|]\!] < \infty. \tag{4.1.6}$$

The aim is to design the coder-estimator tuple $(\gamma^{(1)}, \gamma^{(2)}, \delta)$ such that the resulting estimation errors are *uniformly bounded* in the sense of Def. 15 with respect to the infinity norm. We impose the following assumptions $\forall i, j$ and $h \in \{0,1,2\}$ and $\forall k, t \in \mathbb{Z}_{\geq 0}$.

(A1) Each matrix pair $(C^{(i)}, A^{(i)})$ is observable.

(A2) The noise terms $V^{(i)}(k)$, $W^{(i)}(k)$ are uniformly bounded, i.e., $\sup_{k\geq 0}[\![\|V(k)\|]\!], \sup_{k\geq 0}[\![\|W(k)\|]\!] < \infty$, where $V(k) := (V^{(0)}(k), V^{(1)}(k), V^{(2)}(k))^T$ and $W(k) := (W^{(0)}(k), W^{(1)}(k), W^{(2)}(k))^T$.

(A3) The initial states $X^{(i)}(0)$ and the noise signals $V^{(j)}(k)$, $W^{(h)}(t)$ are all mutually unrelated, $\forall i,\ j,\ h \in \{0,1,2\}$ and $\forall k, t \geq 0$. .

(A4) The channel noise signal $Z(k)$ is unrelated with the combined initial state and noise signals $(X(0), V(0:k-1), W(0:k))$.

(A5) Each system matrix $A^{(i)}$ has one or more unstable eigenvalues $\lambda_\ell^{(i)}$, i.e., $|\lambda_\ell^{(i)}| \geq 1$, where the subscript ℓ corresponds to the index of the unstable eigenvalue.

(A6) The zero signal is a valid realization of measurement and process noise, i.e., $0 \in [\![V^{(i)}]\!], [\![W^{(i)}]\!]$.

Remark 9. *Note that the reason for assumption **(A5)** is that the main focus of this work is instability, i.e., $|\lambda_\ell^{(i)}| \geq 1$. In fact, for strictly stable systems, the spectral radius of the state matrix $A^{(i)}$, denoted by $\rho(A^{(i)})$, is*

$$\begin{aligned} \rho(A^{(i)}) &:= \max\{|\lambda_1^{(i)}|, \cdots, |\lambda_\ell^{(i)}|\} \\ &< 1. \end{aligned} \tag{4.1.7}$$

Furthermore, it is established that matrices whose spectral radius is strictly less then 1 are convergent (see, e.g., [41, pg. 348]), i.e., the following holds

$$\lim_{k\to\infty} (A^{(i)})^k = 0, \tag{4.1.8}$$

*and since also the process noise $V^{(i)}$ is assumed to be bounded (by virtue of **(A2)**), then the estimation error (4.1.5) for such systems is guaranteed to be bounded. Hence, w.l.o.g., we can omit systems that are strictly stable and restrict our attention to state matrices that contain only unstable eigenvalues.*

4.2 Uniformly Bounded State Estimation over MACs

We now present the main result of this chapter. First, for each subsystem matrix $A^{(i)} \in \mathbb{R}^{d_i \times d_i}$, let the *topological entropy* be given by

$$h^{(i)} := \sum_{1 \leq \ell \leq d_i : |\lambda_\ell| \geq 1} \log_2 |\lambda_\ell^{(i)}|. \tag{4.2.1}$$

Theorem 5. *Consider the linear time-invariant systems in (4.1.1a)-(4.1.1b) whose outputs are coded (4.1.2) and estimated (4.1.4) via the two-input single output MAC (4.1.3). Suppose Assumptions (A1)-(A6) hold. If there exists a coder-estimator tuple $(\gamma^{(1)}, \gamma^{(2)}, \delta)$ yielding uniformly bounded estimation errors (Def. 15), then*

$$h := \left(h^{(0)}, h^{(1)}, h^{(2)}\right)^T \in \mathscr{C}_0, \tag{4.2.2}$$

where h is the vector of topological entropies of the corresponding systems and $\mathscr{C}_0$ refers to the zero-error capacity region (3.1.4) of the channel.
Furthermore, if $h \in \operatorname{int}(\mathscr{C}_0)$, then a coder-estimator tuple that achieves uniformly bounded state estimation errors can be constructed.

Remark 10. *Note that the interior region of $\mathscr{C}_0$ is guaranteed to be non-empty, when Requirements **(i)** and **(ii)** of Lemma 1 are fulfilled.*

Remark 11. *This result is an extension of [3, 12], from centralized LTI systems with point-to-point channels, to distributed LTI systems estimated over a MAC.*

Remark 12. *The case where $\boldsymbol{h}$ lies exactly on the boundary of the zero-error capacity region, i.e., $\boldsymbol{h} \in \mathscr{C}_0 \setminus \operatorname{int}(\mathscr{C}_0)$ introduces some technical issues and is not addressed here.*

Remark 13. *Although the formulation with three decoupled subsystems may seem special, it will be shown later in Section 4.3 that this setup is reasonably general for the case of noiseless linear systems.*

Next, we present both necessity and sufficiency proofs of Theorem 5.

4.2.1 Necessity Proof

Without loss of generality assume that for $i \in \{0,1,2\}$ the state matrix $A^{(i)}$ of Plant i is in *real Jordan canonical* form, i.e., it consists of m square blocks on its diagonal such that the j-th block is denoted by $A_j^{(i)} \in \mathbb{R}^{d_{i,j} \times d_{i,j}}$, with $j \in [1:m]$ and $\sum_{j=1}^{m} d_{i,j} = d_i$:

$$A^{(i)} = \begin{pmatrix} A_1^{(i)} & 0 & \cdots & 0 \\ 0 & A_2^{(i)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m^{(i)} \end{pmatrix} \in \mathbb{R}^{d_i \times d_i}. \tag{4.2.3}$$

In the following proof we only consider the unstable system eigenvalues without loss of generality. Furthermore, it follows from Def. 15 that uniformly bounded errors are obtained for any uniformly bounded noise ranges $[\![V(k)]\!]$ and $[\![W(k)]\!]$ as well as any initial state range $[\![X(0)]\!]$ contained in some l-ball $\mathbf{B}_l$. Hence, for the upcoming analysis we set the noise terms to zero, i.e., $[\![V(k)]\!] = [\![W(k)]\!] = \{0\}$ and construct the range of initial states in the following manner. First, select $\varepsilon \in \left(0, 1 - \max_{\ell:|\lambda_\ell^{(i)}|\geq 1} \frac{1}{|\lambda_\ell^{(i)}|}\right)$. Note that this is guaranteed by virtue of Assumption **(A5)**. We then divide the interval $[-l,l]$ on the ℓ-th axis into κ_ℓ equal subintervals of length $2l/\kappa_\ell$ such that

$$\kappa_\ell := \left\lfloor \left| (1-\varepsilon)\,\lambda_\ell^{(i)} \right|^k \right\rfloor, \quad \ell \in [1:d_i]. \tag{4.2.4}$$

Let $p_\ell(s)$ denote the midpoints of the subintervals, for $s = [1:\kappa_\ell]$ and construct an interval $\mathbf{I}_\ell(s)$ centered at $p_\ell(s)$ such that its length is equal to l/κ_ℓ. We define the family of hypercuboids $\mathscr{H}$ as

follows

$$\mathscr{H} = \left\{ \left(\prod_{\ell=1}^{d_i} \mathbf{I}_\ell(s) \right) : s \in [1 : \kappa_\ell],\ \ell \in [1 : d_i] \right\}. \tag{4.2.5}$$

Observe that any two hypercuboids from $\mathscr{H}$ are separated by a distance of l/κ_ℓ along the ℓ-th axis for each $\ell \in [1 : d_i]$. In the following analysis, the superscript i referring to the respective plant is omitted unless otherwise stated. The initial state range is set as $[\![X^{(i)}(0)]\!] = \bigcup_{\mathbf{H}\in\mathscr{H}} \mathbf{H} \subset \mathbf{B}_l(0)$. The operator $\mathrm{dm}(\cdot)$ is defined as the diameter of a set using the ∞-norm. Then, as $[\![E_j^{(i)}(k)]\!] \supseteq [\![E_j^{(i)}(k)|q(0:k-1)]\!]$,

$$\begin{aligned}
\mathrm{dm}([\![E_j^{(i)}(k)]\!]) &\geq \mathrm{dm}\left([\![E_j^{(i)}(k)|q(0:k-1)]\!]\right) && (4.2.6)\\
&= \mathrm{dm}\left([\![X_j^{(i)}(k) - \delta_j^{(i)}(k, q(0:k-1))\,|q(0:k-1)]\!]\right)\\
&= \mathrm{dm}\left([\![(A_j^{(i)})^k X_j^{(i)}(0)|q(0:k-1)]\!]\right) && (4.2.7)\\
&= \sup_{r,p\in[\![X_j^{(i)}(0)|q(0:k-1)]\!]} \left\| (A_j^{(i)})^k (r-p) \right\|\\
&\geq \sup_{r,p\in[\![X_j^{(i)}(0)|q(0:k-1)]\!]} \frac{\|(A_j^{(i)})^k (r-p)\|_2}{\sqrt{d_i}}\\
&\geq \sup_{r,p\in[\![X_j^{(i)}(0)|q(0:k-1)]\!]} \frac{\sigma_{\min}\left(\left(A_j^{(i)}\right)^k\right)\|r-p\|_2}{\sqrt{d_i}}\\
&\geq \sigma_{\min}\left(\left(A_j^{(i)}\right)^k\right) \frac{\mathrm{dm}\left([\![X_j^{(i)}(0)|q(0:k-1)]\!]\right)}{\sqrt{d_i}}, && (4.2.8)
\end{aligned}$$

where $k \in \mathbb{Z}_{\geq 0}$, $q(0:k-1) \in [\![Q(0:k-1)]\!]$, $\|\cdot\|_2$ denotes the Euclidean norm and $\sigma_{\min}(\cdot)$ refers to the smallest singular value. The inequalities (4.2.6) and (4.2.7) hold because conditioning reduces the range and the diameter of a uv range is translation invariant, respectively. Now, note that the Yamamoto identity for asymptotically large k states that

$$\lim_{k\to\infty} \left(\sigma_{\min}((A_j^{(i)})^k)\right)^{1/k} = \left|\lambda_{\min}\left(A_j^{(i)}\right)\right|, \tag{4.2.9}$$

with $\lambda_{\min}$ being the eigenvalue with the smallest magnitude. Then, since $j \in [1 : m] < \infty$, i.e., there are finitely many blocks $A_j^{(i)}$, there exists $k_\varepsilon \in \mathbb{Z}_{\geq 0}$ such that

$$\sigma_{\min}\left((A_j^{(i)})^k\right) \geq \left(1 - \frac{\varepsilon}{2}\right)^k |\lambda_{\min}(A_j^{(i)})|^k, \text{ for } k \geq k_\varepsilon. \tag{4.2.10}$$

Additionally, by uniform boundedness of errors (4.1.6) there exists $\xi > 0$ such that

$$\begin{aligned}
\xi &\geq \sup \left[\!\!\left[\|E^{(i)}(k)\| \right]\!\!\right] \\
&\geq \sup \left[\!\!\left[\|E_j^{(i)}(k)\| \right]\!\!\right] \\
&\geq \frac{1}{2} \mathrm{dm} \left([\![E_j^{(i)}(k)]\!] \right) \\
&\geq \left| \left(1 - \frac{\varepsilon}{2}\right) \lambda_{\min}(A_j^{(i)}) \right|^k \frac{\mathrm{dm} \left([\![X_j^{(i)}(0) | q(0:k-1)]\!] \right)}{2\sqrt{d_i}}.
\end{aligned} \tag{4.2.11}$$

For large enough $k \in \mathbb{N}$, the hypercuboid family $\mathscr{H}$ in (4.2.5) is an $[\![X^{(i)}(0)|Q(0:k-1)]\!]$-overlap isolated partition of $[\![X^{(i)}(0)]\!]$. To show this, we suppose in contradiction that $\exists \mathbf{H} \in \mathscr{H}$ that is overlap connected in the family $[\![X^{(i)}(0)|Q(0:k-1)]\!]$ with another hypercuboid from $\mathscr{H}$. Therefore, there would exist a set $[\![X^{(i)}(0)|q(0:k-1)]\!]$ which contains a point $r_j \in \mathbf{H}$ and a point $p_j \in \mathbf{H}'$, with $\mathbf{H}' \in \mathscr{H} \setminus \mathbf{H}$ such that r_j and p_j are overlap connected. This implies

$$\begin{aligned}
\|p_j - r_j\| &\leq \mathrm{dm} \left([\![X^{(i)}(0) | q(0:k-1]]\!] \right) \\
&\leq \frac{2\sqrt{d_i}\xi}{\left| \left(1 - \frac{\varepsilon}{2}\right) \lambda_{\min}(A_j^{(i)}) \right|^k},
\end{aligned} \tag{4.2.12}$$

for $j \in [1:m]$ and $k \geq k_\varepsilon$. Nonetheless, note that by construction the distance between any two hypercuboids in $\mathscr{H}$ is equal to l/κ_ℓ along the ℓ_i-th axis. Therefore,

$$\begin{aligned}
\|p_j - r_j\| &\geq \frac{l}{\kappa_\ell} \\
&= \frac{l}{\left\lfloor (1-\varepsilon) |\lambda_\ell^{(i)}| \right\rfloor^k} \\
&= \frac{l}{(1-\varepsilon)^k |\lambda_{\min}(A_j^{(i)})|^k}
\end{aligned} \tag{4.2.13}$$

For sufficiently large k, it is possible to obtain $\left((1-\varepsilon/2)/(1-\varepsilon)\right)^k > 2\sqrt{d_i}\xi/l$. Hence, the RHS of (4.2.13) would exceed the RHS of (4.2.12) resulting in a contradiction. Thus, when k is large enough, the family $\mathscr{H}$ is $[\![X^i(0)|Q(0:k-1)]\!]$-overlap isolated partition of $[\![X^i(0)]\!]$. As the cardinality of any $[\![X^{(i)}(0)|Q(0:k-1)]\!]$-overlap isolated partition is upper bounded by the nonstochastic information $\left|[\![X^{(i)}(0)|Q(0:k-1)]\!]_*\right|$, we obtain

$$\begin{aligned}
I_*[X^{(i)}(0); Q(0:k-1)] &= \log |[\![X^{(i)}(0)|Q(0:k-1)]\!]_*| \\
&\geq \log |\mathscr{H}|
\end{aligned}$$

$$
\begin{aligned}
&= \log\left(\prod_{\ell=1}^{d_i} \kappa_\ell\right) \\
&= \log\left(\prod_{\ell=1}^{d_i} \left\lfloor |(1-\varepsilon)\lambda_\ell^{(i)}|^k \right\rfloor\right) \\
&\geq \log\left(\prod_{\ell=1}^{d_i} 0.5 \left|(1-\varepsilon)\lambda_\ell^{(i)}\right|^k\right) \\
&= \log\left(2^{-d_i}(1-\varepsilon)^{kd_i}\left|\prod_{\ell=1}^{d_i}\lambda_\ell^{(i)}\right|^k\right) \\
&= k\left(d_i\log(1-\varepsilon) - \frac{d_i}{k} + \sum_{\ell=0}^{d_i}\log\left|\lambda_\ell^{(i)}\right|\right). \qquad (4.2.14)
\end{aligned}
$$

Hence, for $i \in \{0,1,2\}$ it follows from (4.2.14)

$$
\begin{aligned}
\frac{I_*[X^{(i)}(0:k-1);Q(0:k-1)]}{k} &\geq \frac{I_*[X^{(i)}(0);Q(0:k-1)]}{k} \\
&\geq d_i\log(1-\varepsilon) - \frac{d_i}{k} + \sum_{\ell=0}^{d_i}\log|\lambda_\ell^{(i)}|. \qquad (4.2.15)
\end{aligned}
$$

Thus, for $k \to \infty$ and arbitrarily small ε we obtain

$$
\begin{aligned}
\frac{I_*[X^{(i)}(0:k-1);Q(0:k-1)]}{k} &\geq \frac{I_*[X^{(i)}(0);Q(0:k-1)]}{k} \\
&\geq \sum_{\ell=0}^{d_i}\log|\lambda_\ell^{(i)}|. \qquad (4.2.16)
\end{aligned}
$$

Before proceeding with the proof, we present the following lemma.

Lemma 2. *Let Λ, Ω and Θ denote three uv's such that Λ and Θ are mutually unrelated. Then, the following relationship between $I_*[\Lambda;\Omega|\Theta]$ (2.1.12) and $I_*[\Lambda;\Omega]$ (2.1.17) holds*

$$I_*[\Lambda;\Omega|\Theta] \geq I_*[\Lambda;\Omega]. \qquad (4.2.17)$$

Proof. To prove Lemma 2 we use the common variable interpretation of (conditional) nonstochastic information as introduced in (2.1.12) and (2.1.17), respectively. Consider three uv's Λ, Θ and Ω such that Λ and Θ are unrelated. Let $\mathscr{Z}_\Theta$ denote the set of all uv's Z_Θ such that $Z_\Theta \perp \Theta$ and $Z_\Theta \equiv f(\Lambda,\Theta) = g(\Omega,\Theta)$. Thus,

$$I_*[\Lambda;\Omega|\Theta] = \max_{Z_\Theta \in \mathscr{Z}_\Theta} \log|[\![Z_\Theta]\!]|. \qquad (4.2.18)$$

Additionally, the set $\mathscr{Z}$ consists of all uv's Z such that $Z \equiv \phi(\Lambda) = \psi(\Omega)$, and hence,

$$I_*[\Lambda;\Omega] = \max_{Z\in\mathscr{Z}} \log|[\![Z]\!]|. \tag{4.2.19}$$

Since $\Lambda \perp \Theta$, then $\phi(\Lambda) \perp \Theta$ and subsequently $Z \perp \Theta$. Therefore, $\mathscr{Z} \subseteq \mathscr{Z}_\Theta$, and thus

$$\log|[\![Z]\!]| \leq \log|[\![Z_\Theta]\!]|. \tag{4.2.20}$$

By maximizing both LHS and RHS of (4.2.20) over $\mathscr{Z}$ and $\mathscr{Z}_\Theta$, we obtain

$$I_*[\Lambda;\Omega] \leq I_*[\Lambda;\Omega|\Theta]. \tag{4.2.21}$$

This concludes the proof of Lemma 2. □

Since $X^{(0)}(0:k-1)$ and $X^{(l)}(0)$ are unrelated for $l \in \{1,2\}$, it follows from Lemma 2

$$I_*[X^{(l)}(0);Q(0:k-1)|X^{(0)}(0:k-1)] \geq I_*[X^{(l)}(0);Q(0:k-1)]. \tag{4.2.22}$$

Note that $S^{(1)}(0:k-1) \leftrightarrow X^{(0)}(0:k-1) \leftrightarrow S^{(2)}(0:k-1)$, i.e., $S^{(1)}(0:k-1) \perp S^{(2)}(0:k-1)|X^{(0)}(0:k-1)$. Furthermore, the initial states $\{X^{(i)}(0)\}_{i=0}^2$ and additive noises $\{W^{(i)}(k),V^{(i)}(k)\}_{i=0}^2$ are mutually unrelated. Hence, the requirement **(A3)** results in the Markov chain $X^{(l)}(0) \leftrightarrow S^{(l)}(0:k-1) \leftrightarrow Q(0:k-1)|X^{(0)}(0:k-1)$, for $l \in \{1,2\}$. Thus, the *conditional data processing inequality* [2] yields

$$I_*[X^{(l)}(0);Q(0:k-1)|X^{(0)}(0:k-1)] \leq I_*[S^{(l)}(0:k-1);Q(0:k-1)|X^{(0)}(0:k-1)]. \tag{4.2.23}$$

By combining this lower bound on $I_*[S^{(l)}(0:k-1);Q(0:k-1)|X^{(0)}(0:k-1)]$ with inequalities (4.2.16) and (4.2.22), we obtain

$$\frac{I_*[S^{(l)}(0:k-1);Q(0:k-1)|X^{(0)}(0:k-1)]}{k} > \sum_{\ell=0}^{d_l} \log|\lambda_\ell^{(l)}|, \tag{4.2.24}$$

for $k \to \infty$. From (4.2.16) and (4.2.24), we conclude that

$$\frac{I_*[X^{(0)}(0:k-1);Q(0:k-1)]}{k} > \sum_{\ell=0}^{d_0} \log|\lambda_\ell^{(0)}|,$$
$$\frac{I_*[S^{(l)}(0:k-1);Q(0:k-1)|X^{(0)}(0:k-1)]}{k} > \sum_{\ell=0}^{d_l} \log|\lambda_\ell^{(l)}|,$$

for $l \in \{1,2\}$. This completes the proof of necessity.

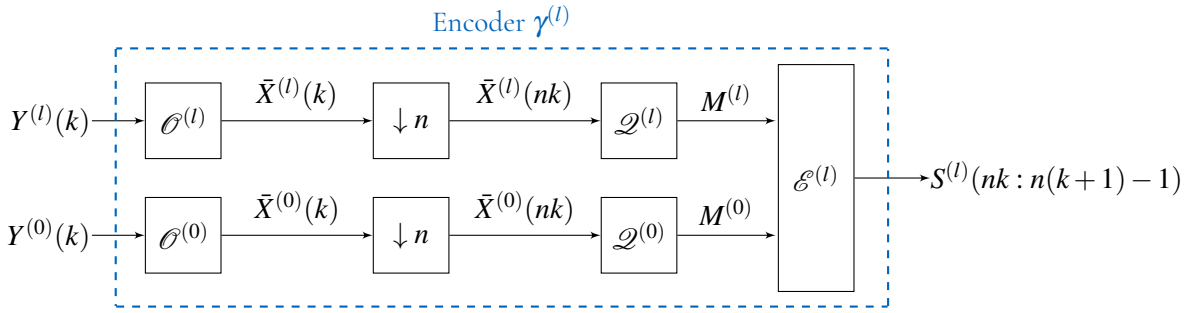


Figure 4.2: Structure of encoder $\gamma^{(l)}$ for $l \in \{1,2\}$. Note that the symbol "$\downarrow n$" denotes the downsampling block by a factor of n.

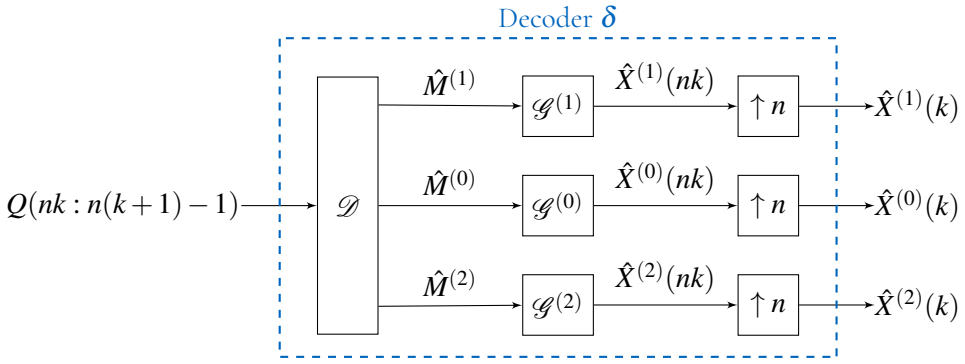


Figure 4.3: Structure of decoder δ. Note that the symbol "$\uparrow n$" denotes the upsampling block by a factor of n.

4.2.2 Sufficiency Proof

The sufficiency of (4.2.2) is now established using a separation structure between source and channel coding. To this end, we discuss in detail the structure of the encoder blocks $\gamma^{(1)}$ and $\gamma^{(2)}$. Firstly, a Luenberger observer $\mathscr{O}^{(i)}$, with $i \in \{0,1,2\}$ generates the signals $\bar{X}^{(i)}(k)$ at time instant k. The state observer $\mathscr{O}^{(i)}$ is defined by the following equation

$$\bar{X}^{(i)}(k+1) = A^{(i)}\bar{X}^{(i)}(k) + L\left(Y^{(i)}(k) - C^{(i)}\bar{X}^{(i)}(k)\right), \tag{4.2.25}$$

where L is a filter matrix of appropriate dimensions. The observability of $(A^{(i)}, C^{(i)})$ for $i \in \{0,1,2\}$ -Assumption **(A1)**- implies that there exists an observer (4.2.25) which guarantees asymptotic boundedness of estimation error. More details regarding linear state observers can be found in [39]. Note

that the process noise corresponds to the innovations fed to the state observer, i.e.,

$$\bar{V}^{(i)}(k) = L(Y^{(i)}(k) - C^{(i)}\bar{X}^{(i)}(k)). \tag{4.2.26}$$

The i-th observer's output $\bar{X}^{(i)}(k+1)$ are then down-sampled by a factor of n to yield

$$\bar{X}^{(i)}(n(k+1)) = (A^{(i)})^n \bar{X}^{(i)}(nk) + \Psi_{n-1}^{(i)}(k), \tag{4.2.27}$$

where it can be shown that the accumulated innovation noise term written as

$$\Psi_r^{(i)}(k) := \sum_{\xi=0}^{r} (A^{(i)})^{n-1-\xi} \bar{V}^{(i)}(nk+\xi),$$

is uniformly bounded $\forall r \in [0 : n-1]$ over $k \in \mathbb{Z}_{\geq 0}$ when an appropriate filter matrix L is used.
Next, each of the down-sampled sequences is processed by the respective adaptive quantizer $\mathscr{Q}^{(i)}$ to generate the messages M^i that are drawn from finite sets $\mathscr{M}^{(i)}$ with cardinalities $|\mathscr{M}^{(i)}|$ for $i \in \{0,1,2\}$, as depicted in Fig. 4.2. Finally, each pair of messages $\{(M^{(0)}, M^{(l)})\}_{l=1}^{2}$ is mapped into the codeword $S^{(l)}(nk : n(k+1)-1)$ with the block-length n by the corresponding channel encoder $\mathscr{E}^{(l)}$ at time instant k. The code rate for each message $M^{(i)}$ is then

$$R^{(i)} = (\log|\mathscr{M}^{(i)}|)/n, \text{ for } i \in \{0,1,2\}. \tag{4.2.28}$$

The receiver δ, shown in Fig. 4.3, is a three-stage process that consists of the reverse operations performed by the encoding block. Firstly, using the channel output $Q(nk : n(k+1)-1)$, the message estimates $\{\hat{M}^{(i)}\}_{i=0}^{2}$ are produced by means of an appropriate channel decoder. Each of these estimates is then processed by an *adaptive dequantizer* $\mathscr{G}^i$ followed by an upsampling operation by n to generate the state estimations $\hat{X}^{(i)}(k)$ at time instant k for $i \in \{0,1,2\}$.
As $\mathbf{int}(\mathscr{C}_0)$ is an open set (by definition) and non-empty (Lemma 1), there exists an open ℓ_∞-ball centered at h and with arbitrarily small radius $\delta > 0$ denoted as $\mathbf{B}_\delta(h) \subseteq \mathbf{int}(\mathscr{C}_0)$. Hence, for the given vector of topological entropies $h = (h^{(0)}, h^{(1)}, h^{(2)})^T \in \mathbf{int}(\mathscr{C}_0)$, there exists a rate vector $R = (R^{(0)}, R^{(1)}, R^{(2)})^T \in \mathbf{B}_\delta(h)$, i.e., for $i \in \{0,1,2\}$:

$$R^{(i)} = h^{(i)} + \delta. \tag{4.2.29}$$

Let $\mathbf{B}_\varepsilon(R)$ be the ℓ_∞-ball of center R and arbitrarily small radius $\varepsilon > 0$ as illustrated in Fig. 4.4. Since $R \in \mathscr{C}_0$, then there exists a rate vector $R_n \in \mathscr{C}_{0,n} \cap \mathbf{B}_\varepsilon(R)$ with $n \in \mathbb{Z}_{\geq 1}$, i.e.,

$$\|R - R_n\|_\infty \leq \varepsilon. \tag{4.2.30}$$

As ε can be arbitrarily chosen, we can select ε small enough such that $\varepsilon < \delta$; and hence, it is guar-

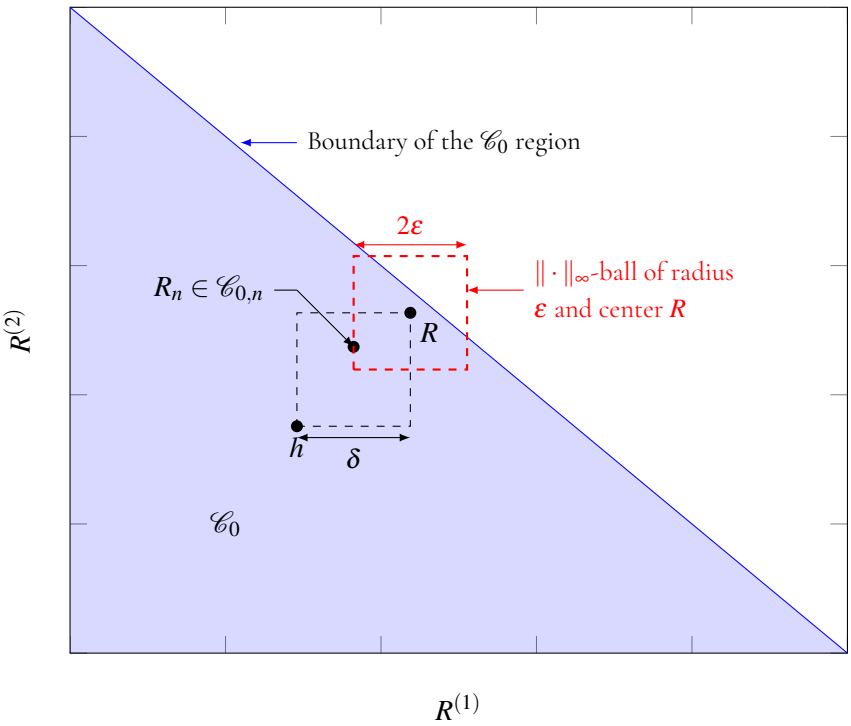


Figure 4.4: Illustrative figure of the zero-error capacity region's boundary in a two-dimensional space, i.e., $R^{(0)} = 0$. The region shadowed in blue corresponds to a part of the channel's [zoomed-in] zero-error capacity region $\mathscr{C}_0$.

anteed that there exists a zero-error code with rate vector R_n that is component-wise strictly larger than h.

Consequently, the communication channel linking each message $M^{(i)}$ to its estimate $\hat{M}^{(i)}$ can be modeled as noiseless. Hence, by the data rate theorem (Proposition 5.2 in [40]) $\forall R_n^{(i)} > h^{(i)}$, $\exists \mathscr{Q}^{(i)}, \mathscr{G}^{(i)}$ such that the prediction error $E^{(i)}(kn) = X^{(i)}(kn) - \hat{X}^{(i)}(kn)$ is uniformly bounded.

Now, every time instant $t \in \mathbb{Z}_{\geq 1}$ can be written for some nonnegative integer k as $t = kn + r$, where $r \in [0 : n-1]$. Furthermore, consider the estimator

$$\hat{X}^{(i)}(t) := (A^{(i)})^r \hat{X}^{(i)}(nk), \text{ for } i \in \{0,1,2\}. \tag{4.2.31}$$

We examine the requirement (4.1.6) for the resulting estimation error $E^{(i)}(t) = X^{(i)}(t) - \hat{X}^{(i)}(t)$, i.e.,

$$\begin{aligned} \sup \|E^{(i)}(t)\| &= \sup \|(A^{(i)})^r X^{(i)}(nk) + \Psi_r^{(i)}(k) - (A^{(i)})^r \hat{X}^{(i)}(nk)\| \\ &\leq \|(A^{(i)})^r\| \sup \|X^{(i)}(nk) - \hat{X}^{(i)}(nk)\| + \|\Psi_r^{(i)}(k)\| \\ &\leq \max_{r \in [0:n-1]} \left\{ \|(A^{(i)})^r\| \right\} \sup \|E^{(i)}(nk)\| + \|\Psi_r^{(i)}(k)\|, \end{aligned} \tag{4.2.32}$$

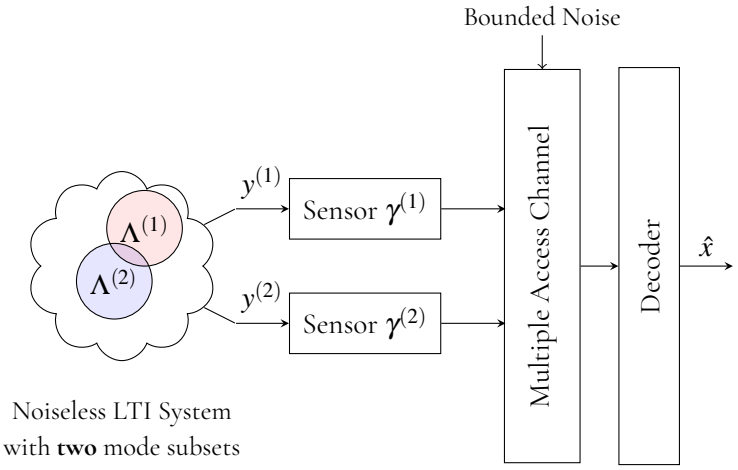


Figure 4.5: The states of a noiseless (process and measurement) LTI system are detected by two distinct sensors. Each sensor incorporates an observer and a channel encoder. In this scenario, we consider uncoordinated access strategies with no apriori resource allocation modeled as a MAC.

where the prediction error $E^{(i)}(nk)$ of the down-sampled system was shown to be uniformly bounded for some coder-estimator tuple, and the accumulated process noise term $\Psi_r^{(i)}(k)$ does also satisfy this condition. The RHS of (4.2.32) is therefore uniformly bounded over $k \in \mathbb{Z}_{\geq 0}$. This completes the sufficiency proof.

4.3 Generality of the Proposed Setup for Noiseless Linear Systems

The dynamic decoupling between the three subsystems in (4.1.1a)-(4.1.1b) may seem like a strict requirement. In this subsection, we show that any noiseless linear plant observed via two sensors can be put into this form, under a mild assumption on the output matrices. Subsystems 1 and 2 represent the plant modes that are observable only at sensors 1 and 2 respectively, while subsystem 0 represents the modes that are observable at each of the two sensors.

Consider a noiseless LTI system with a dynamic matrix $\tilde{A} \in \mathbb{R}^{d\times d}$ that is observed by two different sensors $\gamma^{(1)}$ and $\gamma^{(2)}$ as depicted in Fig. 4.5. The system is described by

$$\tilde{X}(k+1) = \tilde{A}\tilde{X}(k), \tag{4.3.1a}$$

$$Y^{(l)}(k) = \tilde{C}^{(l)}\tilde{X}(k), \text{ for } l \in \{1,2\}, \tag{4.3.1b}$$

where the joint system $((\tilde{C}^{(1)},\tilde{C}^{(2)}),\tilde{A})$ is observable. In order to decouple the states of the observed system, it is possible to put the matrix $\tilde{A}$ into *Jordan canonical form* [41]. Note that we call *system modes* to refer to the states associated with a Jordan block. Supposing that $\tilde{A}$ has p real eigenvalues

$\{\lambda_k\}_{k=1}^{p}$ and $(n-p)/2$ conjugate complex eigenvalues $\{\lambda_k = \alpha_k + j\beta_k\ (\lambda_k^* = \alpha_k - j\beta_k)\}_{k=p+1}^{m}$ with $m = p + (n-p)/2$, then there exists a transformation matrix T such that

$$T = (v_1, \cdots, v_p, \Re(v_{p+1}), \Im(v_{p+1}), \cdots, \Re(v_m), \Im(v_m)), \tag{4.3.2}$$

where v_k refer to the eigenvector corresponding to the eigenvalue λ_k. Using the matrix T, (4.3.1a)-(4.3.1b) become

$$X(k+1) = \underbrace{T^{-1}\tilde{A}T}_{A} X(k), \tag{4.3.3a}$$

$$Y^{(l)}(k) = \underbrace{\tilde{C}^{(l)}T}_{C^{(l)}} X(k), \text{ for } l \in \{1,2\} \tag{4.3.3b}$$

with the initial state $X(0) = T^{-1}\tilde{X}(0)$. The dynamic matrix A has then the following form

$$A = T^{-1}\tilde{A}T = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_\Gamma \end{pmatrix}, \tag{4.3.4}$$

with Γ *Jordan blocks* such that the i-th block can be written as

$$J_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ 0 & & & \lambda_i \end{pmatrix}, \tag{4.3.5}$$

for real eigenvalues, and

$$J_i = \begin{pmatrix} W_i & I_2 & & 0 \\ & \ddots & \ddots & \\ & & W_i & I_2 \\ 0 & & & W_i \end{pmatrix}, \ W_i = \begin{pmatrix} \alpha_i & \beta_i \\ -\beta_i & \alpha_i \end{pmatrix}, \ I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.3.6}$$

for complex conjugate eigenvalues $\alpha_i \pm j\beta_i$. For simplicity we suppose that each of the blocks $\{J_i\}_{i=1}^{\Gamma}$ admits distinct eigenvalues. Note that the output matrix $C^{(l)}$ can be expressed as a concatenation of sub-matrices $C_i^{(l)}$ with $i \in [1:\Gamma]$

$$C^{(l)} = \begin{pmatrix} C_1^{(l)} & C_2^{(l)} & \cdots & C_\Gamma^{(l)} \end{pmatrix} \in \mathbb{R}^{b_l \times d}, \tag{4.3.7}$$

where each sub-matrix $C_i^{(l)}$ is a collection of column vectors such that

$$C_i^{(l)} = \begin{pmatrix} c_{1,i}^{(l)} & c_{2,i}^{(l)} & \cdots & c_{d_i,i}^{(l)} \end{pmatrix} \in \mathbb{R}^{b_l \times d_i}. \tag{4.3.8}$$

Lemma 3. *Without loss of generality, consider the state matrix A to be in Jordan form (4.3.4) and the output matrix $C^{(l)}$ (4.3.7) for $l \in \{1,2\}$. Γ denotes the number of Jordan blocks such that each block admits a distinct eigenvalue. Then the pair $(A, C^{(l)})$ is observable if and only if the leading column of the block $C_i^{(l)}$, namely $c_{1,i}^{(l)}$, is non-zero $\forall i \in [1:\Gamma]$.*

Proof. The following proof is based on Hautus test [39, p. 156] that states the following.

Theorem 6. *A system (A, C) is observable if and only if the matrix $\begin{pmatrix} A - \lambda I \\ C \end{pmatrix}$ has full column rank at every eigenvalue λ of A.*

Let $A \in \mathbb{R}^{d \times d}$ be the state matrix in the sense of (4.3.4) and $C^{(l)} \in \mathbb{R}^{b_l \times d}$ be the output matrix as defined in (4.3.7). For $s \in \mathbb{R}$, consider the matrix $\begin{pmatrix} A - sI \\ C^{(l)} \end{pmatrix}$ with the following form

$$\left(\begin{array}{cccc|cccc|c|cccc}
\lambda_1 - s & 1 & & & & & & & & & & & \\
 & \ddots & \ddots & & & & & & & & & & \\
 & & \lambda_1 - s & 1 & & & & & & & & & \\
 & & & \lambda_1 - s & & & & & & & & & \\
\hline
 & & & & \lambda_2 - s & 1 & & & & & & & \\
 & & & & & \ddots & \ddots & & & & & & \\
 & & & & & & \lambda_2 - s & 1 & & & & & \\
 & & & & & & & \lambda_2 - s & & & & & \\
\hline
 & & & & & & & & \ddots & & & & \\
\hline
 & & & & & & & & & \lambda_\Gamma - s & 1 & & \\
 & & & & & & & & & & \ddots & \ddots & \\
 & & & & & & & & & & & \lambda_\Gamma - s & 1 \\
 & & & & & & & & & & & & \lambda_\Gamma - s \\
\hline
c_{1,1}^{(l)} & c_{2,1}^{(l)} & \cdots & c_{d_{11},1}^{(l)} & c_{1,2}^{(l)} & c_{2,2}^{(l)} & \cdots & c_{d_{21},2}^{(l)} & \cdots & c_{1,\Gamma}^{(l)} & c_{2,\Gamma}^{(l)} & \cdots & c_{d_{\Gamma\ell_\Gamma},\Gamma}^{(l)}
\end{array}\right), \tag{4.3.9}$$

where the empty entries are zeros. In the following analysis, we focus on the first Jordan block without loss of generality. Hence, for $s = \lambda_1$, the matrix (4.3.9) consists of columns whose components are either

- All zeros except for the last b_l elements (the respective column of $C^{(l)}$). This is the leading column of the Jordan block associated with λ_1; or

- All zeros except for one non-zero element (either 1 or $\lambda_j - \lambda_1$ for $j \in [2:\Gamma]$) and the last $b^{(l)}$ elements (the respective column of $C^{(l)}$). This is the leading column of each of the Jordan blocks associated with λ_j for all $j \in [2:\Gamma]$; or

- All zeros except for two non-zero elements (1 and $\lambda_j - \lambda_1$ for $j \in [2:\Gamma]$) and the last $b^{(l)}$ elements (the respective column of $C^{(l)}$).

As the rank of (4.3.9) will not change by elementary row operations, by means of Gaussian elimination we transform it into the following form

$$\left(\begin{array}{cccc|cccc|c|cccc}
0 & 1 & & & & & & & & & & \\
 & \ddots & \ddots & & & & & & & & & \\
 & & 0 & 1 & & & & & & & & \\
 & & & 0 & & & & & & & & \\
\hline
 & & & & \lambda_2-\lambda_1 & 0 & & & & & & \\
 & & & & & \ddots & \ddots & & & & & \\
 & & & & & & \lambda_2-\lambda_1 & 0 & & & & \\
 & & & & & & & \lambda_2-\lambda_1 & & & & \\
\hline
 & & & & & & & & \ddots & & & \\
\hline
 & & & & & & & & & \lambda_\Gamma-\lambda_1 & 0 & \\
 & & & & & & & & & & \ddots & \ddots \\
 & & & & & & & & & & & \lambda_\Gamma-\lambda_1 \\
\hline
c_{1,1}^{(l)} & 0 & \cdots & 0 & c_{1,2}^{(l)} & 0 & \cdots & 0 & \cdots & c_{1,\Gamma}^{(l)} & 0 & \cdots & 0
\end{array}\right), \tag{4.3.10}$$

where $\mathbf{0}$ is a b_l-dimensional all-zero column vector. Clearly, the first column of (4.3.10) consists of d elements that are zeros followed by b_l entries corresponding to $c_{1,1}^{(l)}$. Therefore, to obtain a full-ranked matrix (4.3.10), the column $c_{1,1}^{(l)}$ must be non-zero. Using the same argument with the remaining $\Gamma - 1$ eigenvalues, Lemma 3 is established. □

We now assume that the matrix $C^{(l)}$ has the following structure: If there exists a non-zero coefficient $c_{\ell,i}^{(l)}$, $\ell \in [1:d_i]$, associated with a Jordan block J_i, then the leading column in $C_i^{(l)}$ must be non-zero, i.e., $c_{1,i}^{(l)} \neq 0$. For each of the systems $\{(A, C^{(l)})\}_{l=1}^{2}$, constrain the matrix A to the Jordan blocks that have non-zero contribution to the output $y^{(l)}$. This results w.l.o.g. in a reduced system $(A_*^{(l)}, C_*^{(l)})$ for $l \in \{1,2\}$. Based on the structure of $C^{(l)}$ (and subsequently $C_*^{(l)}$) as well as Lemma 3, both systems are observable and thus the modes associated with the Jordan blocks are reconstructable. To clarify this structure for the reader, an example is provided in the next page.

Without loss of generality, we express (4.3.3b) as

$$Y^{(l)}(k) = \begin{pmatrix} C^{(l)}_{\mathrm{O}^l} & \mathbf{0} \end{pmatrix} \begin{pmatrix} X_{\mathrm{O}^l}(k) \\ X_{\bar{\mathrm{O}}^l}(k) \end{pmatrix}, \text{ for } l \in \{1,2\} \tag{4.3.11}$$

where X_{O^l} and $X_{\bar{\mathrm{O}}^l}$ are the system states observable and unobservable at the l-th sensor, respectively. Let $\mathbb{X}_{\mathrm{O}^l}$ and $\mathbb{X}_{\bar{\mathrm{O}}^l}$ denote the subsets of observable and unobservable states at sensor $l \in \{1,2\}$. Then, by virtue of the discussion after Lemma 3, the system modes can be partitioned into three distinct classes:

- modes common to both encoders $\gamma^{(1)}$ and $\gamma^{(2)}$, i.e., $X^{(0)} = \mathbb{X}_{\mathrm{O}^1} \cap \mathbb{X}_{\mathrm{O}^2}$,
- modes observed *only* by $\gamma^{(1)}$, i.e., $X^{(1)} = \mathbb{X}_{\mathrm{O}^1} \cap \mathbb{X}_{\bar{\mathrm{O}}^2}$,
- modes observed *only* by $\gamma^{(2)}$, i.e., $X^{(2)} = \mathbb{X}_{\mathrm{O}^2} \cap \mathbb{X}_{\bar{\mathrm{O}}^1}$.

Example: In this example, we consider a system matrix A with two Jordan blocks as follows

$$A = \begin{pmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 1 \\ 0 & 0 & 0 & \lambda_2 \end{pmatrix} \in \mathbb{R}^{4\times 4}, \tag{4.3.12}$$

and two output matrices such that $C^{(1)}$ satisfies the assumed structure and $C^{(2)}$ violates it,

$$C^{(1)} = \begin{pmatrix} c^{(1)}_{1,1} & c^{(1)}_{2,1} & 0 & 0 \end{pmatrix} \in \mathbb{R}^{1\times 4}, \tag{4.3.13}$$

$$C^{(2)} = \begin{pmatrix} 0 & c^{2}_{2,1} & 0 & c^{2}_{2,2} \end{pmatrix} \in \mathbb{R}^{1\times 4}, \tag{4.3.14}$$

with $c^{(1)}_{1,1}, c^{(1)}_{2,1}, c^{(2)}_{2,1}$ and $c^{(2)}_{2,2} \neq 0$. Then, the reduced system equations for $l \in \{1,2\}$ are respectively

$$A^{(1)}_{*} = \begin{pmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{pmatrix} \in \mathbb{R}^{2\times 2}, \tag{4.3.15}$$

$$C^{(1)}_{*} = \begin{pmatrix} c^{(1)}_{1,1} & 0 \end{pmatrix} \in \mathbb{R}^{1\times 2}. \tag{4.3.16}$$

and

$$A^{(2)}_{*} = \begin{pmatrix} \lambda_2 & 1 \\ 0 & \lambda_2 \end{pmatrix} \in \mathbb{R}^{2\times 2}, \tag{4.3.17}$$

$$C^{(2)}_{*} = \begin{pmatrix} 0 & c^{(2)}_{2,2} \end{pmatrix} \in \mathbb{R}^{1\times 2}. \tag{4.3.18}$$

According to Lemma 3, the modes (x_1, x_2) fall into the observability space of Sensor 1, whereas (x_3, x_4) are unobservable by both Sensor 1 and 2.

In order to estimate the different states, a dead-beat observer [42] is incorporated at each sensor. By collecting the measurements $(Y^{(l)}(n(k-1)), \cdots, Y^{(l)}(nk))$ for any $k \in \mathbb{Z}_{\geq 1}$ over a period of time n, the l-th sensor recovers its observable states and generates the following measurements

$$\bar{Y}^{(l)}(nk) = X_{\mathbf{O}^l}(nk), \text{ for } l \in \{1,2\}. \tag{4.3.19}$$

Using the mode classes $\boldsymbol{X}^{(0)}$, $\boldsymbol{X}^{(1)}$, and $\boldsymbol{X}^{(2)}$, it is hence possible to decompose the original system into three subsystems with dynamic matrices $\boldsymbol{A}^{(0)}$, $\boldsymbol{A}^{(1)}$ and $\boldsymbol{A}^{(2)}$ and reform (4.3.3a)-(4.3.3b) to obtain the setup depicted in Fig. 4.1 characterized by (4.1.1a)-(4.1.1b).

Hence, this formulation captures some of the essential elements of the problem of distributed state estimation, e.g., where each sensor observes a different subset of the overall system's dynamical modes, although it is limited to three systems.

4.4 Numerical Example: The Binary Adder Channel (BAC)

4.4.1 Zero-Error Capacity Region of the BAC

A well-known example of a MAC in the literature is the *binary adder channel* (BAC). This channel model consists of two independent users communicating with one receiver via a common discrete memoryless channel such that at each time instant $k \in \mathbb{Z}_{\geq 0}$

$$Y_k = X_k^{(1)} + X_k^{(2)} \in \{0,1,2\}, \tag{4.4.1}$$

where $X_k^{(1)}$, $X_k^{(2)} \in \{0,1\}$ are the inputs generated by users 1 and 2 respectively, and Y_k is the corresponding channel output. The i-th user selects a particular *codeword* $x_{1:n}^{(i)}$ from a finite set $\mathscr{X}^{(i)} \subseteq \{0,1\}^n$, where $n \in \mathbb{Z}_{\geq 1}$ is the code block-length. At the receiver side, zero-error decoding is achieved when each element of the sumset

$$\mathscr{Y} = \left\{ x_{1:n}^{(1)} + x_{1:n}^{(2)} : \forall x_{1:n}^{(1)} \in \mathscr{X}^{(1)},\, x_{1:n}^{(2)} \in \mathscr{X}^{(2)} \right\} \tag{4.4.2}$$

appears exactly once, i.e., one-to-one correspondence, and hence unambiguous decoding is always possible. Note that the addition here is carried out over $\mathbb{Z}^n$. At first glance, the problem might seem simple, nevertheless no closed form of the BAC zero-error capacity region $\mathscr{C}_0$ has been found yet. This has led to the emergence of two streams of research: a first group of researchers investigating the *outer bound* of this region and a second one focusing on improving the known *inner bound*. In regards to the outer bound, it has been always clear that $\mathscr{C}_0$ lies inside the ordinary Shannon capacity region which was determined by Ahlswede [19] and Liao [20] as the closure of the convex hull of nonnegative

rate pairs (R^1, R^2) that satisfy

$$\begin{cases} R^{(i)} & \leq 1, \ \text{ for } i \in \{1,2\} \\ R^{(1)} + R^{(2)} & \leq \frac{3}{2}. \end{cases}$$

The most recent upper bound was obtained by Austrin *et al.* in [21]. In their paper, the authors showed that for $R^{(1)} > 1 - \varepsilon$ such that ε is sufficiently small, we have $R^{(2)} < 0.4228$. Nonetheless, this bound quickly converges to $(1-\varepsilon) + R^{(2)} \leq 1.5$ as ε increases, and hence, it is still unclear whether zero-error transmission over BAC at rate $(R^{(1)}, R^{(2)})$ such that $R^{(1)} + R^{(2)} = 1.5$ is possible or not. On the other hand, the current state of the art in terms of lower bound on $\mathscr{C}_0$ was obtained by Mattas and Östergård [23] who found a zero-error code such that the sum rate is

$$R^{(1)} + R^{(2)} = \frac{\log_2(240)}{6} \approx 1.3178 \text{ bits}. \tag{4.4.3}$$

The zero-error code achieving this sum rate is used in the numerical example of the next section.

4.4.2 Application to the State Estimation Problem (Theorem 5)

To appreciate the result established in Theorem 5, we explore a setup where the states of two independent dynamical systems with given topological vector of entropies $h = (h^{(1)}, h^{(2)})^T$ are measured, encoded and transmitted by users $\mathscr{E}^{(1)}$ and $\mathscr{E}^{(2)}$ through a BAC (4.4.1) and eventually recovered at the receiver end.

Simulation Details. We implement two scalar LTI systems described by the following equations

$$x^{(i)}(k+1) = a^{(i)} x^{(i)}(k) + v^{(i)}(k), \tag{4.4.4a}$$

$$y^{(i)}(k) = x^{(i)}(k), \ \ i \in \{1,2\}, \tag{4.4.4b}$$

where $a^{(i)} = 2^{h^{(i)}}$ and $v^{(i)}(k)$ denotes the process noise with range $[-1,1]$. At time instant $k = 0$, the systems' initial states $\{x^{(i)}(0)\}_{i=1}^2$ and noise term $\{v^{(i)}(0)\}_{i=1}^2$ are randomly chosen from $[-1,1]$. At each sampling time nk the encoder produces an estimate $\hat{x}^{(i)}(nk)$ of the true state $x^{(i)}(nk)$. Then, the error $e^{(i)}(nk) := y^{(i)}(nk) - \hat{x}^{(i)}(nk)$ is processed by means of an adaptive uniform quantizer $\mathscr{Q}^{(i)}$ as follows. Firstly, the interval $[-1,1]$ is partitioned into K_i sub-intervals of equal size whose respective midpoints are denoted by $\omega(1), \cdots, \omega(K_i)$. Next, the error $e^{(i)}(nk)$ is scaled and quantized via

$$e_q^{(i)}(nk) := \mathscr{Q}^{(i)}\left(e^{(i)}(nk)/\ell_k^{(i)}\right) \in \{\omega(1), \cdots, \omega(K_i)\}. \tag{4.4.5}$$

Both encoder and decoder agree on the initial value of the scaling factor $\ell_0^{(i)}$ and update it at each instant nk as follows

$$\ell_{k+1}^{(i)} := \frac{(a^{(i)})^n}{K_i}\ell_k^{(i)} + \Psi_{\max}^{(i)}, \tag{4.4.6}$$

where the term $\Psi_{\max}^{(i)} = \sum_{\xi=0}^{n-1}(a^{(i)})^{n-1-\xi}$ forms an upper bound on the accumulated noise $\Psi_{n-1}^{(i)}(k)$ during a time window of duration n. The quantized error $e_q^{(i)}(nk)$ is then encoded by $\mathscr{E}^{(i)}$ which employs a zero-error codebook with block-length n and a rate $R^{(i)} = \log_2(K_i)/n$. The decoder $\mathscr{D}$ uses the received codewords to generate the $e_q^{(i)}(nk)$ with exactly zero decoding errors. Finally, the state estimate $\hat{x}^{(i)}(n(k+1))$ is updated by means of the following rule

$$\hat{x}^{(i)}(n(k+1)) = (a^{(i)})^n\left(\hat{x}^{(i)}(nk) + \ell_k^{(i)}e_q^{(i)}(nk)\right). \tag{4.4.7}$$

Note that the estimate $\hat{x}^{(i)}(k')$ for time instants $k' \in (nk, n(k+1))$ is computed via

$$\hat{x}^{(i)}(k') = (a^{(i)})^{k'-nk}\hat{x}^{(i)}(nk). \tag{4.4.8}$$

Fig. 4.6 illustrates the described construction above.

Numerical Results & Discussion. The system (4.4.4a)-(4.4.4b) is simulated for different parameter settings as outlined in Table 4.1. Fig. 4.7 depicts a realization of the systems' unstable states, i.e., $X^{(1)} = x^{(1)}$ and $X^{(2)} = x^{(2)}$, on both linear and logarithmic scales. Zero-error communication over the BAC is accomplished by means of the following codebooks [23] of block-length $n = 6$:

$$\begin{aligned}\mathscr{X}^{(1)} &= \{0,3,6,15,17,27,36,46,48,57,60,63\},\\ \mathscr{X}^{(2)} &= \{8,9,10,13,21,22,26,28,29,30,33,34,35,37,41,42,50,53,54,55\},\end{aligned}$$

where $\mathscr{X}^{(1)}$ and $\mathscr{X}^{(2)}$ are expressed in the decimal basis. In this experiment, 10^5 realizations of system (4.4.4a)-(4.4.4b) with different initial conditions and noise values uniformly chosen from the range $[-1,1]$ are simulated. At each time step k, the estimation error with the maximum norm over all realizations is selected and plotted in Fig. 4.8-4.9 for both scenarios 1 and 2.

When the vector of topological entropies h lies within the zero-error capacity region of the BAC, it is possible to find code rates such that $R^{(i)} > h^{(i)}$ for $i \in \{1,2\}$. Thus, the estimation error is bounded in accordance with Theorem 5 as exhibited in Fig. 4.8. Scenario 2 on the other hand illustrates the case where $h \notin \mathbf{int}(\mathscr{C}_0)$, and hence, it becomes impossible to reconstruct the state estimates at the receiver with zero error. One can see in Fig. 4.9 that the estimation error grows exponentially with time.

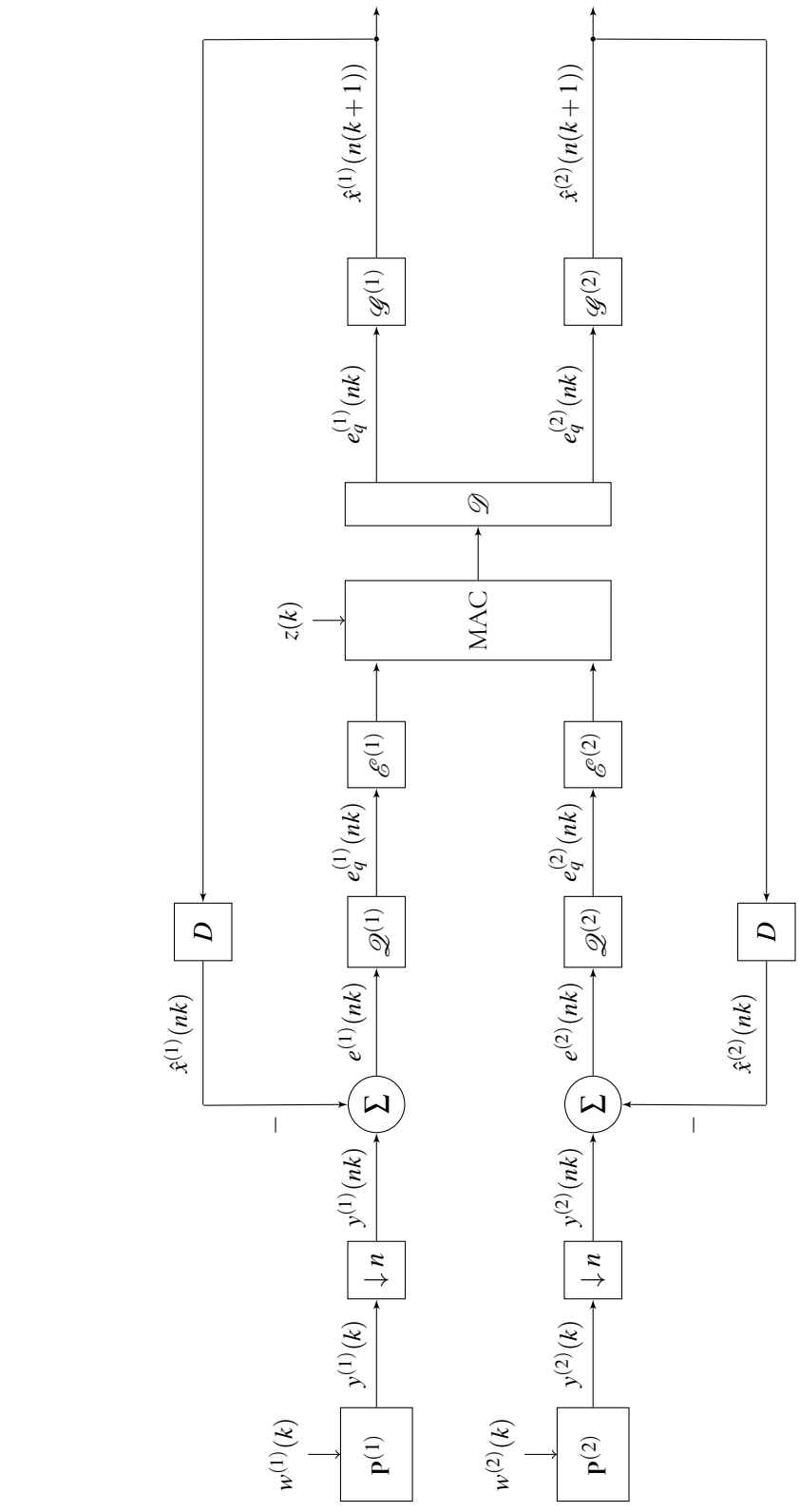


Figure 4.6: State estimation of LTI scalar systems with disturbances over two-input MAC using zero-error communication scheme as described in Section 5.4.1. For $i \in \{1,2\}$, the blocks $\mathscr{Q}^{(i)}$ and $\mathscr{G}^{(i)}$ denote the adaptive quantizers (5.4.4) and dequantizers (5.4.7), respectively. Additionally, D is a delay block. It is worth noting that the initial estimate $\hat{x}^{(i)}(0)$ is randomly selected from some bounded range, and that the quantizer and dequantizer both know the corresponding plant's dynamics $a^{(i)}$ as well as the process noise range $[\![V^{(i)}]\!]$.

Scenario Nº	1	2
$h^{(1)}$	0.15	0.85
$R^{(1)}$	0.597	0.597
$h^{(2)}$	0.18	0.95
$R^{(2)}$	0.720	0.720
$[\![V^{(i)}]\!]_{i=1,2}$	$[-1,1]$	$[-1,1]$

Table 4.1: Simulation Parameters. *The code block-length is set to* $n = 6$.

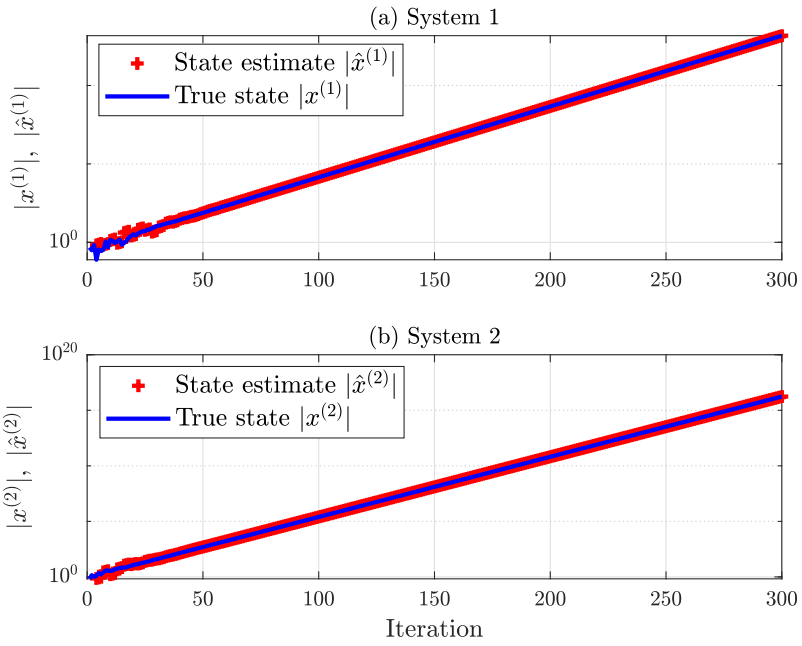
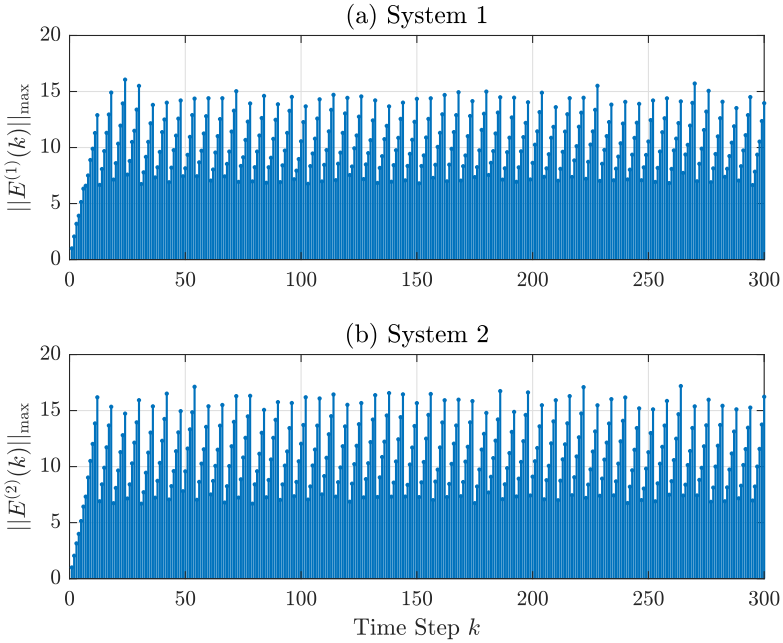
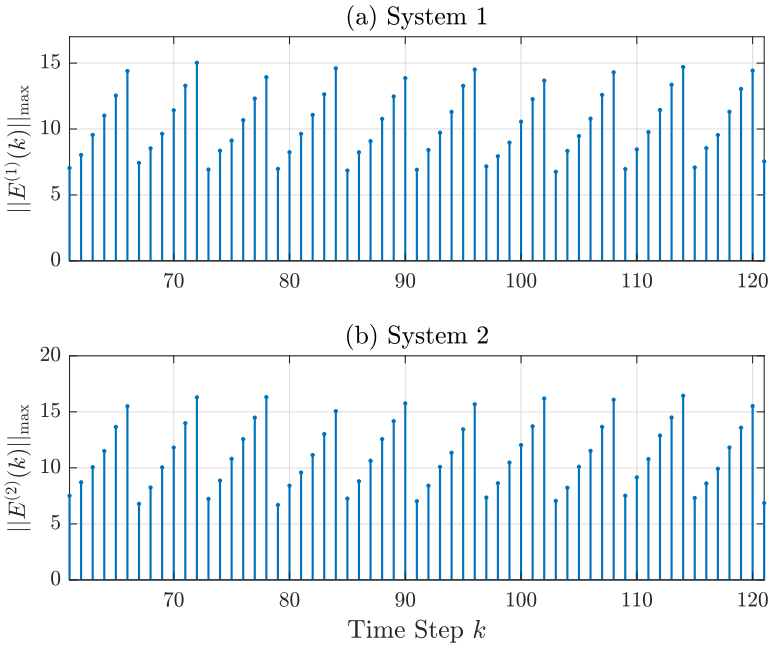


Figure 4.7: Example of state realizations $x^{(1)}, x^{(2)}$ and their corresponding estimations $\hat{x}^{(1)}, \hat{x}^{(2)}$ for unstable systems 1 and 2 on a *logarithmic* scale.



(a) Empirical maximum error norms on a *linear* scale.



(b) Zoomed version of (a).

Figure 4.8: Scenario 1. The topological entropy vector $h \in \mathrm{int}(\mathscr{C}_0)$, and the code rates $R^{(1)} > h^{(1)}$ and $R^{(2)} > h^{(2)}$.

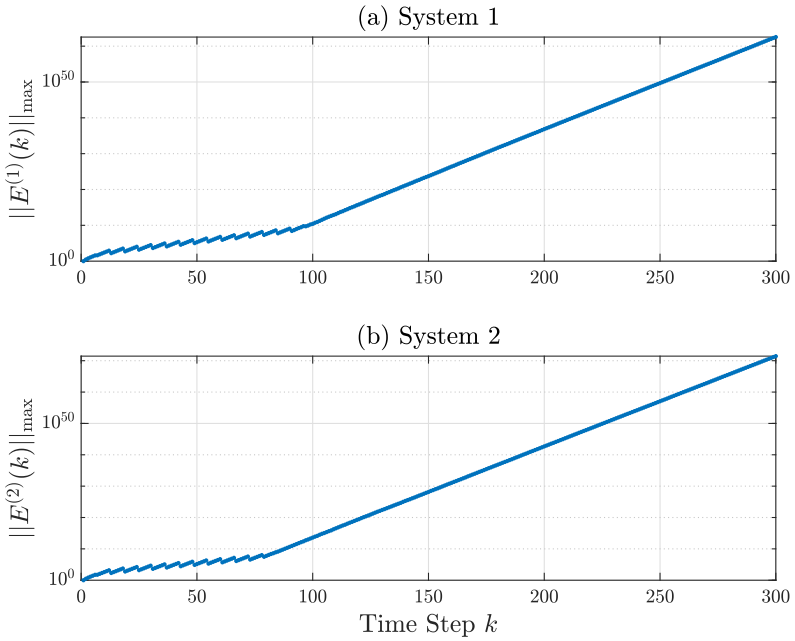


Figure 4.9: Empirical maximum error norms on a *logarithmic* scale for Scenario 2. The topological entropy vector $h \notin \mathrm{int}(\mathscr{C}_0)$, and the zero-error code rates $R^{(1)} < h^{(1)}$ and $R^{(2)} < h^{(2)}$.

4.5 Summary

We studied, in this chapter, the problem of distributed state estimation across a two-input single-output MAC with bounded noise. Using the notions of nonstochastic information as well as nonstochastic conditional information, tight necessary and sufficient conditions to ensure uniformly bounded errors at the receiver end were proven. It was shown that in order to achieve this goal, the vector of the plants' topological entropies must lie within the zero-error capacity region of the communication link. This result establishes a connection between the intrinsic properties of the linear systems and the channel characteristics. Afterwards, we have demonstrated that the considered setup is indeed general in the case of one-dimensional dynamical systems. Finally, the example of state estimation of scalar plants communicating over a two-input BAC was thoroughly discussed and corresponding simulation results were presented.

"Knowledge rests not upon truth alone, but upon error also."

— Carl Jung (1875–1961)

5

Universal Lower Bounds on the State Estimation Errors

After having established the necessary and sufficient conditions to achieve uniformly bounded estimation errors over two-input single-output MAC in Chapter 4, our aim now is to characterize the fundamental trade-off between the communication data rate, code-length, system dynamics and the state estimation performance. The performance of the coder-estimator is quantified by the asymptotic worst-case estimation error max norm. Using volume-based analysis and the well-known Brunn-Minkowski inequality [35, pg. 675], we derive universal lower bounds on the performance of any applicable estimation policy in Theorem 7. As this result is obtained on the overall performance of the entire system by treating the communication channel as a point-to-point one, we call it a *centralized* lower bound. We then refine this lower bound and derive *decentralized* lower bounds in Theorem 9, which capture the distributed nature of the considered setup. It is subsequently shown that these bounds are indeed tight by constructing an encoder-estimator tuple for scalar systems for which the lower bound is achieved. Next, the centralized lower bound is generalized to a system with an arbitrary state dimension (Lemma 4). The obtained lower bound connects between the code block-length, average entropy per dimension, the available rate and the volume of noise. This result is then analyzed for asymptotically large code block-lengths, and we show in Lemma 5 that the estimation performance in this regime is driven by the worst-case system eigenvalue.

5.1 Problem Formulation

The state estimation problem studied here is that of Chapter 4 and illustrated in Fig. 4.1. In fact, recall that for the i-th plant $\mathbf{P}^{(i)}$, with $i \in \{0,1,2\}$, at time instant nk, where $n \in \mathbb{Z}_{\geq 1}$ is the

code block-length and $k \in \mathbb{Z}_{\geq 0}$, the estimator generates

$$\hat{X}^{(i)}(nk) = \delta^{(i)}(nk, Q(0:nk)), \tag{5.1.1}$$

where $Q(0:nk)$ is the channel output sequence.
The estimation error of the entire system $(\mathbf{P}^{(0)}, \mathbf{P}^{(1)}, \mathbf{P}^{(2)})$ at sampled time instant nk is defined as

$$E(nk) := \begin{pmatrix} E^{(0)}(nk) \\ E^{(1)}(nk) \\ E^{(2)}(nk) \end{pmatrix} = \begin{pmatrix} X^{(0)}(nk) - \hat{X}^{(0)}(nk) \\ X^{(1)}(nk) - \hat{X}^{(1)}(nk) \\ X^{(2)}(nk) - \hat{X}^{(2)}(nk) \end{pmatrix} = X(nk) - \hat{X}(nk) \in \mathbb{R}^d, \tag{5.1.2}$$

where d refers to the entire system's state dimension, i.e., $d = \sum_{i=0}^{2} d^{(i)}$. Recall also that, at time instant $n(k+1)$, the state $X^{(i)}$ can be written as

$$X^{(i)}(n(k+1)) = \left(A^{(i)}\right)^n X^{(i)}(nk) + \sum_{\xi=0}^{n-1} (A^{(i)})^{n-1-\xi} V^{(i)}(nk+\xi). \tag{5.1.3}$$

In the following analysis, we derive a lower bound on the max-norm of the downsampled prediction error $E(nk)$ in (5.1.2). Note that the system matrix $A := \mathrm{diag}(A^{(0)}, A^{(1)}, A^{(2)})$ and process noise term $V(t) = (V^{(0)}(t), V^{(1)}(t), V^{(2)}(t))^T$.
In this chapter, we use volume-based analysis to derive centralized and decentralized lower bounds on time-asymptotic error norm for any possible estimation policy. The qualifying terms *"centralized"* and *"decentralized"* allow to distinguish between two types of lower bounds obtained here, namely

- **Centralized bound:** is a bound on the max-norm of the estimation error achieved by the whole system, i.e., E as defined in (5.1.2); and
- **Decentralized bounds:** refer to the bounds obtained on the estimation errors associated with the individual subsystems $\mathbf{P}^{(i)}$ with $i \in \{0, 1, 2\}$.

5.2 Centralized Lower Bound

We start by deriving the centralized bounds that apply to any coder-estimator construction. To this end, we present an analysis of the evolution of the estimation error volume with time. Such approach of error analysis, namely the study of error volume's evolution over time, was also adopted in previous works both in a deterministic setting, see, e.g., [43], as well as stochastic setting, such as [44]. The difference here with the techniques used in [43] is that we exploit the non-stochastic information theoretic framework to formulate the key argument.

Theorem 7. *Consider the linear time-invariant systems in (4.1.1a)-(4.1.1b) with topological entropy sum* $H = \sum_{i=0}^{2} h^{(i)}$ *and whose outputs are encoded by (4.1.2) and estimated by (4.1.4) over the two-input single output MAC with bounded noise in (4.1.3). Then, for any coder-estimator tuple with data rate sum* $R = \sum_{i=0}^{2} R^{(i)}$*, the following bound holds at any sampled time instant* nk*:*

(i) If $R \leq H$*, then*

$$\limsup_{k\to\infty} [\![\|E(nk)\|]\!] = \infty. \tag{5.2.1}$$

(ii) Else if $R > H$*, then*

$$\limsup_{k\to\infty} [\![\|E(nk)\|]\!] \geq \frac{1}{2} \cdot \frac{2^{nH/d} - 1}{2^{H/d} - 1} \cdot \frac{\mathrm{vol}([\![V]\!])^{1/d}}{1 - 2^{-n(R-H)/d}}, \tag{5.2.2}$$

where d *is the sum of the subsystems' individual dimensions.*

Proof. It is evident that for any bounded measurable $[\![E]\!] \subset \mathbb{R}^d$, it holds

$$\mathrm{vol}([\![E]\!]) \leq (2 \cdot \sup[\![\|E\|]\!])^d. \tag{5.2.3}$$

Set $\eta_{nk} := \sup[\![\|E(nk)\|]\!]$, where $E(nk)$ is the state estimation error obtained at time instant nk, with $k \in \mathbb{Z}_{\geq 0}$ and $n \in \mathbb{Z}_{\geq 1}$. Then, it follows that

$$(2\eta_{nk})^d \geq \mathrm{vol}([\![E(nk)]\!]). \tag{5.2.4}$$

Observe that when the channel output sequence $Q(0:nk)$ takes particular values $q(0:nk) \in \mathscr{Q}^{nk+1}$, then we know that $[\![E(nk)|Q(0:nk) = q(0:nk)]\!] \subseteq [\![E(nk)]\!]$. Subsequently, we obtain

$$\underbrace{\max_{q(0:nk)\in\mathscr{Q}^{nk+1}} \mathrm{vol}([\![E(nk)|q(0:nk)]\!])^{1/d}}_{:=m(nk)} \leq 2\eta_{nk}, \tag{5.2.5}$$

with $m(nk)$ denoting the d-th root of the maximum error uncertainty volume. For the $(k+1)$-th

step, i.e., at time $n(k+1)$, we obtain

$$\begin{aligned} m(n(k+1)) &= \max_{q(0:n(k+1))} \mathrm{vol}(\llbracket E(n(k+1))|q(0:n(k+1))\rrbracket)^{1/d} \\ &= \max_{q(0:n(k+1))} \mathrm{vol}(\llbracket X(n(k+1)) - \hat{X}(n(k+1))|q(0:n(k+1))\rrbracket)^{1/d} \\ &\overset{(5.1.3)}{=} \max_{q(0:n(k+1))} \mathrm{vol}(\llbracket A^n X(nk) + \sum_{\xi=0}^{n-1} A^{n-1-\xi} V(nk+\xi) - \\ &\qquad \underbrace{\delta(n(k+1), q(0:n(k+1)))}_{\text{constant translation}} | q(0:n(k+1))\rrbracket)^{1/d} \\ &= \max_{q(0:n(k+1))} \mathrm{vol}(\llbracket A^n X(nk) + \sum_{\xi=0}^{n-1} A^{n-1-\xi} V(nk+\xi)|q(0:n(k+1))\rrbracket)^{1/d}, \end{aligned} \tag{5.2.6}$$

where (5.2.6) follows from the fact that the set volume is invariant under linear transformation. Next, since the process noise V does not depend on the channel output $q(0:k)$, then we re-write the RHS of (5.2.6) as

$$\begin{aligned} \llbracket A^n X(nk) + \sum_{\xi=0}^{n-1} A^{n-1-\xi} V(nk+\xi)|q(0:n(k+1))\rrbracket = \llbracket A^n X(nk)|q(0:n(k+1))\rrbracket + \\ \llbracket \sum_{\xi=0}^{n-1} A^{n-1-\xi} V(nk+\xi)\rrbracket. \end{aligned} \tag{5.2.7}$$

Before proceeding with the derivation, we now state the Brunn-Minkowski inequality [35, p. 675], as it plays a central role in the proof.

Theorem 8 (Brunn-Minkowski Inequality). *Let $n \in Z_{\geq 1}$ and μ be the Lebesgue measure on $\mathbb{R}^n$. Furthermore, consider two Lebesgue measurable sets $\mathscr{A}, \mathscr{B} \subset \mathbb{R}^n$, and define the set sum $\mathscr{A} + \mathscr{B} := \{a+b : a \in \mathscr{A},\ b \in \mathscr{B}\}$. Then, the following inequality holds*

$$\mu(\mathscr{A} + \mathscr{B})^{(1/n)} \geq \mu(\mathscr{A})^{(1/n)} + \mu(\mathscr{B})^{(1/n)}. \tag{5.2.8}$$

By using Theorem 8 and the fact that $\llbracket V(k)\rrbracket = \llbracket V\rrbracket$ at any time instant $k \in \mathbb{Z}_{\geq 0}$, we obtain the following lower bound on $m(n(k+1))$

$$\begin{aligned} m(n(k+1)) &\geq \max_{q(0:n(k+1))} \mathrm{vol}(\llbracket A^n X(nk)|q(0:n(k+1))\rrbracket)^{1/d} + \mathrm{vol}(\llbracket \sum_{\xi=0}^{n-1} A^{n-1-\xi} V(nk+\xi)\rrbracket)^{1/d} \\ &= |\det A|^{n/d} \max_{q(0:n(k+1))} \mathrm{vol}(\llbracket X(nk)|q(0:n(k+1))\rrbracket)^{1/d} + \mathrm{vol}(\llbracket \sum_{\xi=0}^{n-1} A^{n-1-\xi} V(nk+\xi)\rrbracket)^{1/d} \end{aligned}$$

$$
\begin{aligned}
&= |\det A|^{n/d} \max_{q(0:nk)} \max_{q(nk+1:n(k+1))} \mathrm{vol}(\llbracket X(nk)|q(0:n(k+1))\rrbracket)^{1/d} + \\
&\quad \mathrm{vol}(\llbracket \sum_{\xi=0}^{n-1} A^{n-1-\xi} V(nk+\xi)\rrbracket)^{1/d} \\
&\geq |\det A|^{n/d} \max_{q(0:nk)} \max_{q(nk+1:n(k+1))} \mathrm{vol}(\llbracket X(nk)|q(0:n(k+1))\rrbracket)^{1/d} + \\
&\quad \sum_{\xi=0}^{n-1} |\det A|^{(n-1-\xi)/d} \mathrm{vol}(\underbrace{\llbracket V(nk+\xi)\rrbracket}_{=\llbracket V\rrbracket})^{1/d} \qquad (5.2.9) \\
&= |\det A|^{n/d} \max_{q(0:nk)} \max_{q(nk+1:n(k+1))} \mathrm{vol}(\llbracket X(nk)|q(0:n(k+1))\rrbracket)^{1/d} + \\
&\quad \frac{|\det A|^{n/d}-1}{|\det A|^{1/d}-1} \mathrm{vol}(\llbracket V\rrbracket)^{1/d}, \qquad (5.2.10)
\end{aligned}
$$

where (5.2.9) follows from the application of the Brunn-Minkowski inequality (Theorem 8).

Now, note that since the channel output $\boldsymbol{Q}$ is a function of the current sensor measurements as well as channel noise (4.1.3), we have

$$
\llbracket X(nk)|q(0:nk)\rrbracket = \bigcup_{q(nk+1:n(k+1))\in\mathscr{Q}^n} \llbracket X(nk)|q(0:n(k+1))\rrbracket. \qquad (5.2.11)
$$

This yields the following

$$
\begin{aligned}
\mathrm{vol}(\llbracket X(nk)|q(0:nk)\rrbracket) &\overset{(5.2.11)}{=} \mathrm{vol}\left(\bigcup_{q(nk+1:n(k+1))\in\mathscr{Q}^n} \llbracket X(nk)|q(0:n(k+1))\rrbracket\right) \\
&\leq \sum_{q(nk+1:n(k+1))\in\mathscr{Q}^n} \mathrm{vol}(\llbracket X(nk)|q(0:n(k+1))\rrbracket) \qquad (5.2.12) \\
&\leq |\mathscr{Q}^n| \max_{q(nk+1:n(k+1))} \mathrm{vol}(\llbracket X(nk)|q(0:n(k+1))\rrbracket) \\
&\leq 2^{n(R^{(0)}+R^{(1)}+R^{(2)})} \max_{q(nk+1:n(k+1))} \mathrm{vol}(\llbracket X(nk)|q(0:n(k+1))\rrbracket), \qquad (5.2.13)
\end{aligned}
$$

where (5.2.12) is a result of the countable subadditivity property of measures on $\mathbb{R}^d$, e.g., see [45, p. 43] and (5.2.13) follows from the fact that $|\mathscr{Q}^n| \leq 2^{n(R^{(0)}+R^{(1)}+R^{(2)})}$. Note that the individual data rate $R^{(i)}$ here encodes the output of the respective subsystem $\mathbf{P}^{(i)}$ and is defined in bits/channel use as

$$
R^{(i)} := \log_2 |\mathscr{M}^{(i)}|/n, \ \forall i \in \{0,1,2\}. \qquad (5.2.14)
$$

with n the code block-length and $\mathscr{M}^{(i)}$ is the finite set of messages to which the plant output $\boldsymbol{Y}^{(i)}$ is mapped after undergoing the appropriate signal processing steps, i.e., downsampling and quantization.

Let $R = \sum_{i=0}^{2} R^{(i)}$ and substitute (5.2.13) in (5.2.10) to obtain

$$\begin{aligned} m(n(k+1)) &\geq 2^{-nR/d} |\det A|^{n/d} \max_{q(0:nk)} \operatorname{vol}(\llbracket X(nk) | q(0:nk) \rrbracket)^{1/d} + \frac{|\det A|^{n/d} - 1}{|\det A|^{1/d} - 1} \operatorname{vol}(\llbracket V \rrbracket)^{1/d} \\ &= 2^{-nR/d} |\det A|^{n/d} \max_{q(0:nk)} \operatorname{vol}(\llbracket X(nk) - \underbrace{\delta(nk, q(0:nk))}_{=\hat{X}(nk)} | q(0:nk) \rrbracket)^{1/d} + \\ &\quad \frac{|\det A|^{n/d} - 1}{|\det A|^{1/d} - 1} \operatorname{vol}(\llbracket V \rrbracket)^{1/d} \qquad (5.2.15) \\ &= 2^{-nR/d} |\det A|^{n/d} m(nk) + \frac{|\det A|^{n/d} - 1}{|\det A|^{1/d} - 1} \operatorname{vol}(\llbracket V \rrbracket)^{1/d} \\ &= 2^{-n(R-H)/d} m(nk) + \frac{2^{nH/d} - 1}{2^{H/d} - 1} \operatorname{vol}(\llbracket V \rrbracket)^{1/d}, \qquad (5.2.16) \end{aligned}$$

where (5.2.15) holds as the volume of a set remains invariant when constant translation is applied (which is, in this case, $\delta(nk, q(0:nk))$ given a particular realization $q(0:nk)$), and $H = \sum_{i=0}^{2} h^{(i)}$ is the sum of topological entropies. If $R \leq H$, it is clear then that $m(k) \to \infty$.
On the other hand, if $R > H$ then solving the recursive inequality (5.2.16) yields

$$m(nk) \geq \frac{2^{nH/d} - 1}{2^{H/d} - 1} \frac{\operatorname{vol}(\llbracket V \rrbracket)|^{1/d}}{1 - 2^{-n(R-H)/d}} + \underbrace{\left(m_0 - \frac{2^{nH/d} - 1}{2^{H/d} - 1} \frac{\operatorname{vol}(\llbracket V \rrbracket)^{1/d}}{1 - 2^{-(R-H)/d}} \right) 2^{-kn(R-H)/d}}_{\to 0 \text{ as } k \to \infty}. \qquad (5.2.17)$$

Using this in (5.2.5), we obtain

$$\limsup_{k \to \infty} \llbracket \| E(nk) \| \rrbracket \geq \frac{1}{2} \cdot \frac{2^{nH/d} - 1}{2^{H/d} - 1} \cdot \frac{\operatorname{vol}(\llbracket V \rrbracket)^{1/d}}{1 - 2^{-n(R-H)/d}}. \qquad (5.2.18)$$

This completes the proof of Theorem 7. □

5.3 Decentralized Lower Bounds

After having found a lower bound on the prediction error for the totality of the system, i.e., $(\mathbf{P}^{(0)}, \mathbf{P}^{(1)}, \mathbf{P}^{(2)})$, we proceed now to obtain similar performance measures on each of the individual subsystems. Observe that sensors γ^1 and γ^2 encode the measurements $(Y^{(0)}, Y^{(1)})$ and $(Y^{(0)}, Y^{(2)})$, respectively. Since all of the plants are mutually unrelated, we have

- At most $R^{(0)}$ bits/use to estimate $X^{(0)}$,
- At most $R^{(1)}$ bits/use to estimate $X^{(1)}$,
- At most $R^{(2)}$ bits/use to estimate $X^{(2)}$.

Theorem 9. *Consider the linear time-invariant systems in (4.1.1a)-(4.1.1b) with topological entropies* $\{h^{(i)}\}_{i=0}^{2}$ *and whose outputs are encoded by (4.1.2) and estimated by (4.1.4) over the two-input single output MAC with bounded noise in (4.1.3). Then, for any coder-estimator tuple with data rates* $\{R^{(i)}\}_{i=0}^{2}$*, the following bound holds at any sampled time instant* nk *and for* $i \in \{0,1,2\}$*:*

(i) If $R^{(i)} \leq h^{(i)}$*, then*

$$\limsup_{k\to\infty} [\![\|E^{(i)}(nk)\|]\!] = \infty. \tag{5.3.1}$$

(ii) Else if $R^{(i)} > h^{(i)}$*, then*

$$\limsup_{k\to\infty} [\![\|E^{(i)}(nk)\|]\!] \geq \frac{1}{2} \cdot \frac{2^{nh^{(i)}/d^{(i)}} - 1}{2^{h^{(i)}/d^{(i)}} - 1} \cdot \frac{\mathrm{vol}([\![V^{(i)}]\!])^{1/d^{(i)}}}{1 - 2^{-n(R^{(i)}-h^{(i)})/d^{(i)}}}, \tag{5.3.2}$$

where $d^{(i)}$ *is the state dimension of the* i*-th subsystem and* $E^{(i)}(nk)$ *is the estimation error associated with it, i.e., plant* $P^{(i)}$*, with* $i \in \{0,1,2\}$*.*

Corollary 1. *It follows from Theorem 9 that*

(i) If there exists at least one subsystem $P^{(i)}$ *with* $R^{(i)} \leq h^{(i)}$*, then*

$$\limsup_{k\to\infty} [\![\|E(nk)\|]\!] = \infty. \tag{5.3.3}$$

(ii) Else if, $\forall i \in \{0,1,2\}$*,* $R^{(i)} > h^{(i)}$*, then*

$$\limsup_{k\to\infty} [\![\|E(nk)\|]\!] \geq \max_{i\in\{0,1,2\}} \left\{ \frac{1}{2} \cdot \frac{2^{nh^{(i)}/d^{(i)}} - 1}{2^{h^{(i)}/d^{(i)}} - 1} \cdot \frac{\mathrm{vol}([\![V^{(i)}]\!])^{1/d^{(i)}}}{1 - 2^{-n(R^{(i)}-h^{(i)})/d^{(i)}}} \right\}. \tag{5.3.4}$$

Proof. Consider the estimation error of the states of $\mathbf{P}^{(1)}$, denoted by $E^{(1)}(nk)$ at the sampled time instant nk. We define $m^{(1)}(nk)$ as

$$\begin{aligned} m^{(1)}(nk) &:= \max_{q(0:nk)\in\mathcal{Q}^{nk+1}} \mathrm{vol}([\![E^{(1)}(nk) \mid q(0:nk),\, X^{(0)}(nk) = 0,\, X^{(2)}(nk) = 0]\!])^{1/d^{(1)}} \\ &\leq 2\eta_{nk}^{(1)}, \end{aligned} \tag{5.3.5}$$

where $\eta_{nk}^{(1)} := [\![\|E^{(1)}(nk)\|]\!]$. At time instant $n(k+1)$ the $d^{(1)}$-th root uncertainty volume $m^{(1)}(n(k+1))$ is

$$m^{(1)}(n(k+1)) = \max_{q(0:n(k+1))} \mathrm{vol}([\![E^{(1)}(n(k+1)) \mid q(0:n(k+1)),$$

$$
\begin{aligned}
&X^{(0)}(n(k+1))=0,\, X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}}\\
&= \max_{q(0:n(k+1))} \mathrm{vol}([\![X^{(1)}(n(k+1))-\hat{X}^{(1)}(n(k+1)) \mid q(0:n(k+1)),\\
&\qquad X^{(0)}(n(k+1))=0,\, X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}}\\
&\overset{(5.1.3)}{=} \max_{q(0:n(k+1))} \mathrm{vol}([\![(A^{(1)})^n X^{(1)}(nk)+\sum_{\xi=0}^{n-1}(A^{(1)})^{n-1-\xi}V^{(1)}(nk+\xi)-\\
&\qquad \underbrace{\delta^{(1)}(n(k+1),q(0:n(k+1)))}_{\text{constant translation}} \mid q(0:n(k+1)),\\
&\qquad X^{(0)}(0:n(k+1))=0,\, X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}}\\
&= \max_{q(0:n(k+1))} \mathrm{vol}([\![(A^{(1)})^n X^{(1)}(nk)+\sum_{\xi=0}^{n-1}(A^{(1)})^{n-1-\xi}V^{(1)}(nk+\xi) \mid\\
&\qquad q(0:n(k+1)),\, X^{(0)}(n(k+1))=0,\, X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}}\\
&\geq \max_{q(0:n(k+1))} \mathrm{vol}([\![(A^{(1)})^n X^{(1)}(nk) \mid q(0:n(k+1)),\, X^{(0)}(n(k+1))=0,\\
&\qquad X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}} \mathrm{vol}([\![\sum_{\xi=0}^{n-1}(A^{(1)})^{n-1-\xi}V^{(1)}(nk+\xi) \mid\\
&\qquad q(0:n(k+1)),\, X^{(0)}(n(k+1))=0,\, X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}}
\end{aligned}
\tag{5.3.6}
$$

$$
\begin{aligned}
&= \max_{q(0:n(k+1))} \mathrm{vol}([\![(A^{(1)})^n X^{(1)}(nk) \mid q(0:n(k+1)),\, X^{(0)}(n(k+1))=0,\\
&\qquad X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}}+\\
&\qquad \mathrm{vol}([\![\sum_{\xi=0}^{n-1}(A^{(1)})^{n-1-\xi}V^{(1)}(nk+\xi)]\!])^{1/d^{(1)}}\\
&\geq |\det A^{(1)}|^{n/d^{(1)}} \max_{q(0:n(k+1))} \mathrm{vol}([\![X^{(1)}(nk) \mid q(0:n(k+1)),\, X^{(0)}(n(k+1))=0,\\
&\qquad X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}}+\\
&\quad \sum_{\xi=0}^{n-1}|\det A^{(1)}|^{(n-1-\xi)/d^{(1)}} \mathrm{vol}(\underbrace{[\![V^{(1)}(nk+\xi)]\!]}_{=[\![V^{(1)}]\!]})^{1/d^{(1)}}\\
&= |\det A^{(1)}|^{n/d^{(1)}} \max_{q(0:n(k+1))} \mathrm{vol}([\![X^{(1)}(nk) \mid q(0:n(k+1)),\, X^{(0)}(n(k+1))=0,\\
&\qquad X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}}+\\
&\quad \frac{|\det A^{(1)}|^{n/d^{(1)}}-1}{|\det A^{(1)}|^{1/d^{(1)}}-1}\mathrm{vol}([\![V^{(1)}]\!])^{1/d^{(1)}}\\
&= |\det A^{(1)}|^{n/d^{(1)}} \max_{q(0:nk)} \max_{q(nk+1:n(k+1))} \mathrm{vol}([\![X^{(1)}(nk) \mid q(0:n(k+1)),\, X^{(0)}(n(k+1))=0,
\end{aligned}
$$

$$X^{(2)}(n(k+1))=0]\!])^{1/d^{(1)}}+$$
$$\frac{|\det A^{(1)}|^{n/d^{(1)}}-1}{|\det A^{(1)}|^{1/d^{(1)}}-1}\mathrm{vol}([\![V^{(1)}]\!])^{1/d^{(1)}}, \tag{5.3.7}$$

where (5.3.6) holds by virtue of Brunn-Minkowski inequality for volumes of set sums (Theorem 8). Similar to (5.2.11), we have

$$[\![X^{(1)}(nk)\mid q(0:nk),\,X^{(0)}(n(k+1))=0,\,X^{(2)}(n(k+1))=0]\!]=$$
$$\bigcup_{q(nk+1:n(k+1))\in\mathscr{Q}^n}[\![X^{(1)}(nk)\mid q(0:n(k+1)),\,X^{(0)}(n(k+1))=0,\,X^{(2)}(n(k+1))=0]\!]. \tag{5.3.8}$$

This yields

$$\begin{aligned}
&\mathrm{vol}([\![X^{(1)}(nk)\mid q(0:nk),\,X^{(0)}(n(k+1))=0,\,X^{(2)}(n(k+1))=0]\!])\\
&\le\sum_{q(nk+1:n(k+1))\in\mathscr{Q}^n}\mathrm{vol}([\![X^{(1)}(nk)\mid q(0:n(k+1)),\,X^{(0)}(n(k+1))=0,\,X^{(2)}(n(k+1))=0]\!])\\
&\le|\mathscr{Q}^n|\max_{q(nk+1:n(k+1))}\mathrm{vol}([\![X^{(1)}(nk)\mid q(0:n(k+1)),\,X^{(0)}(n(k+1))=0,\,X^{(2)}(n(k+1))=0]\!])\\
&\le 2^{nR^{(1)}}\max_{q(nk+1:n(k+1))}\mathrm{vol}([\![X^{(1)}(nk)\mid q(0:n(k+1)),\,X^{(0)}(n(k+1))=0,\,X^{(2)}(n(k+1))=0]\!]),
\end{aligned} \tag{5.3.9}$$

where (5.3.9) follows from the observation that the states of subsystems $\mathbf{P}^{(0)}$ and $\mathbf{P}^{(2)}$ are fixed to null, and hence, the number of channel output symbols is at most equal to the number of messages generated by $\mathbf{P}^{(1)}$, namely $2^{nR^{(1)}}$. By substituting (5.3.9) into (5.3.7), we obtain

$$\begin{aligned}
m^{(1)}(n(k+1))&\ge 2^{-nR^{(1)}/d^{(1)}}|\det A^{(1)}|^{n/d^{(1)}}\max_{q(0:nk)}\mathrm{vol}([\![X^{(1)}(nk)|q(0:nk)]\!])^{1/d^{(1)}}\\
&\quad+\frac{|\det A^{(1)}|^{n/d^{(1)}}-1}{|\det A^{(1)}|^{1/d^{(1)}}-1}\mathrm{vol}([\![V^{(1)}]\!])^{1/d^{(1)}}\\
&=2^{-nR^{(1)}/d^{(1)}}|\det A^{(1)}|^{n/d^{(1)}}\max_{q(0:nk)}\mathrm{vol}([\![X^{(1)}(nk)-\underbrace{\delta^{(1)}(nk,q(0:nk))}_{=\hat{X}^{(1)}(nk)}|q(0:nk)]\!])^{1/d^{(1)}}\\
&\quad+\frac{|\det A^{(1)}|^{n/d^{(1)}}-1}{|\det A^{(1)}|^{1/d^{(1)}}-1}\mathrm{vol}([\![V^{(1)}]\!])^{1/d^{(1)}}\\
&=2^{-nR^{(1)}/d^{(1)}}|\det A^{(1)}|^{n/d^{(1)}}m^{(1)}(nk)+\frac{|\det A^{(1)}|^{n/d^{(1)}}-1}{|\det A^{(1)}|^{1/d^{(1)}}-1}\mathrm{vol}([\![V^{(1)}]\!])^{1/d^{(1)}}\\
&=2^{-n(R^{(1)}-h^{(1)})/d^{(1)}}m^{(1)}(nk)+\frac{2^{nh^{(1)}/d^{(1)}}-1}{2^{h^{(1)}/d^{(1)}}-1}\mathrm{vol}([\![V^{(1)}]\!])^{1/d^{(1)}}.
\end{aligned} \tag{5.3.10}$$

Thus, if $R^{(1)} \leq h^{(1)}$, it follows that $m^{(1)}(nk) \to \infty$. On the other hand, if $R^{(1)} > h^{(1)}$ then we obtain

$$m^{(1)}(nk) \geq \frac{2^{nh^{(1)}/d^{(1)}}-1}{2^{h^{(1)}/d^{(1)}}-1}\frac{\mathrm{vol}(\llbracket V^{(1)} \rrbracket)^{1/d^{(1)}}}{1-2^{-n(R^{(1)}-h^{(1)})/d^{(1)}}} + \underbrace{\left(m_0 - \frac{2^{nh^{(1)}/d^{(1)}}-1}{2^{h^{(1)}/d^{(1)}}-1}\frac{\mathrm{vol}(\llbracket V^{(1)} \rrbracket)^{1/d^{(1)}}}{1-2^{-(R^{(1)}-h^{(1)})/d^{(1)}}}\right)2^{-nk(R^{(1)}-h^{(1)})/d^{(1)}}}_{\to 0 \text{ as } k\to\infty}. \tag{5.3.11}$$

Thus,

$$\limsup_{k\to\infty} \llbracket \|E^{(1)}(nk)\| \rrbracket \geq \frac{1}{2}\cdot\frac{2^{nh^{(1)}/d^{(1)}}-1}{2^{h^{(1)}/d^{(1)}}-1}\cdot\frac{\mathrm{vol}(\llbracket V^{(1)} \rrbracket)^{1/d^{(1)}}}{1-2^{-n(R^{(1)}-h^{(1)})/d^{(1)}}}. \tag{5.3.12}$$

In a similar manner, a lower bound on the max norm of the estimation error for subsystems $\mathbf{P}^{(0)}$ and $\mathbf{P}^{(2)}$ can be established. □

5.4 Tightness of Decentralized Lower Bounds for Scalar Systems

5.4.1 System Model

Consider three unrelated and fully observable LTI scalar systems, $\mathbf{P}^{(0)}$, $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$, described by the following system of equations

$$x^{(i)}(k+1) = a^{(i)}x^{(i)}(k) + v^{(i)}(k), \tag{5.4.1a}$$

$$y^{(i)}(k) = x^{(i)}(k), \text{ for } i \in \{0,1,2\},\ \forall k \in \mathbb{Z}_{\geq 0}, \tag{5.4.1b}$$

where $a^{(i)} = 2^{h^{(i)}}$ and $v^{(i)}(k)$ denotes the process noise with range $\llbracket V^{(i)} \rrbracket$. At each sampled time instant nk, the encoder produces an estimate $\hat{x}^{(i)}(nk)$ of the true state $x^{(i)}(nk)$. Then, the error $e^{(i)}(nk) := x^{(i)}(nk) - \hat{x}^{i}(nk)$ is processed by means of an adaptive uniform quantizer $\mathscr{Q}^{(i)}$ as follows. Initially, the interval $[-1,1]$ is partitioned into $K^{(i)}$ sub-intervals of equal size $2\ell_0^{(i)}$. Both encoder and decoder agree on the initial value of the scaling factor $\ell_0^{(i)}$, and update it recursively at instant $n(k+1)$ by

$$\ell_{k+1}^{(i)} := \frac{(a^{(i)})^n}{K^{(i)}}\ell_k^{(i)} + \frac{\Psi_{\max}^{(i)}}{2}, \tag{5.4.2}$$

where the term $\Psi_{\max}^{(i)} = \sum_{\xi=0}^{n-1} (a^{(i)})^{n-1-\xi} \mathrm{vol}(\llbracket V^{(i)} \rrbracket)$ forms an upper bound on the accumulated noise during a time window of duration n. Observe that

$$\Psi_{\max}^{(i)} = \frac{2^{nh^{(i)}} - 1}{2^{h^{(i)}} - 1} \mathrm{vol}(\llbracket V^{(i)} \rrbracket). \tag{5.4.3}$$

The error $e^{(i)}(nk)$ is then scaled and quantized as follows

$$e_q^{(i)}(nk) := \mathscr{Q}^{(i)} \left(e^i(nk) / \ell_k^{(i)} \right) \in \{\omega(1), \cdots, \omega(K^{(i)})\}. \tag{5.4.4}$$

Next, each pair of quantized errors $(e_q^{(0)}(nk), e_q^{(j)}(nk))$ is encoded by $\mathscr{E}^{(j)}$, with $j \in \{1,2\}$. The encoders employ zero-error codebooks with block-length n, and the code rates are defined as

$$R^{(i)} = \frac{\log_2 K^{(i)}}{n}, \ \forall i \in \{0,1,2\}. \tag{5.4.5}$$

We assume in this construction that $(h^{(0)}, h^{(1)}, h^{(2)})^T \in \mathscr{C}_0$ and $R^{(i)} > h^{(i)}$, with $\mathscr{C}_0$ being the zero-error capacity region of the MAC and $i \in \{0,1,2\}$. At the other end of the channel, the decoder $\mathscr{D}$ uses the received codewords to generate the estimated error $\hat{e}_q^{(i)}(nk)$ with exactly zero decoding errors. Hence, we have

$$\hat{e}_q^{(i)}(nk) = e_q^{(i)}(nk), \ \forall i \in \{0,1,2\}. \tag{5.4.6}$$

Finally, the state estimate $\hat{x}^{(i)}(n(k+1))$ is updated by means of the following rule:

$$\hat{x}^{(i)}(n(k+1)) = (a^{(i)})^n \left(\hat{x}^i(nk) + \ell_k^{(i)} e_q^{(i)}(nk) \right), \ \forall i \in \{0,1,2\}. \tag{5.4.7}$$

5.4.2 Proof of Tightness of Decentralized Lower Bounds

From (5.4.2) and (5.4.3), we write $\ell_k^{(i)}$ as

$$\begin{aligned} \ell_{k+1}^{(i)} &= \frac{(a^{(i)})^n}{K^{(i)}} \ell_k^{(i)} + \frac{2^{nh^{(i)}} - 1}{2^{h^{(i)}} - 1} \frac{\mathrm{vol}(\llbracket V^{(i)} \rrbracket)}{2} \\ &\overset{(5.4.5)}{=} \frac{2^{nh^{(i)}}}{2^{nR^{(i)}}} \ell_k^{(i)} + \frac{2^{nh^{(i)}} - 1}{2^{h^{(i)}} - 1} \frac{\mathrm{vol}(\llbracket V^{(i)} \rrbracket)}{2}. \end{aligned} \tag{5.4.8}$$

Solving the recursive equation (5.4.8) yields

$$\begin{aligned}\ell_k^{(i)} &= \left(2^{n(h^{(i)}-R^{(i)})}\right)^k \ell_0^{(i)} + \sum_{\xi=0}^{k-1} \left(2^{n(h^{(i)}-R^{(i)})}\right)^{\xi} \frac{2^{nh^{(i)}}-1}{2^{h^{(i)}}-1} \frac{\mathrm{vol}(\llbracket V^{(i)} \rrbracket)}{2} \\ &= 2^{nk(h^{(i)}-R^{(i)})} \ell_0^{(i)} + \frac{1-2^{nk(h^{(i)}-R^{(i)})}}{1-2^{n(h^{(i)}-R^{(i)})}} \frac{2^{nh^{(i)}}-1}{2^{h^{(i)}}-1} \frac{\mathrm{vol}(\llbracket V^{(i)} \rrbracket)}{2}. \end{aligned} \tag{5.4.9}$$

Subsequently, as $\boldsymbol{R}^{(i)} > h^{(i)}$, we obtain the following for asymptotically large k:

$$\lim_{k\to\infty} \ell_k^{(i)} = \frac{1}{2} \frac{2^{nh^{(i)}}-1}{2^{h^{(i)}}-1} \frac{1}{1-2^{-n(R^{(i)}-h^{(i)})}} \mathrm{vol}(\llbracket V^{(i)} \rrbracket). \tag{5.4.10}$$

As $\ell_k^{(i)}$ is indeed an upper bound on the worst-case estimation error, i.e.,

$$\sup \llbracket \| E^{(i)}(nk) \| \rrbracket \leq \ell_{nk}^{(i)}, \ \forall k \in \mathbb{Z}_{\geq 0}.$$

Thus, the time-asymptotic worst-case estimation error max norm is upper bounded as follows

$$\begin{aligned}\lim_{k\to\infty} \sup \llbracket \| E^{(i)}(nk) \| \rrbracket &\leq \lim_{k\to\infty} \ell_{nk}^{(i)} \\ &= \frac{1}{2} \frac{2^{nh^{(i)}}-1}{2^{h^{(i)}}-1} \frac{1}{1-2^{-n(R^{(i)}-h^{(i)})}} \mathrm{vol}(\llbracket V^{(i)} \rrbracket). \end{aligned} \tag{5.4.11}$$

Therefore, this construction achieves the universal lower bound stated in Theorem 9 as an upper bound as well, and this shows the tightness of the decentralized lower bound.

5.5 Comparison between Centralized and Decentralized Lower Bounds

In the following numerical analysis, we consider two scalar LTI systems whose states are estimated over a two-user BAC. To compare the performance of the previously obtained lower bounds in Theorem 7 and 9, we examine two different scenarios for different feasible rate pairs $(\boldsymbol{R}^{(1)}, \boldsymbol{R}^{(2)})$. Firstly, we are interested in the setting where the topological entropies of both plants are of the same order (Scenario 1). Next, we investigate the case of having one plant with a dominant eigenvalue, i.e., its topological entropy is much larger than that of the second system. The relevant fixed parameters of both scenarios are given in Table 5.1. For the sake of simplicity, we refer to the centralized and decentralized lower bounds by $\mathbf{LB}^{\mathrm{cent}}$ and $\mathbf{LB}^{\mathrm{decent}}$, respectively.

Fixed Parameters	Scenario 1	Scenario 2
$d^{(1)}$	1	1
$d^{(2)}$	1	1
$h^{(1)}$	0.08	0.1
$h^{(2)}$	0.07	0.01
n	6	6

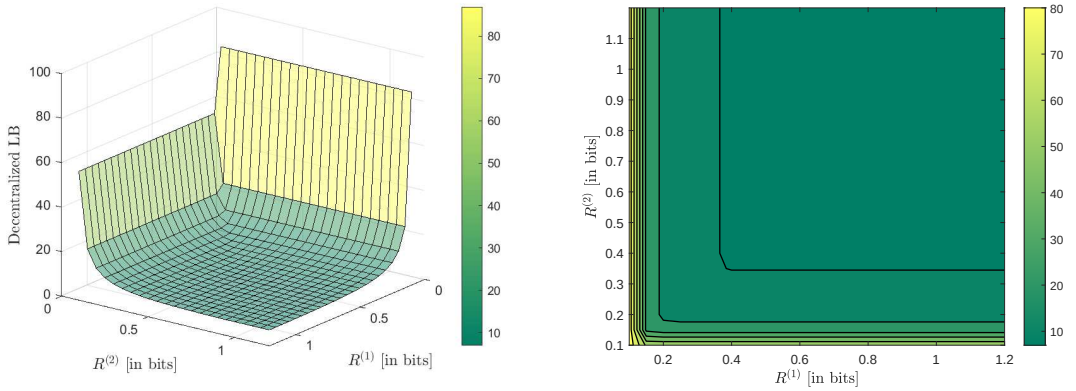
Table 5.1: System parameters of Scenario 1 and 2. The system parameters are chosen such that the topological entropies of the plants are of the same order in Scenario 1, whereas, in Scenario 2, the second plant has a tenfold larger topological entropy. The code block-length is set to 6 to be consistent with the numerical examples and analysis presented in previous parts of this thesis.

5.5.1 Numerical Results for Scenario 1

As it was theoretically proven in Section 5.4, the decentralized bound $\mathbf{LB}^{\text{decent}}$ shown in Fig. 5.1 turns out to be tighter than the centralized counterpart $\mathbf{LB}^{\text{cent}}$ illustrated in Fig. 5.2. This can be particularly seen when one of the rates is fixed to null. In this case, $\mathbf{LB}^{\text{decent}}$ remains constant (≈ 80 when $R^{(1)} = 0$ bit, and ≈ 52 when $R^{(2)} = 0$ bit) as the other rate increases, whereas the performance of $\mathbf{LB}^{\text{cent}}$ deteriorates exponentially. Another observation is the symmetry of $\mathbf{LB}^{\text{cent}}$ with respect to the line described by the equation $\mathscr{L} : R^{(2)} = R^{(1)}$. This is particularly obvious in Fig. 5.2(b) showing the contours of $\mathbf{LB}^{\text{cent}}$ in the $(R^{(1)}, R^{(2)})$ plane. This behavior stems from the fact that the expression of $\mathbf{LB}^{\text{cent}}$ depends on the sum of topological entropies h and the sum rate R, unlike $\mathbf{LB}^{\text{decent}}$ which is computed by selecting the subsystem with the parameters maximizing the RHS of (5.3.4), as shown in Corollary 1.

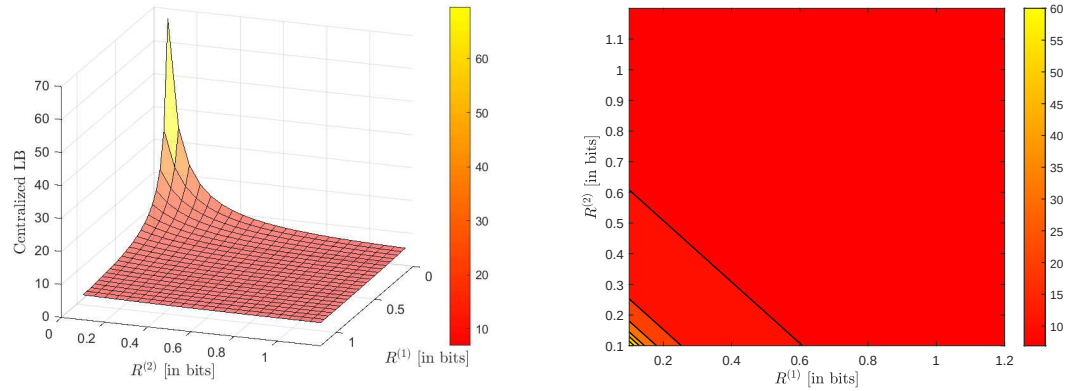
5.5.2 Numerical Results for Scenario 2

In this scenario, we are interested in the shape of $\mathbf{LB}^{\text{cent}}$ and $\mathbf{LB}^{\text{decent}}$ when there is a subsystem with a dominant eigenvalue. To this end, we set the topological entropies $h^{(1)} = 10 \cdot h^{(2)}$, as outlined in Table 5.1. Whilst the shape of $\mathbf{LB}^{\text{cent}}$, shown in Fig. 5.4, does not exhibit any particular change in comparison to the previous scenario, the order difference between $h^{(1)}$ and $h^{(2)}$ is reflected in the non-symmetry of $\mathbf{LB}^{\text{decent}}$ (with respect to $\mathscr{L} : R^{(2)} = R^{(1)}$) illustrated in Fig. 5.3. In fact, one can see that, for a fixed sufficiently small $R^{(1)}$, $\mathbf{LB}^{\text{decent}} \approx 38$. On the other hand, if $R^{(2)}$ is fixed to a small value, then the lower bound $\mathbf{LB}^{\text{decent}}$ becomes nearly 20. This suggests that a larger topological entropy results in a degradation in the performance of the state estimator, i.e., the corresponding state estimation error's magnitude increases.



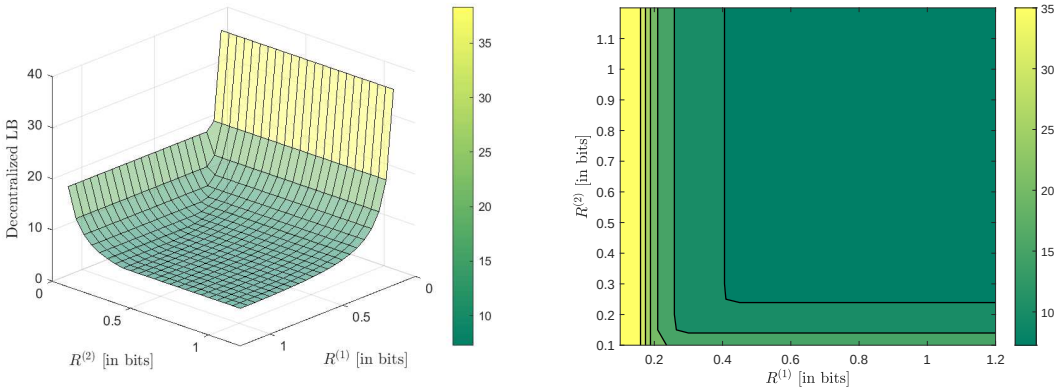
(a) Surface plot (b) Contour plot

Figure 5.1: Decentralized lower bound $\mathbf{LB}^{\text{decent}}$ for different $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ using the *max norm*, where $\boldsymbol{h}^{(1)}$ and $\boldsymbol{h}^{(2)}$ being of the same order.



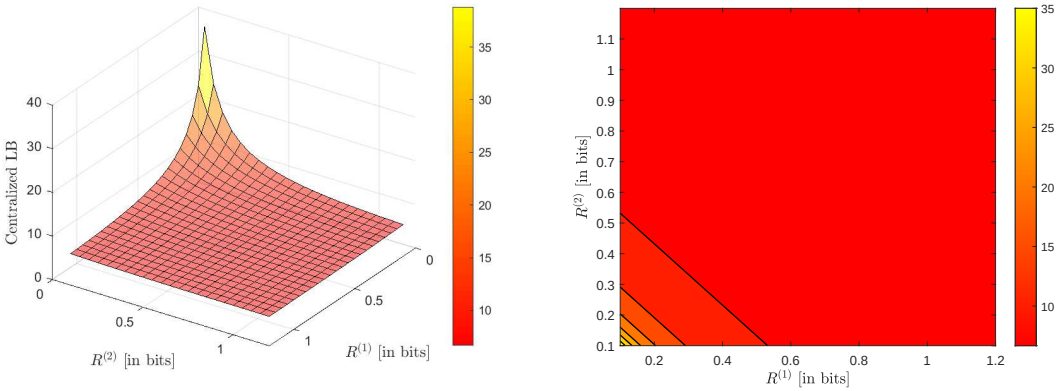
(a) Surface plot (b) Contour plot

Figure 5.2: Centralized lower bound $\mathbf{LB}^{\text{cent}}$ for different $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ using the *max norm*, with $\boldsymbol{h}^{(1)}$ and $\boldsymbol{h}^{(2)}$ being of the same order.



(a) Surface plot (b) Contour plot

Figure 5.3: Decentralized lower bound for different $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ using the *max norm* and $h^{(1)} = 10 \cdot h^{(2)}$.



(a) Surface plot (b) Contour plot

Figure 5.4: Centralized lower bound for different $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ using the *max norm* and $h^{(1)} = 10 \cdot h^{(2)}$.

5.6 Generalization of the Centralized Lower Bound

In our earlier analysis, we have focused on a system that consists of three unrelated LTI plants whose states are estimated over a two-user MAC. It is now natural to extend this setup, and derive a generalized lower bound on the estimation error norm. To this end, consider a dynamic system with d decoupled states and a state matrix $A = \mathrm{diag}(\lambda^{(1)}, \cdots, \lambda^{(d)})$ that is governed by the following equations

$$X(k+1) = AX(k) + V(k) \in \mathbb{R}^d, \tag{5.6.1a}$$

$$Y(k) = X(k) \in \mathbb{R}^d, \ \ \forall k \in \mathbb{Z}_{\geq 0}, \tag{5.6.1b}$$

where $V(k) \in [\![V]\!]$ denotes the process noise disturbing the plant at time instant $k \in \mathbb{Z}_{\geq 0}$. After being appropriately downsampled and encoded by means of a zero-error code with block-length n, the measurements are mapped to

$$S(k) = \gamma(k, Y(0:k)). \tag{5.6.2}$$

The encoded sequence is then communicated over some channel, which can be either point-to-point or MAC, operating at a *total* rate R. Using the received channel output $Q(0:nk)$, the estimator uses the decoding law $\delta(\cdot)$ to produce

$$\hat{X}(nk) = \delta(nk, Q(0:nk)). \tag{5.6.3}$$

Similar to previous parts of this thesis, the estimation error $E(nk)$, at sampled time nk, is defined as the difference between the true state and the estimate, i.e.,

$$E(nk) := X(nk) - \hat{X}(nk). \tag{5.6.4}$$

In the following, we derive a lower bound on the max norm of any vector of elements of the estimation error $E(nk)$ when $k \to \infty$ (Lemma 4), which then allows us to find the best lower bound on the time-asymptotic worst-case estimation error of the entire system.

5.6.1 Analytical Analysis

By virtue of the same argument used in Section 5.3 to show Theorem 9 and, subsequently, Corollary 1, it is also possible to derive a family of lower bounds on the worst-case max norm of the estimation errors $E^{\mathfrak{J}}(nk)$ associated with the subset of state indices $\mathfrak{J} \subseteq \mathfrak{D} := \{1, \cdots, d\}$ as stated in the following lemma:

Lemma 4 (Generalized Centralized Lower Bound)**.** *Consider a linear time-invariant system obeying*

(5.6.1a)-(5.6.1b) with a d-dimensional topological vector $h = (h^{(1)}, \cdots, h^{(d)})^T$ and whose outputs are encoded by (5.6.2) and estimated by (5.6.3) over a communication channel operating at a total rate R. Furthermore, let $\mathfrak{J}$ be a subset of state indices $\{1, \cdots, d\}$. If $R > \sum_{j\in\mathfrak{J}} h^{(j)}$, then

$$\limsup_{k\to\infty} [\![\|E^{\mathfrak{J}}(nk)\|]\!] \geq \frac{1}{2} \cdot \frac{2^{n\sum_{j\in\mathfrak{J}} h^{(j)}/|\mathfrak{J}|} - 1}{2^{\sum_{j\in\mathfrak{J}} h^{(j)}/|\mathfrak{J}|} - 1} \cdot \frac{\mathrm{vol}([\![V^{\mathfrak{J}}]\!])^{1/|\mathfrak{J}|}}{1 - 2^{-n(R-\sum_{j\in\mathfrak{J}} h^{(j)})/|\mathfrak{J}|}}, \tag{5.6.5}$$

where n is the code block-length and $[\![V^{\mathfrak{J}}]\!]$ is the range of noise disturbing the states with indices from $\mathfrak{J}$.

Proof. Consider a plant with d states associated with the eigenvalues $\Lambda = \{\lambda^{(1)}, \cdots, \lambda^{(d)}\}$. The plant's measurements are communicated over a channel operating at a total rate R and using a code with block length n. Hence, any state $X^{(i)}$, with $i \in [1:d]$, can be estimated using *at most* R bits/use. Let $\mathfrak{J}$ denote a subset of the eigenvalues' indices and its complementary set $\tilde{\mathfrak{J}} = \mathfrak{D} \setminus \mathfrak{J}$. The proof follows the same steps of Theorem 9, except that now, we fix the states associated with all modes with indices from $\tilde{\mathfrak{J}}$ to null, to obtain the following bound

$$\begin{aligned}
&\mathrm{vol}([\![X^{\mathfrak{J}}(nk)|q(0:nk),\, X^{\tilde{\mathfrak{J}}}(n(k+1)) = 0]\!]) \\
&\qquad \leq \sum_{q(nk+1:n(k+1))\in\mathscr{Q}^n} \mathrm{vol}([\![X^{\mathfrak{J}}(nk)|q(0:n(k+1)),\, X^{\tilde{\mathfrak{J}}}(n(k+1)) = 0]\!]) \\
&\qquad \leq |\mathscr{Q}^n| \max_{q(nk+1:n(k+1))} \mathrm{vol}([\![X^{\mathfrak{J}}(nk)|q(0:n(k+1)),\, X^{\tilde{\mathfrak{J}}}(n(k+1)) = 0]\!]) \\
&\qquad \leq 2^{nR} \max_{q(nk+1:n(k+1))} \mathrm{vol}([\![X^{\mathfrak{J}}(nk)|q(0:n(k+1)),\, X^{\tilde{\mathfrak{J}}}(n(k+1)) = 0]\!]),
\end{aligned} \tag{5.6.6}$$

where $|\mathscr{Q}|$ is the cardinality of the channel output alphabet $\mathscr{Q}$. Using (5.6.6) and proceeding to a similar manner as in the proof of (5.3.12), we obtain

$$\limsup_{k\to\infty} [\![\|E^{\mathfrak{J}}(nk)\|]\!] \geq \frac{1}{2} \cdot \frac{2^{n\sum_{j\in\mathfrak{J}} h^{(j)}/|\mathfrak{J}|} - 1}{2^{\sum_{j\in\mathfrak{J}} h^{(j)}/|\mathfrak{J}|} - 1} \cdot \frac{\mathrm{vol}([\![V^{\mathfrak{J}}]\!])^{1/|\mathfrak{J}|}}{1 - 2^{-n(R-\sum_{j\in\mathfrak{J}} h^{(j)})/|\mathfrak{J}|}}.$$

This completes the proof of (5.6.5). □

Using this result and the fact that $\mathfrak{J}$ can be *any* desired subset of indices from $\mathfrak{D}$, it is possible to improve the lower bound on the estimation error's norm by partitioning $\mathfrak{D}$ and maximizing the max norm of the RHS of (5.6.5) over *all* possible subsets of indices $\{\mathscr{P}_1, \cdots, \mathscr{P}_{\varpi}\}$. Notice that ϖ denotes the number of all possible subsets, and thus, it can be computed as follows

$$\varpi = \sum_{\xi=1}^{d} \binom{d}{\xi} = 2^d - 1. \tag{5.6.7}$$

State Dimension	Entropy vector h	Maximizing Set	Eval. of RHS (5.6.10)
$d=1$	(1)	$\{h^{(1)}\}$	43.4768
$d=2$	$(0.1,0.9)$	$\{h^{(2)}\}$	29.0860
	$(0.2,0.8)$	$\{h^{(2)}\}$	20.5888
	$(0.3,0.7)$	$\{h^{(1)},h^{(2)}\}$	17.7836
	$(0.5,0.5)$	$\{h^{(1)},h^{(2)}\}$	17.7836
$d=3$	$(0.1,0.1,0.8)$	$\{h^{(3)}\}$	20.5888
	$(0.13,0.13,0.74)$	$\{h^{(3)}\}$	17.0399
	$(0.2,0.2,0.6)$	$\{h^{(1)},h^{(2)},h^{(3)}\}$	16.5202
	$(0.1,0.45,0.45)$	$\{h^{(1)},h^{(2)},h^{(3)}\}$	16.5202
	$(0.333,0.333,0.333)$	$\{h^{(1)},h^{(2)},h^{(3)}\}$	16.5202

Table 5.2: Evaluation of RHS (5.6.10) for systems with the same total entropy $H=1$ bit, but different state dimensions d and entropy vectors $h:=(h^{(j)})_{j=1}^{d}$.

Define the function $\mathfrak{f}$ such that

$$\mathfrak{f}(\mathfrak{J}) := \frac{1}{2}\cdot\frac{2^{n\bar{h}_{\mathfrak{J}}}-1}{2^{\bar{h}_{\mathfrak{J}}}-1}\cdot\frac{\mathrm{vol}(\llbracket V^{\mathfrak{J}}\rrbracket)^{1/|\mathfrak{J}|}}{1-2^{-n(R/|\mathfrak{J}|-\bar{h}_{\mathfrak{J}})}}, \tag{5.6.8}$$

where $\bar{h}_{\mathfrak{J}}$ is the *average entropy per dimension* defined as

$$\bar{h}_{\mathfrak{J}} := \sum_{j\in\mathfrak{J}} h^{(j)}/|\mathfrak{J}|. \tag{5.6.9}$$

As we are considering the max norm in this analysis, the lower bound on the asymptotic worst-case estimation error magnitude of the entire system is then

$$\limsup_{k\to\infty}\llbracket\|E(nk)\|\rrbracket \geq \max_{i\in[1:\varpi]}|\mathfrak{f}(\mathscr{P}_i)|. \tag{5.6.10}$$

5.6.2 Numerical Example

To gain better understanding of the impact of the system parameters, such as the state dimension and the dominant eigenvalue, we evaluate numerically the RHS of (5.6.10) for different configurations. The fixed parameters in this computation are the coding block-length chosen as $n=6$, the total rate set to $R=1.31$ bits/use and the topological entropy sum $H=\sum_{i=1}^{d}h^{(i)}=1$ bit. Note that the individual topological entropies $h^{(i)}$ are selected differently in each experiment with the constraint that their total sum is equal to 1 bit. Table 5.2 summarizes the experiments' outcomes for 10 different scenarios.

The interesting fact, that one can observe in Table 5.2, is that the largest topological entropy does not always maximize the lower bound (5.6.10). From these results, two different regimes appear to

emerge, namely a regime where the lower bound performance is driven by the topological entropy with the largest magnitude, and another one in which all topological entropies of the system maximize (5.6.10).

To gain more insights into the behavior of the lower bound (5.6.10), we plot the function $\mathfrak{f}(\mathfrak{J})$ for a two-dimensional system with topological entropies $h^{(1)}$ and $h^{(2)}$ in Fig. 5.5. The entropy sum is fixed to a given value, namely **(a)** 1.2 bits and **(b)** 0.12 bits. Likewise, the coding block-length n and rate R are set to 6 and 1.31 bits/use, respectively. The purple line in both figures corresponds to the obtained lower bound, i.e., the maximum $\mathfrak{f}(\cdot)$ over all possible partitions of indices, namely $\{\{1\};\{2\};\{1,2\}\}$. One can see from Fig. 5.6.10 that two different regimes can be distinguished: In the first regime, the worst-case entropy drives the performance of the lower bound. This is observed when **(a)** $h_{\max} \leq 0.16$, and **(b)** $h_{\max} \leq 0.016$, where $h_{\max}$ is the dominant topological entropy, i.e., $h_{\max} = \max\{h^{(1)}, h^{(2)}\}$. On the other hand, in the second regime the function $\mathfrak{f}(\mathfrak{J})$ is maximized by the entropy sum, i.e., $\mathfrak{J} = \{1,2\}$.

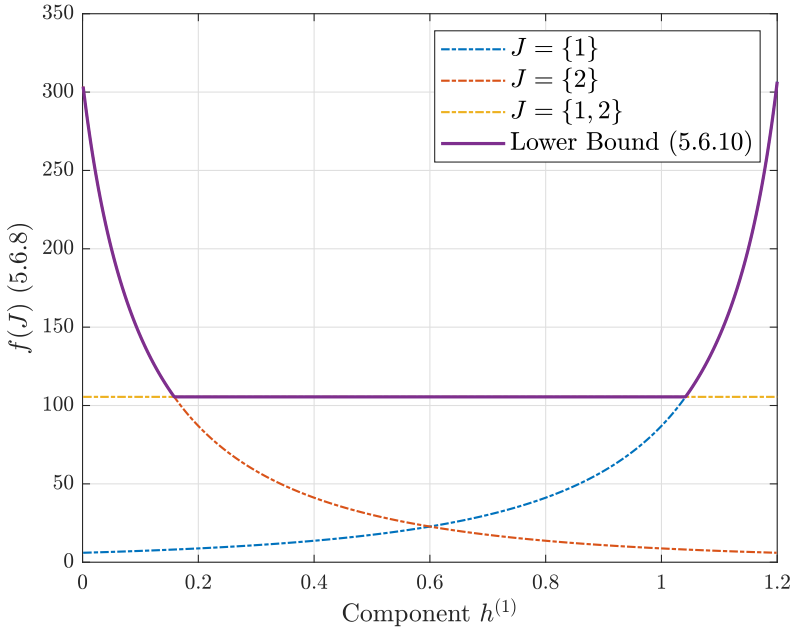
5.6.3 Analysis of the Large n Regime

Assuming that $R > H$, then as $n \to \infty$, the lower bound established in Lemma 4 becomes

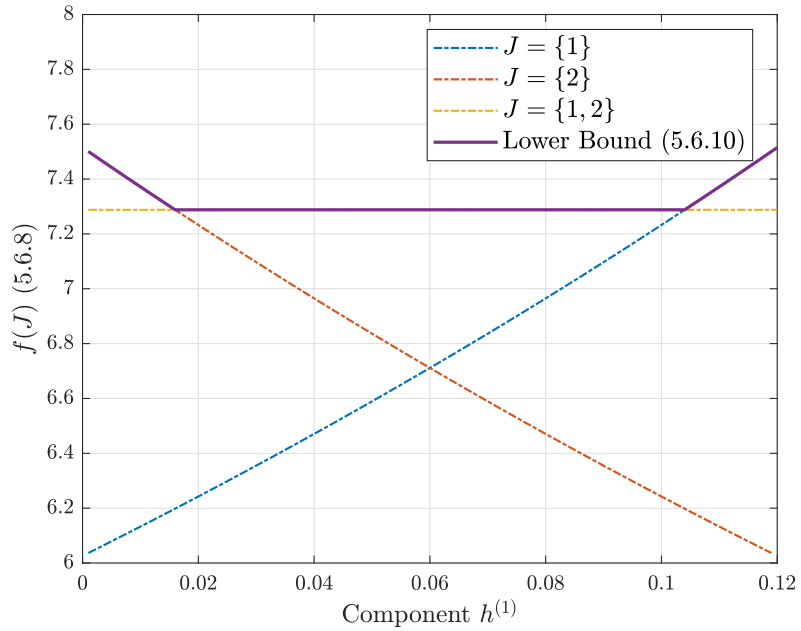
$$\begin{aligned}
\limsup_{k\to\infty} [\![\|E^{\mathfrak{J}}(nk)\|]\!] &\geq \frac{1}{2} \cdot \frac{2^{n\bar{h}_{\mathfrak{J}}} - 1}{2^{\bar{h}_{\mathfrak{J}}} - 1} \cdot \frac{\mathrm{vol}([\![V^{\mathfrak{J}}]\!])^{1/|\mathfrak{J}|}}{1 - 2^{-n(R/|\mathfrak{J}| - \bar{h}_{\mathfrak{J}})}} \\
&= \frac{1}{2} \cdot \frac{\mathrm{vol}([\![V^{\mathfrak{J}}]\!])^{1/|\mathfrak{J}|}}{2^{\bar{h}_{\mathfrak{J}}} - 1} (2^{n\bar{h}_{\mathfrak{J}}} - 1) \cdot \\
&\quad \left(1 + 2^{-n(R/|\mathfrak{J}| - \bar{h}_{\mathfrak{J}})} + 2^{-2n(R/|\mathfrak{J}| - \bar{h}_{\mathfrak{J}})} + O(2^{-3n(R/|\mathfrak{J}| - \bar{h}_{\mathfrak{J}})})\right) \qquad (5.6.11) \\
&= \frac{1}{2} \cdot \frac{\mathrm{vol}([\![V^{\mathfrak{J}}]\!])^{1/|\mathfrak{J}|}}{2^{\bar{h}_{\mathfrak{J}}} - 1} \left(2^{n\bar{h}_{\mathfrak{J}}} + 2^{n(2\bar{h}_{\mathfrak{J}} - R/|\mathfrak{J}|)} + O(2^{n(3\bar{h}_{\mathfrak{J}} - R/|\mathfrak{J}|)})\right) \\
&= \frac{1}{2} \cdot \frac{2^{n\bar{h}_{\mathfrak{J}}}}{2^{\bar{h}_{\mathfrak{J}}} - 1} \cdot \mathrm{vol}([\![V^{\mathfrak{J}}]\!])^{1/|\mathfrak{J}|} + o(2^{n\bar{h}_{\mathfrak{J}}}), \qquad (5.6.12)
\end{aligned}$$

where (5.6.11) is obtained using the asymptotic expansion of $\frac{1}{1-2^{-n(R/|\mathfrak{J}| - \bar{h}_{\mathfrak{J}})}}$. Fig. 5.6 exhibits the behavior of the RHS of (5.6.12) as the state dimension $|\mathfrak{J}|$ increases, for different values of n. The interesting fact, that one can see in this plot, is the drastic deterioration of the lower bound performance for medium-sized systems. For instance, for a system employing $n = 50$, the lower bound drops from a value slightly above 200 when the system is scalar to only 30 in the case of an 8-dimensional system. This performance deterioration means that the lower bound obtained in (5.6.12) becomes loose as the state dimension increases, and, if possible, it is better to assess the different states of the system individually.

Lemma 5 (Lower Bound in the Asymptotically Large Code Block-Length Regime)**.** *Given that $R > H$, then for asymptotically large code block-length n, i.e., $n \to \infty$, the lower bound on the time-asymptotic worst-*



(a) Entropy sum is fixed to $\sum_{j=1}^{2} h^{(j)} = 1.2$ bits.



(b) Entropy sum is fixed to $\sum_{j=1}^{2} h^{(j)} = 0.12$ bits.

Figure 5.5: The function $\mathfrak{f}(\mathfrak{J})$ evaluated for different system settings. We consider a two-dimensional system with entropies $h^{(1)}$ and $h^{(2)}$, whose sum is maintained fixed in each case **(a)** and **(b)**. The coding block-length is $n = 6$ and the rate sum is set as 1.31 bits/use. Each line depicts the function $\mathfrak{f}(\mathfrak{J})$ for $\mathfrak{J} = \{1\}$ (blue line), $\mathfrak{J} = \{2\}$ (red line) and $\mathfrak{J} = \{1, 2\}$ (yellow line). The lower bound (5.6.10) is shown as a solid line (purple).

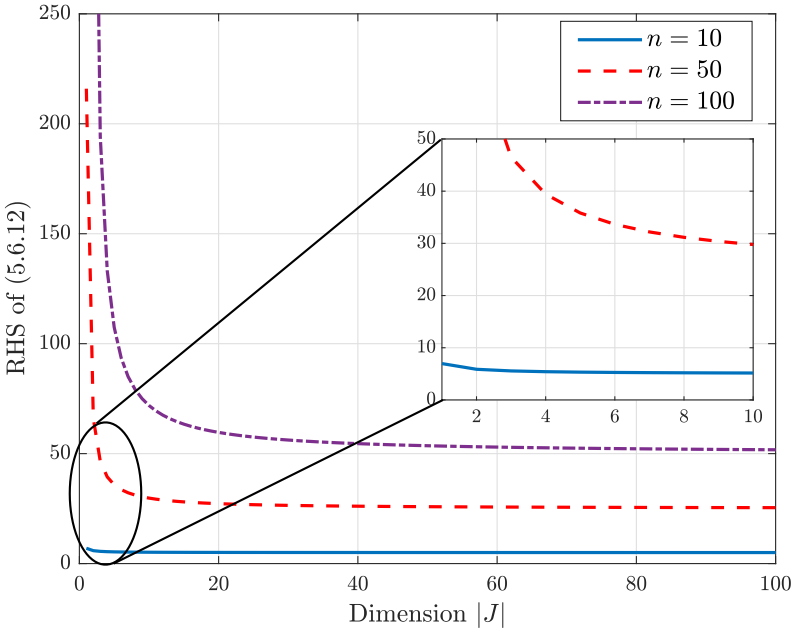


Figure 5.6: Behavior of the lower bound (5.6.12) for different state dimensions $|\mathfrak{J}|$. For the purpose of this plot, the entropy sum was fixed to $\sum_{j\in\mathfrak{J}} h^{(j)} = 0.1$.

case estimation error max norm (5.6.10) becomes

$$\limsup_{k\to\infty} [\![\|E(nk)\|]\!] \geq \frac{1}{2}\frac{2^{nh_{\max}}}{2^{h_{\max}}-1}\cdot \mathrm{vol}([\![V^{\max}]\!]) + o(2^{nh_{\max}}), \tag{5.6.13}$$

where $\boldsymbol{h}_{\max}$ is the largest element of the topological entropy vector $\boldsymbol{h}$, and $V^{\max}$ refers to the process noise affecting the corresponding state.

Proof. As $n \to \infty$, the RHS of (5.6.12) is maximized by the index subset $\mathfrak{J}$ with the largest $\bar{h}_{\mathfrak{J}} > 0$, since this yields the largest exponential factor $2^{n\bar{h}_{\mathfrak{J}}}$ in the leading term. This is achieved for the largest element of the topological entropy vector $\boldsymbol{h}$, that is when $\boldsymbol{h}_{\mathfrak{J}}$ corresponds to the worst-case component. The proof thereof is straightforward. Consider the topological entropy vector $\boldsymbol{h}$ with d components, i.e., $\boldsymbol{h} := (h^{(1)}, h^{(2)}, \cdots, h^{(d)})$. We assume w.l.o.g. that $h^{(1)} \geq h^{(2)} \geq \cdots \geq h^{(d)}$. As $h^{(1)} \geq h^{(j)}$ for any $j \in [1:d]$, then it holds, for any combination of components of $\boldsymbol{h}$ with indices $\mathfrak{J} \subseteq \{1, \cdots, d\}$, that

$$|\mathfrak{J}| \cdot h^{(1)} \geq \sum_{j\in\mathfrak{J}} h^{(j)}. \tag{5.6.14}$$

Subsequently, it follows

$$\underbrace{\frac{h^{(1)}}{1}}_{=\bar{h}_1} \geq \underbrace{\frac{\sum_{j\in\mathfrak{J}} h^{(j)}}{\mathfrak{J}}}_{=\bar{h}_{\mathfrak{J}}}, \ \forall \mathfrak{J} \subseteq \{1,\cdots,d\}. \tag{5.6.15}$$

As $h^{(1)}$ by assumption is the dominant topological entropy, this proves that the RHS of (5.6.12) is indeed always maximized by the worst-case component of $\boldsymbol{h}$, i.e.,

$$\limsup_{k\to\infty} [\![\|E(nk)\|]\!] \geq \frac{1}{2}\frac{2^{nh_{\max}}}{2^{h_{\max}}-1} \cdot \mathrm{vol}([\![V^{\max}]\!]) + o(2^{nh_{\max}}), \tag{5.6.16}$$

where $h_{\max}$ is the largest element of the topological entropy vector $\boldsymbol{h}$, and $V^{\max}$ refers to the process noise affecting the corresponding state. Thus, Lemma 5 is established. □

Remark 14. *Observe that if the d-dimensional process noise V has a hypercuboidal structure such that its range $[\![V]\!]$ is*

$$[\![V]\!] = [-\bar{V},\bar{V}]^d, \tag{5.6.17}$$

then the volume of $[\![V^{\max}]\!]$, where $V^{\max}$ is the noise term disturbing the state corresponding to the eigenvalue with the largest magnitude, i.e., $|\lambda_{\max}|$, becomes

$$\mathrm{vol}([\![V^{\max}]\!]) = 2\bar{V}. \tag{5.6.18}$$

Thus, the lower bound obtained in Lemma 5 can be further simplified to

$$\limsup_{k\to\infty} [\![\|E(nk)\|]\!] \geq \frac{2^{nh_{\max}}}{2^{h_{\max}}-1}\bar{V} + o(2^{nh_{\max}}). \tag{5.6.19}$$

5.7 Summary

By analyzing the evolution of the estimation error uncertainty volume, we firstly derived a centralized lower bound on the worst-case max norm of the asymptotic error (Theorem 7). We then further refined this result to obtain a decentralized lower bound on the norm of estimation errors associated with each of the three systems in Theorem 9 and Corollary 1. These lower bounds are qualified as universal, since they are applicable to any encoder-estimator tuple achieving bounded estimation errors regardless of their implementation or the employed technique.
Next, we showed that the obtained decentralized lower bounds are tight for the case of a scalar system by constructing a scheme achieving the bound of Theorem 9. Then, after generalizing the centralized lower bound to the case of a system with an arbitrary number of decoupled states, we studied the

behavior of the bound when large code lengths are employed. In our analysis, it was revealed that the dominant eigenvalue, i.e., the eigenvalue with the largest magnitude, is the decisive parameter in terms of driving the system's performance.

In order to provide a guarantee on the estimation performance, our aim in the next chapter is to derive an upper bound on the error when the measurements are quantized. Along with the lower bound, this will allow us to gain a better understanding of the estimation performance, and subsequently, characterize the fundamental trade-off between the communication data rate, code-length and system dynamics.

This page intentionally left blank.

"To know that we know what we know, and to know that we do not know what we do not know, that is true knowledge."

— Nicolaus Copernicus (1473–1543)

6

Asymptotically Tight Upper Bounds on the State Estimation Error

A natural step that follows the derivation of lower bounds on the state estimation error in the previous chapter is to examine if it is possible to delimit the error max-norm with an upper bound as well, i.e., to find guarantees on the estimation performance. To address this question, we consider the same setup as in the preceding chapters of this thesis, where the measurements are processed by means of a uniformly adaptive quantizer. Additionally, each state is estimated using a portion of the total available rate as it is explained in Section 6.1. Using this construction, upper bounds on the time-asymptotic max-norm of the estimation error are then presented in Theorems 10, 11 and 12. The first result (Theorem 10) is obtained for scalar Jordan blocks, whereas Theorems 11 and 12 address the more generalized case of non-scalar Jordan blocks with real and complex eigenvalues, respectively. Afterwards, we mainly focus on the case of one-dimensional Jordan blocks. As it turns out, the obtained upper bound depends on the individual rates allocated to estimate each state, and hence, the problem of finding the tightest upper bound boils down to deriving the optimal rate allocation strategy as shown in Theorem 13. Thus, we formulate the appropriate optimization problem and study it in Section 6.5. In the same section, the obtained solution of the optimization problem is studied in the asymptotically large code block-length regime (Theorem 6), and it is shown that the upper and lower bounds in this regime converge to the same limit in Proposition 1. Finally, a numerical example is presented in Section 6.6.

6.1 System Construction

Consider the three LTI systems $(\mathbf{P}^{(0)}, \mathbf{P}^{(1)}, \mathbf{P}^{(2)})$ which are coded and estimated via a two-user MAC with common message (Def. 8) such that the system's topological entropy vector $\boldsymbol{H} \in \mathbf{int}(\mathscr{C}_0)$. Recall that $\mathscr{C}_0$ denotes the zero-error capacity region of the deployed MAC, and is defined in (3.1.4). We are interested in deriving an upper bound on the max-norm of the state estimation error $\boldsymbol{E}(k)$ in (5.1.2). Note that the system employs an $(\lceil 2^{nR^{(0)}} \rceil, \lceil 2^{nR^{(1)}} \rceil, \lceil 2^{nR^{(2)}} \rceil, n)$ zero-error code (Def. 9). In order to simplify notation, we assume, without loss of generality, that $2^{nR^{(i)}} \in \mathbb{Z}_{\geq 0}$ for all $i \in \{0,1,2\}$. In the following analysis, we further assume that the systems' states are fully observed by the sensors, and are governed by the following equations:

$$X^{(i)}(k+1) = A^{(i)}X^{(i)}(k) + V^{(i)}(k), \tag{6.1.1a}$$

$$Y^{(i)}(k) = X^{(i)}(k), \ \ \forall i \in \{0,1,2\}, \tag{6.1.1b}$$

where the state matrices $A^{(0)}$, $A^{(1)}$ and $A^{(2)}$ are in Jordan canonical form [41], i.e.,

$$A^{(i)} = \begin{pmatrix} J_1^{(i)} & & \\ & \ddots & \\ & & J_{B^{(i)}}^{(i)} \end{pmatrix} \in \mathbb{R}^{d_i \times d_i}, \tag{6.1.2}$$

with $B^{(i)}$ Jordan blocks, such that β-th block is associated to eigenvalue $\lambda_\beta^{(i)}$, $\forall \beta \in [1 : B^{(i)}]$, and is written as

$$J_\beta^{(i)} = \begin{pmatrix} \lambda_\beta^{(i)} & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_\beta^{(i)} & 1 \\ & & & \lambda_\beta^{(i)} \end{pmatrix} \in \mathbb{R}^{b_\beta^{(i)} \times b_\beta^{(i)}}, \ \forall \beta \in [1 : B^{(i)}], \tag{6.1.3}$$

for real-valued eigenvalues[1]. Note that the total number of states associated with the system matrix $A^{(i)}$ is $d_i = \sum_{\beta=1}^{B^{(i)}} b_\beta^{(i)}$.

At each sampled time nk, the decoder generates an estimate $(\hat{X}^{(0)}, \hat{X}^{(1)}, \hat{X}^{(2)})$ of the state vector $(X^{(0)}, X^{(1)}, X^{(2)})$. Note that the system states $X^{(i)}$ at sampled time instant nk have the expression (5.1.3). The state estimation error $E^{(i)}(nk)$, defined as

$$E^{(i)}(nk) := \left(X^{(i)}(nk) - \hat{X}^{(i)}(nk) \right) \in \mathbb{R}^{d_i}, \ \forall i \in \{0,1,2\}, \tag{6.1.4}$$

is then quantized at each sampled time instant nk, for $k \in \mathbb{Z}_{\geq 0}$, by means of an adaptive quantizer

[1] A detailed discussion of Jordan blocks with *complex* eigenvalues is presented in Section 6.4

$\mathscr{Q}^{(i)}(\cdot)$. The quantization process occurs as follows: At $k=0$, a d_i-dimensional unit hypercube $\mathscr{I}^{(i)}(0)$, whose radius (defined using the max-norm) is $\|\mathscr{I}^{(i)}(0)\|=1$ and volume $\mathrm{vol}(\mathscr{I}^{(i)}(0))=2^{d_i}$, is partitioned into $2^{nR^{(i)}}$ hypercubes $\{\mathbf{I}_s^{(i)}(0)\}_{s=1}^{2^{nR^{(i)}}}$ such that, for all $s\in[1:2^{nR^{(i)}}]$, $\mathrm{vol}(\mathbf{I}_s^{(i)}(0))=2^{-nR^{(i)}}\mathrm{vol}(\mathscr{I}^{(i)}(0))$. Let $\boldsymbol{\sigma}^{(i)}(s)$ denote the center of $\mathbf{I}_s^{(i)}(nk)$ at time instant nk, $\forall s\in[1:2^{nR^{(i)}}]$. The encoder maps the scaled estimation error vector into the nearest point $\boldsymbol{\sigma}^{(i)}(s)$, i.e.,

$$\bar{E}^{(i)}(nk):=\mathscr{Q}^{(i)}\left(\left(L^{(i)}(k)\right)^{-1}\cdot E^{(i)}(nk)\right)\in\left\{\boldsymbol{\sigma}^{(i)}(1),\cdots,\boldsymbol{\sigma}^{(i)}(2^{nR^{(i)}})\right\}, \tag{6.1.5}$$

where $L^{(i)}(k)$ is a $d_i\times d_i$-diagonal matrix that comprises the scaling factors for the components of $E^{(i)}(nk)$ such that

$$L^{(i)}(k):=\begin{pmatrix}\ell^{(i,1)}(k) & & & \\ & \ell^{(i,2)}(k) & & \\ & & \ddots & \\ & & & \ell^{(i,d_i)}(k)\end{pmatrix}. \tag{6.1.6}$$

Each of the on-diagonal entries of $L^{(i)}(k)$, namely $\{\ell^{(i,\zeta)}(k)\}_{\zeta=1}^{d_i}$, forms an upper bound on the norm of the corresponding element of $E^{(i)}(nk)$, and is known to both the encoder and the decoder. The update equation of $\ell^{(i,\zeta)}(k)$ will be derived in the next part of this section. The subsequent state estimate $\hat{X}^{(i)}(n(k+1))$ is computed as follows

$$\hat{X}^{(i)}(n(k+1))=(A^{(i)})^n\hat{X}^{(i)}(nk)+L^{(i)}(k)\bar{E}^{(i)}(nk). \tag{6.1.7}$$

Rate Allocation for Systems $\mathbf{P}^{(0)}$, $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$. We define the topological entropy $H^{(i)}$ of the i-th system, where $i\in\{0,1,2\}$, as

$$\begin{aligned}H^{(i)}&:=\sum_{\beta:|\lambda_\beta^{(i)}|\geq 1}\sum_{\gamma=1}^{b_\beta^{(i)}}\log|\lambda_\beta^{(i)}|\\&=\sum_{\beta:|\lambda_\beta^{(i)}|\geq 1}b_\beta^{(i)}\log|\lambda_\beta^{(i)}|.\end{aligned} \tag{6.1.8}$$

Without loss of generality, assume that $|\lambda_\beta^{(i)}|\geq 1$ for all $\beta\in[1:B^{(i)}]$ and $i\in\{0,1,2\}$. Also, recall that the messages of Systems 1 and 2 are private and encoded by $\mathscr{E}^{(1)}$ and $\mathscr{E}^{(2)}$, respectively. On the other hand, information generated by System 0 is common to both encoders.

To achieve uniformly bounded estimation, the following set of inequalities must be satisfied, for

$i \in \{0,1,2\}$:

$$r_{\gamma,\beta}^{(i)} > \log|\lambda_\beta^{(i)}|,\ \forall \gamma \in [1 : b_\beta^{(i)}],\ \beta \in [1 : B^{(i)}], \tag{6.1.9a}$$

$$R^{(i)} \geq \sum_{\beta=1}^{B^{(i)}} \sum_{\gamma=1}^{b_\beta^{(i)}} r_{\gamma,\beta}^{(i)}, \tag{6.1.9b}$$

where $r_{\gamma,\beta}^{(i)}$ denotes the rate at which the γ-th element of the $b_\beta^{(i)}$-dimensional estimation error vector $E_\beta^{(i)}$, associated with the β-th Jordan block of the i-th system, is transmitted. Next, we proceed to derive upper bounds on the time-asymptotic max-norm of the estimation error for three different cases:

- *Scalar* Jordan blocks, i.e., $b_\beta^{(i)} = 1, \forall \beta \in [1 : B^{(i)}]$.
- *Non-scalar* Jordan blocks with *real* eigenvalues.
- *Non-scalar* Jordan blocks with *complex* eigenvalues.

6.2 Upper Bound on the State Estimation Error for Scalar Jordan Blocks

The first case that we consider in our analysis is when the Jordan blocks are one-dimensional, i.e., $b_\beta^{(i)} = 1$. We start by presenting the main result in Theorem 10 followed by its detailed proof.

Theorem 10. *Given three dynamical LTI systems $(P^{(0)}, P^{(1)}, P^{(2)})$ in (6.1.1a)-(6.1.1b) that are coded and estimated over a two-input single-output MAC with common message (Def. 8) such that the systems' topological entropy vector $H := (H^{(0)}, H^{(1)}, H^{(2)})^T$ in (6.1.8) lies within $\mathrm{int}(\mathscr{C}_0)$. Furthermore, assume that the Jordan blocks of the state matrix $A^{(i)}$ are scalars, i.e., $b_\beta^{(i)} = 1$, and the conditions (6.1.9a)-(6.1.9b) are satisfied. Then, the time-asymptotic max-norm of the estimation error $E^{(i)}$ associated with the i-th system can be upper bounded as*

$$\lim_{k\to\infty} \|E^{(i)}(nk)\| \leq \max_{\beta\in[1:B^{(i)}]} \left\{ \frac{1+|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}}} \cdot \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)} \right\}, \tag{6.2.1}$$

where $\bar{V}^{(i)} := \sup[\![\|V^{(i)}\|]\!]$ such that $V^{(i)}$ is the process noise affecting the i-th system, n refers to the code block-length, the number of Jordan blocks $B^{(i)}$ is here exactly equal to the state dimension d_i and $r_\beta^{(i)}$ the rate allocated for the estimation of the β-th state, with $\beta \in [1 : B^{(i)}]$.

Proof. The aim here is to derive an upper bound on the max-norm of the estimation error $E_\beta^{(i)}$, where $\beta \in [1 : B^{(i)}]$ and $i \in \{0,1,2\}$. At time instant $n(k+1)$, the error term $E_\beta^{(i)}(n(k+1))$ can be written

as follows

$$E_\beta^{(i)}(n(k+1)) = \left(J_\beta^{(i)}\right)^n E_\beta^{(i)}(nk) + \sum_{\xi=0}^{n-1} \left(J_\beta^{(i)}\right)^{n-1-\xi} V_\beta^{(i)}(nk+\xi) - \left(J_\beta^{(i)}\right)^n L_\beta^{(i)}(k)\bar{E}_\beta^{(i)}(nk), \tag{6.2.2}$$

where $V_\beta^{(i)}$ and $L_\beta^{(i)}$ are the noise term and the update matrix associated with $J_\beta^{(i)}$, respectively. Note here that since we are considering the error term associated with a single Jordan block $J_\beta^{(i)}$, the update matrix $L_\beta^{(i)}(k)$ consists of the same element denoted by $\ell_\beta^{(i)}(k)$ on its diagonal and zeros in the remaining entries, i.e., it is the product of the scalar $\ell_\beta^{(i)}(k)$ and the unit matrix of appropriate dimension. Hence, it is possible to simplify $L_\beta^{(i)}(k)\bar{E}_\beta^{(i)}(nk)$ in (6.2.2) with $\ell_\beta^{(i)}(k)\bar{E}_\beta^{(i)}(nk)$ to obtain

$$E_\beta^{(i)}(n(k+1)) = \left(J_\beta^{(i)}\right)^n \left(E_\beta^{(i)}(nk) - \ell_\beta^{(i)}(k)\bar{E}_\beta^{(i)}(nk)\right) + \sum_{\xi=0}^{n-1} \left(J_\beta^{(i)}\right)^{n-1-\xi} V_\beta^{(i)}(nk+\xi). \tag{6.2.3}$$

We define $\bar{V}^{(i)}$ as follows

$$\bar{V}^{(i)} := \sup[\![\|V^{(i)}\|]\!]. \tag{6.2.4}$$

The max-norm of $E_\beta^{(i)}(n(k+1))$ can therefore be upper bounded as follows

$$\begin{aligned}
\|E_\beta^{(i)}(n(k+1))\| &\leq \left\| \left(J_\beta^{(i)}\right)^n \left(E_\beta^{(i)}(nk) - \ell_\beta^{(i)}(k)\bar{E}_\beta^{(i)}(nk)\right) + \sum_{\xi=0}^{n-1} \left(J_\beta^{(i)}\right)^{n-1-\xi} V_\beta^{(i)}(nk+\xi) \right\| \\
&\leq \left\| \left(J_\beta^{(i)}\right)^n \right\| \cdot \left\| E_\beta^{(i)}(nk) - \ell_\beta^{(i)}(k)\bar{E}_\beta^{(i)}(nk) \right\| + \left\| \sum_{\xi=0}^{n-1} \left(J_\beta^{(i)}\right)^{n-1-\xi} V_\beta^{(i)}(nk+\xi) \right\| \qquad (6.2.5) \\
&\leq \left\| \left(J_\beta^{(i)}\right)^n \right\| \cdot |\ell_\beta^{(i)}(k)| \cdot \left\| \frac{1}{\ell_\beta^{(i)}(k)} E_\beta^{(i)}(nk) - \bar{E}_\beta^{(i)}(nk) \right\| + \\
&\quad \sum_{\xi=0}^{n-1} \left\| \left(J_\beta^{(i)}\right)^{n-1-\xi} \right\| \|V_\beta^{(i)}(nk+\xi)\| \qquad (6.2.6) \\
&\leq \left\| \left(J_\beta^{(i)}\right)^n \right\| \cdot |\ell_\beta^{(i)}(k)| \cdot \left\| \frac{1}{\ell_\beta^{(i)}(k)} E_\beta^{(i)}(nk) - \bar{E}_\beta^{(i)}(nk) \right\| + \sum_{\xi=0}^{n-1} \left\| \left(J_\beta^{(i)}\right)^{n-1-\xi} \right\| \bar{V}^{(i)} \qquad (6.2.7)
\end{aligned}$$

$$= \left\| \left(J_\beta^{(i)}\right)^n \right\| \cdot \ell_\beta^{(i)}(k) \cdot \left\| \frac{1}{\ell_\beta^{(i)}(k)} E_\beta^{(i)}(nk) - \bar{E}_\beta^{(i)}(nk) \right\| + \sum_{\xi=0}^{n-1} \left\| \left(J_\beta^{(i)}\right)^{n-1-\xi} \right\| \bar{V}^{(i)}, \tag{6.2.8}$$

where (6.2.5) and (6.2.6) follow from the triangle inequality, (6.2.7) is by virtue of the definition of $\bar{V}^{(i)}$ (6.2.4), and (6.2.8) follows from the assumption that $\ell_\beta^{(i)}(k)$ is an upper bound on the max-norm of $E_\beta^{(i)}(nk)$, and hence, it is positive. Based on a result in [41, pg. 138], it turns out that $\exists K > 0$ such that for the Jordan block $J_\beta^{(i)} \in \mathbb{R}^{b_\beta^{(i)} \times b_\beta^{(i)}}$ it holds

$$\left\| \left(J_\beta^{(i)}\right)^n \right\| \leq K n^{b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^n, \tag{6.2.9}$$

where $\lambda_\beta^{(i)} = \text{eig}(J_\beta^{(i)})$. Note that, in this inequality, we do not necessarily need to assume that $b_\beta^{(i)} = 1$.

Hence, it follows from (6.2.8) and (6.2.9)

$$\begin{aligned}
&\|E_\beta^{(i)}(n(k+1))\| \\
&\leq K n^{b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^n \ell_\beta^{(i)}(k) \cdot \left\| \frac{1}{\ell_\beta^{(i)}(k)} E_\beta^{(i)}(nk) - \bar{E}_\beta^{(i)}(nk) \right\| + \sum_{\xi=0}^{n-1} K(n-1-\xi)^{b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^{n-1-\xi} \bar{V}^{(i)} \\
&\leq K n^{b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^n \ell_\beta^{(i)}(k) \cdot \left\| \frac{1}{\ell_\beta^{(i)}(k)} E_\beta^{(i)}(nk) - \bar{E}_\beta^{(i)}(nk) \right\| + K(n-1)^{b_\beta^{(i)}-1} \sum_{\xi=0}^{n-1} |\lambda_\beta^{(i)}|^{n-1-\xi} \bar{V}^{(i)},
\end{aligned} \tag{6.2.10}$$

where (6.2.10) results from the fact that $(n-1-\xi)^{b_\beta^{(i)}-1} \leq (n-1)^{b_\beta^{(i)}-1}, \forall \xi \in [0 : n-1]$ as $n, b_\beta^{(i)} \in \mathbb{Z}_{\geq 1}$.

As the RHS of (6.2.10) is an upper bound on the max-norm of $E_\beta^{(i)}(n(k+1))$, and consequently, it is an upper bound on all of its components $\{E_{\gamma,\beta}^{(i)}(n(k+1))\}_{\gamma=1}^{b_\beta^{(i)}}$, we now aim to derive an upper bound on the term $\|E_{\gamma,\beta}^{(i)}(nk) - \ell_{\gamma,\beta}^{(i)}(k)\bar{E}_{\gamma,\beta}^{(i)}(nk)\|$, with $\gamma \in [1 : b_\beta^{(i)}]$. Firstly, recall that $\bar{E}_{\gamma,\beta}^{(i)}(nk)$ is associated with the γ-th component of the estimation error vector, and corresponds to

$$\bar{E}_{\gamma,\beta}^{(i)}(nk) := \mathscr{Q}\left(E_{\gamma,\beta}^{(i)}(nk)/\ell_{\gamma,\beta}^{(i)}(k)\right) \in \left\{ \sigma_{\gamma,\beta}^{(i)}(1), \cdots, \sigma_{\gamma,\beta}^{(i)}(2^{nr_{\gamma,\beta}^{(i)}}) \right\}, \tag{6.2.11}$$

where $\sigma_{\gamma,\beta}^{(i)}(s)$ is the center of the s-th hypercube $\mathbf{I}_s^{(i,\gamma)}(nk)$, $\forall s \in [1 : 2^{nr_{\gamma,\beta}^{(i)}}]$, which is restricted to the γ-th component of the i-th system. The diameter of each hypercube is $2^{-nr_{\gamma,\beta}^{(i)}} \cdot 2$, meaning that the radius is $\|\mathbf{I}_s^{(i,\gamma)}(nk)\| = 2^{-nr_{\gamma,\beta}^{(i)}}$. By construction, to be mapped to $\bar{E}_{\gamma,\beta}^{(i)}(nk)$, the scaled version of

the real error, namely $(E^{(i)}_{\gamma,\beta}(nk)/\ell^{(i)}_{\gamma,\beta}(k))$ must lie within the hypercube with center $\bar{E}^{(i)}_{\gamma,\beta}(nk)$ and radius $2^{-nr^{(i)}_{\gamma,\beta}}$. Thus, we obtain

$$\begin{aligned}\left\| \frac{1}{\ell^{(i)}_{\gamma,\beta}(k)} E^{(i)}_{\gamma,\beta}(nk) - \bar{E}^{(i)}_{\gamma,\beta} \right\| &\leq \|\mathbf{I}_s^{(i,\gamma)}(nk)\| \\ &= 2^{-nr^{(i)}_{\gamma,\beta}}. \end{aligned} \tag{6.2.12}$$

Hence, by plugging (6.2.12) into (6.2.10), we obtain the following upper bound on $\|E^{(i)}_{\gamma,\beta}(n(k+1))\|$:

$$\|E^{(i)}_{\gamma,\beta}(n(k+1))\| \leq Kn^{b^{(i)}_\beta - 1}|\lambda^{(i)}_\beta|^n \ell^{(i)}_{\gamma,\beta}(k) 2^{-nr^{(i)}_{\gamma,\beta}} + K(n-1)^{b^{(i)}_\beta - 1} \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1} \bar{V}^{(i)}. \tag{6.2.13}$$

Now, the RHS of (6.2.13) is used as the update equation of $\ell^{(i)}_{\gamma,\beta}(k)$, i.e.,

$$\begin{aligned}\ell^{(i)}_{\gamma,\beta}(k) &:= Kn^{b^{(i)}_\beta - 1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}} \ell^{(i)}_{\gamma,\beta}(k-1) + K(n-1)^{b^{(i)}_\beta - 1} \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1} \bar{V}^{(i)} \\ &= \left(Kn^{b^{(i)}_\beta - 1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}} \right)^k \ell^{(i)}_{\gamma,\beta}(0) \\ &\quad + K(n-1)^{b^{(i)}_\beta - 1} \sum_{\xi=0}^{k-1} (Kn^{b^{(i)}_\beta - 1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}})^{k-1-\xi} \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1} \bar{V}^{(i)}.\end{aligned}$$

Thus, we write $\ell^{(i)}_{\gamma,\beta}(k)$ as follows

$$\begin{aligned}\ell^{(i)}_{\gamma,\beta}(k) &= \left(Kn^{b^{(i)}_\beta - 1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}} \right)^k \ell^{(i)}_{\gamma,\beta}(0) \\ &\quad + K(n-1)^{b^{(i)}_\beta - 1} \frac{\left(Kn^{b^{(i)}_\beta - 1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}} \right)^k - 1}{Kn^{b^{(i)}_\beta - 1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}} - 1} \cdot \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1} \bar{V}^{(i)}.\end{aligned} \tag{6.2.14}$$

As $\ell^{(i)}_{\gamma,\beta}(k)$ is an upper bound on $\|E^{(i)}_{\gamma,\beta}(nk)\|$, we therefore obtain the following result $\forall k \in \mathbb{Z}_{\geq 1}$

$$\|E^{(i)}_{\gamma,\beta}(nk)\| \leq \left(Kn^{b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}}\right)^k \ell^{(i)}_{\gamma,\beta}(0) + K(n-1)^{b^{(i)}_\beta-1} \frac{\left(Kn^{b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}}\right)^k - 1}{Kn^{b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}} - 1} \cdot \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1}\bar{V}^{(i)}. \tag{6.2.15}$$

When $b^{(i)}_\beta = 1$, the system is completely diagonalizable. In this case, $K = 1$, and since $\gamma \in [1 : b^{(i)}_\beta]$ can only be 1 here, we write $r^{(i)}_\beta \equiv r^{(i)}_{\gamma,\beta}$ and $E^{(i)}_\beta(n(k+1)) \equiv E^{(i)}_{\gamma,\beta}(n(k+1))$. Thus, (6.2.15) becomes

$$\|E^{(i)}_\beta(n(k+1))\| \leq \left(|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta}\right)^{k+1} \ell^{(i)}_\beta(0) + \left(|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta} + 1\right) \frac{\left(|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta}\right)^k - 1}{|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta} - 1} \cdot \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1}\bar{V}^{(i)}. \tag{6.2.16}$$

As $k \to \infty$, the RHS of (6.2.16) becomes

$$\begin{aligned}\lim_{k\to\infty} &\left\{ \Big(\underbrace{|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta}}_{<1,\ \text{as } r^{(i)}_\beta > \log|\lambda^{(i)}_\beta|} \Big)^{k+1} \ell^{(i)}_\beta(0) + \left(|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta} + 1\right) \frac{|\lambda^{(i)}_\beta|^{kn} 2^{-nkr^{(i)}_\beta} - 1}{|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta} - 1} \cdot \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1}\bar{V}^{(i)} \right\} \\ &= \frac{1 + |\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta}}{1 - |\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta}} \cdot \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1}\bar{V}^{(i)}. \end{aligned} \tag{6.2.17}$$

Therefore, the max-norm of the estimation error $E^{(i)}$ associated with the i-th system can be upper bounded as

$$\lim_{k\to\infty} \|E^{(i)}(nk)\| \leq \max_{\beta \in [1:B^{(i)}]} \left\{ \frac{1 + |\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta}}{1 - |\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_\beta}} \cdot \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1}\bar{V}^{(i)} \right\}. \tag{6.2.18}$$

This completes the proof of Theorem 10. □

6.3 Upper Bound on the State Estimation Error for Non-Scalar Jordan Blocks with Real Eigenvalues

After having studied the case of scalar Jordan blocks, we continue now with a more general case, namely when the Jordan blocks are matrices with real eigenvalues and $b_\beta^{(i)} > 1$. The upper bound on the max-norm of the estimation error in this setting is shown in Theorem 11.

Theorem 11. *Given three dynamical LTI systems $(\boldsymbol{P}^{(0)}, \boldsymbol{P}^{(1)}, \boldsymbol{P}^{(2)})$ in (6.1.1a)-(6.1.1b) that are coded and estimated over a two-input single-output MAC with common message (Def. 8) such that the systems' topological entropy vector $\boldsymbol{H} := (H^{(0)}, H^{(1)}, H^{(2)})^T$ in (6.1.8) lies within $\operatorname{int}(\mathscr{C}_0)$. Furthermore, the state matrix $A^{(i)}$ (6.1.2) consists of non-scalar Jordan blocks of the form (6.1.3), and the following conditions hold*

$$r_{\gamma,\beta}^{(i)} > \log|\lambda_\beta^{(i)}| + \frac{1}{n}\log\left(\frac{1-|\lambda_\beta^{(i)}|^{-b_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^{-1}}\right) + (b_\beta^{(i)}-1)\frac{\log n}{n}, \tag{6.3.1a}$$

$$R^{(i)} \geq \sum_{\beta=1}^{B^{(i)}} \sum_{\gamma=1}^{b_\beta^{(i)}} r_{\gamma,\beta}^{(i)}, \ \forall \gamma \in [1 : b_\beta^{(i)}], \ \beta \in [1 : B^{(i)}]. \tag{6.3.1b}$$

Then, we have

$$\lim_{k\to\infty} \|E^{(i)}(nk)\| \leq \max_{\beta\in[1:B^{(i)}]} \max_{\gamma\in[1:b_\beta^{(i)}]} \left\{ \frac{1-|\lambda_\beta^{(i)}|^{-b_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^{-1}} \frac{(n-1)^{b_\beta^{(i)}-1}}{1-\frac{1-|\lambda_\beta^{(i)}|^{-b_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^{-1}} n^{b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n 2^{-nr_{\gamma,\beta}^{(i)}}} \cdot \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)} \right\}, \tag{6.3.2}$$

with $\bar{V}^{(i)} := \sup[\![\|V^{(i)}\|]\!]$.

Remark 15. *Observe here that the required condition on the rates $r_{\gamma,\beta}^{(i)}$, $\forall \gamma \in [1 : b_\beta^{(i)}]$, $\beta \in [1 : B^{(i)}]$, namely (6.3.1a), is more restrictive than (6.1.9a), when the Jordan blocks are scalar. This condition reflects the trade-off involving the Jordan block size $b_\beta^{(i)}$, the code length n, the eigenvalue magnitude, and the allocated rate $r_{\gamma,\beta}^{(i)}$:*

- *If the code length n remains fixed, while the Jordan block size $b_\beta^{(i)}$ is increased, one can see that the RHS of (6.3.1a) would also increase, implying the need to allocate more data resources to reliably estimate the states associated with this Jordan block.*
- *For a fixed Jordan block size and asymptotically large code length n, the RHS of (6.3.1a) approaches $\log|\lambda_\beta^{(i)}|$, suggesting that by making n large enough it is possible to maintain the same estimation performance while reducing the allocated data rate.*

Proof. Firstly, it should be pointed out that the proof of (6.2.15) presented in Section 6.2 is valid for non-scalar Jordan blocks with real eigenvalues, as $b_\beta^{(i)}$ was not assumed to be necessarily equal to 1. Hence, the max-norm of each element of the estimation error vector is upper bounded as follows

$$\begin{aligned}\|E_{\gamma,\beta}^{(i)}(nk)\| \leq &\left(Kn^{b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n 2^{-nr_{\gamma,\beta}^{(i)}}\right)^k \ell_{\gamma,\beta}^{(i)}(0)+ \\ &K(n-1)^{b_\beta^{(i)}-1}\frac{\left(Kn^{b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n 2^{-nr_{\gamma,\beta}^{(i)}}\right)^k-1}{Kn^{b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n 2^{-nr_{\gamma,\beta}^{(i)}}-1}\cdot\frac{|\lambda_\beta^{(i)}|^n-1}{|\lambda_\beta^{(i)}|-1}\bar{V}^{(i)},\end{aligned} \tag{6.3.3}$$

where $\gamma \in [1 : b_\beta^{(i)}]$, $\beta \in [1 : B^{(i)}]$. Observe that for the RHS of (6.3.3) to converge, the following constraint must be satisfied

$$Kn^{b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n 2^{-nr_{\gamma,\beta}^{(i)}} < 1, \tag{6.3.4}$$

which leads to

$$\begin{aligned}nr_{\gamma,\beta}^{(i)} &> \log\left(Kn^{b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n\right) \\ &= \log K + (b_\beta^{(i)}-1)\log n + n\log|\lambda_\beta^{(i)}|.\end{aligned} \tag{6.3.5}$$

This condition is more restrictive than (6.1.9a) which requires that $r_{\gamma,\beta}^{(i)} > \log|\lambda_\beta^{(i)}|$.

The first step in this proof is to find an expression of K in terms of the problem's parameters. Recall that K is a strictly positive real number for which the max-norm of the Jordan block $J_\beta^{(i)} \in \mathbb{R}^{b_\beta^{(i)}\times b_\beta^{(i)}}$ to the n-th power is upper bounded as follows [41, pg. 138]

$$\left\|\left(J_\beta^{(i)}\right)^n\right\| \leq Kn^{b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n, \tag{6.3.6}$$

where $\lambda_\beta^{(i)} = \mathrm{eig}(J_\beta^{(i)})$. For real eigenvalues $\lambda_\beta^{(i)}$, the Jordan matrix can be expressed as

$$J_\beta^{(i)} = \lambda_\beta^{(i)} I_{b_\beta^{(i)}} + N_{b_\beta^{(i)}}, \tag{6.3.7}$$

where $N_{b_\beta^{(i)}}$ is the $b_\beta^{(i)} \times b_\beta^{(i)}$ of the following form

$$N_{b_\beta^{(i)}} = \begin{pmatrix} 0 & 1 & & \mathbf{0} \\ & \ddots & \ddots & \\ & & & 1 \\ \mathbf{0} & & & 0 \end{pmatrix}. \tag{6.3.8}$$

Then, $\|(J_\beta^{(i)})^n\|$ can be expressed as

$$\begin{aligned} \left\| \left(J_\beta^{(i)}\right)^n \right\| &= \left\| \sum_{\xi=0}^{n} \binom{n}{\xi} (\lambda_\beta^{(i)})^\xi N_{b_\beta^{(i)}}^{n-\xi} \right\| \\ &= \left\| \sum_{\xi=n-b_\beta^{(i)}+1}^{n} \binom{n}{\xi} (\lambda_\beta^{(i)})^\xi N_{b_\beta^{(i)}}^{n-\xi} \right\| \qquad (6.3.9) \\ &\leq \sum_{\xi=n-b_\beta^{(i)}+1}^{n} \left| \binom{n}{\xi} (\lambda_\beta^{(i)})^\xi \right| \left\| N_{b_\beta^{(i)}}^{n-\xi} \right\| \\ &= \sum_{r=0}^{b_\beta^{(i)}-1} \left| \binom{n}{n-r} (\lambda_\beta^{(i)})^{n-r} \right| \left\| N_{b_\beta^{(i)}}^{r} \right\| \qquad (6.3.10) \\ &\leq \sum_{r=0}^{b_\beta^{(i)}-1} \left| \binom{n}{n-r} (\lambda_\beta^{(i)})^{n-r} \right|, \qquad (6.3.11) \end{aligned}$$

where Step (6.3.9) follows from the fact that $N_{b_\beta^{(i)}}^n = \mathbf{0}$ for all $n \geq b_\beta^{(i)}$ and (6.3.10) is a straightforward consequence of variable change, i.e., $\xi = n - r$. Additionally, for each $r \in [0 : b_\beta^{(i)} - 1]$, it holds

$$\begin{aligned} \left| \binom{n}{n-r} (\lambda_\beta^{(i)})^{n-r} \right| &= \left| \frac{n(n-1)\cdots(n-r+1)(\lambda_\beta^{(i)})^n}{r!(\lambda_\beta^{(i)})^r} \right| \\ &\leq \left| \frac{n^r (\lambda_\beta^{(i)})^n}{r!(\lambda_\beta^{(i)})^r} \right|. \qquad (6.3.12) \end{aligned}$$

Therefore, $\|(J_\beta^{(i)})^n\|$ can be upper bounded as follows

$$\left\| \left(J_\beta^{(i)}\right)^n \right\| \leq \sum_{r=0}^{b_\beta^{(i)}-1} \left| \binom{n}{n-r} (\lambda_\beta^{(i)})^{n-r} \right|$$

$$\leq \sum_{r=0}^{b_\beta^{(i)}-1} \left| \frac{n^r (\lambda_\beta^{(i)})^n}{r!(\lambda_\beta^{(i)})^r} \right|$$
$$\leq n^{b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^n \sum_{r=0}^{b_\beta^{(i)}-1} \frac{|\lambda_\beta^{(i)}|^{-r}}{r!}. \tag{6.3.13}$$

We can now upper bound the finite sum $\sum_{r=0}^{b_\beta^{(i)}-1} \frac{|\lambda_\beta^{(i)}|^{-r}}{r!}$ as follows. Since $\frac{|\lambda_\beta^{(i)}|^{-r}}{r!} \leq |\lambda_\beta^{(i)}|^{-r}$ for any $r \in [0 : b_\beta^{(i)} - 1]$, an upper bound on the finite sum is then

$$\sum_{r=0}^{b_\beta^{(i)}-1} \frac{|\lambda_\beta^{(i)}|^{-r}}{r!} \leq \sum_{r=0}^{b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^{-r}$$
$$= \frac{1 - |\lambda_\beta^{(i)}|^{-b_\beta^{(i)}}}{1 - |\lambda_\beta^{(i)}|^{-1}}. \tag{6.3.14}$$

Thus, K from (6.3.6) is found to be

$$K = \frac{1 - |\lambda_\beta^{(i)}|^{-b_\beta^{(i)}}}{1 - |\lambda_\beta^{(i)}|^{-1}}. \tag{6.3.15}$$

Hence, the requirement (6.3.5) can now be expressed as

$$n r_{\gamma,\beta}^{(i)} > n \log |\lambda_\beta^{(i)}| + \log \left(\frac{1 - |\lambda_\beta^{(i)}|^{-b_\beta^{(i)}}}{1 - |\lambda_\beta^{(i)}|^{-1}} \right) + (b_\beta^{(i)} - 1) \log n. \tag{6.3.16}$$

Now, recall that the generalized upper bound on $\|E_{\gamma,\beta}^{(i)}\|$ (6.3.3) is

$$\|E_{\gamma,\beta}^{(i)}(nk)\| \leq \left(\underbrace{K n^{b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^n 2^{-n r_{\gamma,\beta}^{(i)}}}_{<1, \text{ by } (6.3.16)} \right)^k \ell_{\gamma,\beta}^{(i)}(0)$$
$$+ K(n-1)^{b_\beta^{(i)}-1} \frac{\left(K n^{b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^n 2^{-n r_{\gamma,\beta}^{(i)}} \right)^k - 1}{K n^{b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^n 2^{-n r_{\gamma,\beta}^{(i)}} - 1} \cdot \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}. \tag{6.3.17}$$

By substituting the expression of $\boldsymbol{K}$ (6.3.15) into (6.3.17) and taking the limit $k \to \infty$, we obtain

$$\lim_{k\to\infty} \|E^{(i)}_{\gamma,\beta}(nk)\| \leq \frac{1-|\lambda^{(i)}_\beta|^{-b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} \frac{(n-1)^{b^{(i)}_\beta-1}}{1-\frac{1-|\lambda^{(i)}_\beta|^{-b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} n^{b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}}} \cdot \frac{|\lambda^{(i)}_\beta|^n-1}{|\lambda^{(i)}_\beta|-1}\bar{V}^{(i)}. \tag{6.3.18}$$

Consequently, the upper bound on the max-norm of the estimation error associated with the i-th system is

$$\begin{aligned}\lim_{k\to\infty} \|E^{(i)}(nk)\| &\leq \max_{\beta\in[1:B^{(i)}]} \max_{\gamma\in[1:b^{(i)}_\beta]} \left\{ \lim_{k\to\infty} \|E^{(i)}_{\gamma,\beta}(nk)\| \right\} \\ &\leq \max_{\beta\in[1:B^{(i)}]} \max_{\gamma\in[1:b^{(i)}_\beta]} \left\{ \frac{1-|\lambda^{(i)}_\beta|^{-b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} \frac{(n-1)^{b^{(i)}_\beta-1}}{1-\frac{1-|\lambda^{(i)}_\beta|^{-b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} n^{b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}}} \cdot \right. \\ &\qquad \left. \frac{|\lambda^{(i)}_\beta|^n-1}{|\lambda^{(i)}_\beta|-1}\bar{V}^{(i)} \right\}. \end{aligned} \tag{6.3.19}$$

Hence, Theorem 11 is established. □

6.4 Upper Bound on the State Estimation Error for Non-Scalar Jordan Blocks with Complex Eigenvalues

We now broaden the horizon of our study by considering non-scalar Jordan blocks with *complex* eigenvalues $\lambda^{(i)}_\beta$, namely $\lambda^{(i)}_\beta = \Re(\lambda^{(i)}_\beta) \pm j\Im(\lambda^{(i)}_\beta)$. In this case, the real Jordan form is

$$J^{(i)}_\beta = \begin{pmatrix} W^{(i)}_\beta & I_2 & & 0 \\ & \ddots & \ddots & \\ & & W^{(i)}_\beta & I_2 \\ 0 & & & W^{(i)}_\beta \end{pmatrix} \in \mathbb{R}^{2b^{(i)}_\beta \times 2b^{(i)}_\beta}, \tag{6.4.1}$$

$$W^{(i)}_\beta = \begin{pmatrix} \Re(\lambda^{(i)}_\beta) & \Im(\lambda^{(i)}_\beta) \\ -\Im(\lambda^{(i)}_\beta) & \Re(\lambda^{(i)}_\beta) \end{pmatrix}, \; I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6.4.2}$$

Firstly, we present the main findings of this section in Theorem 12 which is shown afterwards.

Theorem 12. *Given three dynamical LTI systems $(\boldsymbol{P}^{(0)}, \boldsymbol{P}^{(1)}, \boldsymbol{P}^{(2)})$ in (6.1.1a)-(6.1.1b) that are coded and estimated over a two-input single-output MAC with common message (Def. 8) such that the systems' topological*

entropy vector $\boldsymbol{H} := (\boldsymbol{H}^{(0)}, \boldsymbol{H}^{(1)}, \boldsymbol{H}^{(2)})^T$ *in (6.1.8) lies within* $\mathrm{int}(\mathscr{C}_0)$. *Furthermore, the state matrix* $\boldsymbol{A}^{(i)}$ *(6.1.2) consists of non-scalar Jordan blocks of the form (6.4.1), and the following conditions hold*

$$r_{\gamma,\beta}^{(i)} > (2b_\beta^{(i)} - 1)\frac{\log n}{n} + \log|\lambda_\beta^{(i)}| + \frac{1}{n}\log\left(\frac{1-|\lambda_\beta^{(i)}|^{-2b_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^{-1}}\right), \tag{6.4.3a}$$

$$R^{(i)} \geq \sum_{\beta=1}^{B^{(i)}} \sum_{\gamma=1}^{b_\beta^{(i)}} r_{\gamma,\beta}^{(i)}, \ \forall \gamma \in [1 : b_\beta^{(i)}], \ \beta \in [1 : B^{(i)}]. \tag{6.4.3b}$$

Then, we have

$$\begin{aligned}\lim_{k\to\infty}\|E^{(i)}(nk)\| &\leq \max_{\beta\in[1:B^{(i)}]}\max_{\gamma\in[1:b_\beta^{(i)}]}\left\{\lim_{k\to\infty}\|E_{\gamma,\beta}^{(i)}(nk)\|\right\} \\ &\leq \max_{\beta\in[1:B^{(i)}]}\max_{\gamma\in[1:b_\beta^{(i)}]}\left\{\frac{1-|\lambda_\beta^{(i)}|^{-2b_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^{-1}}\frac{(n-1)^{2b_\beta^{(i)}-1}}{1-\frac{1-|\lambda_\beta^{(i)}|^{-2b_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^{-1}}n^{2b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n 2^{-nr_{\gamma,\beta}^{(i)}}}\cdot\right. \\ &\qquad \left.\frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}\right\},\end{aligned} \tag{6.4.4}$$

with $\bar{V}^{(i)} := \sup[\![\|V^{(i)}\|]\!]$.

Remark 16. *When considering the more general case of Jordan blocks with complex-valued eigenvalues, the condition on the rates* $r_{\gamma,\beta}^{(i)}$ *(6.4.3a) slightly changes in comparison with (6.3.1a). Although the main observations mentioned in Remark 15 remain valid here as well, the Jordan block size that arises in this expression is now multiplied by a factor of 2.*

Proof. We start by noting that $\boldsymbol{W}_\beta^{(i)}$ (6.4.2) is given by

$$\boldsymbol{W}_\beta^{(i)} = \boldsymbol{S}\boldsymbol{D}(\lambda_\beta^{(i)})\boldsymbol{S}^{-1}, \tag{6.4.5}$$

where, as explained in [41, §3.4.1], we have

$$D(\lambda_\beta^{(i)}) = \begin{pmatrix} \lambda_\beta^{(i)} & \\ & \bar{\lambda}_\beta^{(i)} \end{pmatrix}, \tag{6.4.6}$$

$$S = \begin{pmatrix} -j & -j \\ 1 & -1 \end{pmatrix}, \tag{6.4.7}$$

$$S^{-1} = \frac{1}{2j} \begin{pmatrix} -1 & j \\ -1 & -j \end{pmatrix}, \tag{6.4.8}$$

such that $\bar{\lambda}_\beta^{(i)}$ denotes the complex conjugate of $\lambda_\beta^{(i)}$. Hence, we can write $J_\beta^{(i)}$ as follows

$$\begin{aligned} J_\beta^{(i)} &= W_\beta^{(i)} \otimes I_{b_\beta^{(i)}} + N_{2b_\beta^{(i)}} \\ &= SD(\lambda_\beta^{(i)})S^{-1} \otimes I_{b_\beta^{(i)}} + N_{2b_\beta^{(i)}}, \end{aligned} \tag{6.4.9}$$

where $\otimes$ refers to the Kronecker product. By raising $J_\beta^{(i)}$ to the power of n and computing the max-norm, we obtain

$$\begin{aligned} \left\| \left(J_\beta^{(i)}\right)^n \right\| &= \left\| \sum_{\xi=0}^{n} \binom{n}{\xi} \left(W_\beta^{(i)} \otimes I_{b_\beta^{(i)}}\right)^\xi N_{2b_\beta^{(i)}}^{n-\xi} \right\| \\ &= \left\| \sum_{\xi=n-2b_\beta^{(i)}+1}^{n} \binom{n}{\xi} \left(W_\beta^{(i)} \otimes I_{b_\beta^{(i)}}\right)^\xi N_{2b_\beta^{(i)}}^{n-\xi} \right\| \qquad (6.4.10) \\ &\leq \sum_{\xi=n-2b_\beta^{(i)}+1}^{n} \binom{n}{i} \left\| W_\beta^{(i)} \otimes I_{b_\beta^{(i)}} \right\|^\xi \left\| N_{2b_\beta^{(i)}}^{n-\xi} \right\| \\ &= \sum_{r=0}^{2b_\beta^{(i)}-1} \binom{n}{n-r} \left\| W_\beta^{(i)} \otimes I_{b_\beta^{(i)}} \right\|^{n-r} \left\| N_{2b_\beta^{(i)}}^{r} \right\| \qquad (6.4.11) \\ &\leq \sum_{r=0}^{2b_\beta^{(i)}-1} \binom{n}{n-r} \left\| W_\beta^{(i)} \otimes I_{b_\beta^{(i)}} \right\|^{n-r}, \qquad (6.4.12) \end{aligned}$$

where (6.4.10) holds because $N_{2b_\beta^{(i)}}^n = 0$, for all $n \geq 2b_\beta^{(i)}$, and (6.4.11) follows from changing the variables $r = n - \xi$.

Observe that $\|\cdot\|$ is the max-norm and $I_{b_\beta^{(i)}}$ is the identity matrix of dimension $b_\beta^{(i)} \times b_\beta^{(i)}$. Thus,

$$\begin{aligned}\left\|W_\beta^{(i)} \otimes I_{b_\beta^{(i)}}\right\| &= \left\|W_\beta^{(i)}\right\| \\ &= \left\|SD(\lambda_\beta^{(i)})S^{-1}\right\| \\ &= \left\|D(\lambda_\beta^{(i)})\right\| \\ &= \max\{|\lambda_\beta^{(i)}|, |\bar{\lambda}_\beta^{(i)}|\} \\ &= |\lambda_\beta^{(i)}|. \end{aligned} \tag{6.4.13}$$

Hence, we have

$$\left\|\left(J_\beta^{(i)}\right)^n\right\| \leq \sum_{r=0}^{2b_\beta^{(i)}-1} \binom{n}{n-r} |\lambda_\beta^{(i)}|^{n-r}. \tag{6.4.14}$$

Using (6.3.12) for each $r \in [0 : 2b_\beta^{(i)} - 1]$, we obtain

$$\begin{aligned}\left\|\left(J_\beta^{(i)}\right)^n\right\| &\leq \sum_{r=0}^{2b_\beta^{(i)}-1} \binom{n}{n-r} |\lambda_\beta^{(i)}|^{n-r} \\ &\leq \sum_{r=0}^{2b_\beta^{(i)}-1} \left|\frac{n^r(\lambda_\beta^{(i)})^n}{r!(\lambda_\beta^{(i)})^r}\right| \\ &\leq n^{2b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n \sum_{r=0}^{2b_\beta^{(i)}-1} \frac{|\lambda_\beta^{(i)}|^{-r}}{r!} \\ &\leq n^{2b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n \sum_{r=0}^{2b_\beta^{(i)}-1} |\lambda_\beta^{(i)}|^{-r} \\ &= n^{2b_\beta^{(i)}-1}|\lambda_\beta^{(i)}|^n \frac{1-|\lambda_\beta^{(i)}|^{-2b_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^{-1}}. \end{aligned} \tag{6.4.15}$$

Hence, from (6.4.15) we can identify K as

$$K = \frac{1-|\lambda_\beta^{(i)}|^{-2b_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^{-1}}. \tag{6.4.16}$$

By virtue of (6.4.15), the upper bound on the time-asymptotic estimation error $\|E_{\gamma,\beta}^{(i)}(nk)\|$ (6.3.3)

then becomes

$$\|E^{(i)}_{\gamma,\beta}(nk)\| \leq \left(n^{2b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n \frac{1-|\lambda^{(i)}_\beta|^{-2b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} 2^{-nr^{(i)}_{\gamma,\beta}} \right)^k \ell^{(i)}_{\gamma,\beta}(0) + (n-1)^{2b^{(i)}_\beta-1} \frac{\left(\frac{1-|\lambda^{(i)}_\beta|^{-2b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} n^{2b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}} \right)^k - 1}{\frac{1-|\lambda^{(i)}_\beta|^{-2b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} n^{2b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}} - 1} \frac{1-|\lambda^{(i)}_\beta|^{-2b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} \frac{|\lambda^{(i)}_\beta|^n-1}{|\lambda^{(i)}_\beta|-1} \bar{V}^{(i)}. \tag{6.4.17}$$

Notice that (6.4.17) converges when

$$\begin{aligned} & 1 > n^{2b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n \frac{1-|\lambda^{(i)}_\beta|^{-2b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} 2^{-nr^{(i)}_{\gamma,\beta}} \\ \Leftrightarrow \quad & nr^{(i)}_{\gamma,\beta} > (2b^{(i)}_\beta-1)\log n + n\log|\lambda^{(i)}_\beta| + \log\left(\frac{1-|\lambda^{(i)}_\beta|^{-2b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} \right). \end{aligned} \tag{6.4.18}$$

Hence, if (6.4.18) is satisfied, then the upper bound on the time-asymptotic max-norm of the estimation error associated with the i-th system turns out to be

$$\begin{aligned} \lim_{k\to\infty}\|E^{(i)}(nk)\| &\leq \max_{\beta\in[1:B^{(i)}]} \max_{\gamma\in[1:b^{(i)}_\beta]} \left\{ \lim_{k\to\infty}\|E^{(i)}_{\gamma,\beta}(nk)\| \right\} \\ &\leq \max_{\beta\in[1:B^{(i)}]} \max_{\gamma\in[1:b^{(i)}_\beta]} \left\{ \frac{1-|\lambda^{(i)}_\beta|^{-2b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} \frac{(n-1)^{2b^{(i)}_\beta-1}}{1-\frac{1-|\lambda^{(i)}_\beta|^{-2b^{(i)}_\beta}}{1-|\lambda^{(i)}_\beta|^{-1}} n^{2b^{(i)}_\beta-1}|\lambda^{(i)}_\beta|^n 2^{-nr^{(i)}_{\gamma,\beta}}} \cdot \right. \\ &\qquad \left. \frac{|\lambda^{(i)}_\beta|^n-1}{|\lambda^{(i)}_\beta|-1} \bar{V}^{(i)} \right\}. \end{aligned} \tag{6.4.19}$$

Thus, the proof of Theorem 12 is completed. □

6.5 The Optimal Rate Allocation Policy for Scalar Jordan Blocks

In the sequel, we restrict our attention to the case where the Jordan blocks are *scalars*, i.e., $b^{(i)}_\beta = 1$. The objective now is to find the rate allocation minimizing (6.2.18) over $r^{(i)}_\beta$ while satisfying

(6.1.9a)-(6.1.9b). We start by presenting the main result of this section in Theorem 13, and then, proceed to proving it in Subsections 6.5.1 and 6.5.2.

Theorem 13. *Consider a system* $\mathbf{P}^{(i)}$ *with state matrix* $A^{(i)} = \mathrm{diag}(\lambda_1^{(i)}, \cdots, \lambda_{B^{(i)}}^{(i)})$, *topological entropy* $H^{(i)}$ *in (6.1.8), disturbed by process noise* $V^{(i)}$, *and* $\bar{V}^{(i)}$ *is defined in (6.2.4). Furthermore, assume that the* β*-th state* $X_\beta^{(i)}$, *with* $\beta \in [1 : B^{(i)}]$, *is processed by means of an adaptive uniform quantizer and estimated at a rate* $r_\beta^{(i)}$. *Let* n *be the code block-length,* $R^{(i)}$ *the total available rate and* $\delta^{(i)}$ *a non-negative pull-back factor such that*

$$R^{(i)} - \delta^{(i)} > H^{(i)}. \tag{6.5.1}$$

Then, the optimum rate allocation can be determined via a two-stage process:

a) Firstly, solve the following implicit equation with respect to $U^{(i)}$

$$H^{(i)} - R^{(i)} + \delta^{(i)} = \frac{1}{n} \sum_{\beta=1}^{B^{(i)}} \log_2 \left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}} \right), \tag{6.5.2}$$

with $U^{(i)} \geq \frac{|\lambda_{\max}^{(i)}|^n - 1}{|\lambda_{\max}^{(i)}| - 1} \bar{V}^{(i)}$.

b) Afterwards, compute the rates $r_\beta^{(i)}$ *using*

$$r_\beta^{(i)} = \log_2 |\lambda_\beta^{(i)}| - \frac{1}{n} \log_2 \left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}} \right), \quad \forall \beta \in [1 : B^{(i)}]. \tag{6.5.3}$$

In the next part, we prove the convexity of the RHS of (6.2.18). Subsequently, we formulate the rate allocation problem as a constrained convex optimization problem whose solution(s) represent the optimal strategy (Theorem 13).

6.5.1 Convexity of the RHS of (6.2.18)

Observe that the cost function in terms of $r_\beta^{(i)}$ in the RHS of (6.2.18) has the following form

$$f(x) = L \cdot \frac{1 + ae^{-bx}}{1 - ae^{-bx}}, \tag{6.5.4}$$

where $L > 0$, $a > 1$ (as it is assumed that $|\lambda_\beta^{(i)}| > 1$) and $b \geq 1$. To study the convexity of $f(x)$, we compute the second derivative $f''(x)$ (f is twice differentiable on the interval of interest $(\log(a)/b, \infty)$). Thus, we have

$$f'(x) = -2abL\frac{e^{-bx}}{(1-ae^{-bx})^2}. \tag{6.5.5}$$

Subsequently, the second derivative of f is

$$f''(x) = 2ab^2Le^{-bx}\frac{(1+ae^{-bx})}{(1-ae^{-bx})^3}. \tag{6.5.6}$$

When $x > \frac{\log(a)}{b}$, $f''(x) > 0$, and hence, we conclude that f is convex on the interval of interest.

6.5.2 Analysis of the Optimization Problem

Recall that in this particular analysis, we are assuming $b_\beta^{(i)} = 1$, with $\beta \in [1 : B^{(i)}]$ and $i \in \{0,1,2\}$. Define the function $\mathfrak{J}(r_\beta^{(i)})$ such that

$$\mathfrak{J}(r_\beta^{(i)}) := \frac{1+|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}}} \cdot \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}. \tag{6.5.7}$$

Then, using an auxiliary cost variable $U^{(i)}$, we formulate the following convex optimization problem, for $i \in \{0,1,2\}$,

$$\min \ U^{(i)} \tag{6.5.8a}$$

$$\text{s. t. } U^{(i)} \geq \mathfrak{J}(r_\beta^{(i)}), \ \forall \beta \in [1 : B^{(i)}], \tag{6.5.8b}$$

$$r_\beta^{(i)} - \varepsilon^{(i)} \geq \log_2 |\lambda_\beta^{(i)}|, \ \forall \beta \in [1 : B^{(i)}], \tag{6.5.8c}$$

$$R^{(i)} - \delta^{(i)} \geq \sum_{\beta=1}^{B^{(i)}} r_\beta^{(i)}, \tag{6.5.8d}$$

where the pull-back factors $\varepsilon^{(i)} > 0$ and $\delta^{(i)} \geq 0$ are constants. Subsequently, we write

$$\min \; U^{(i)} \tag{6.5.9a}$$

$$\text{s. t. } U^{(i)} - \mathfrak{J}(r_\beta^{(i)}) \geq 0, \; \forall \beta \in [1 : B^{(i)}], \tag{6.5.9b}$$

$$r_\beta^{(i)} - \varepsilon^{(i)} - \log_2 |\lambda_\beta^{(i)}| \geq 0, \; \forall \beta \in [1 : B^{(i)}], \tag{6.5.9c}$$

$$R^{(i)} - \delta^{(i)} - \sum_{\beta=1}^{B^{(i)}} r_\beta^{(i)} \geq 0, \tag{6.5.9d}$$

To derive the Karush-Kuhn-Tucker (KKT) conditions, we form the Lagrangian $\mathbf{L}(r_\beta^{(i)}, p)$ as follows

$$\begin{aligned}\mathbf{L}(r_\beta^{(i)}, p) = U^{(i)} &- \sum_{\beta=1}^{B^{(i)}} p_{1,\beta} \left(U^{(i)} - \mathfrak{J}(r_\beta^{(i)}) \right) - \sum_{\beta=1}^{B^{(i)}} p_{2,\beta} \left(r_\beta^{(i)} - \varepsilon^{(i)} - \log_2 |\lambda_\beta^{(i)}| \right) \\ &- p_3 \left(R^{(i)} - \delta^{(i)} - \sum_{\beta=1}^{B^{(i)}} r_\beta^{(i)} \right).\end{aligned} \tag{6.5.10}$$

Thus, the KKT conditions consist of

$$\frac{\partial \mathbf{L}}{\partial r_\beta^{(i)}} = 0, \; \forall \beta \in [1 : B^{(i)}], \tag{6.5.11a}$$

$$\frac{\partial \mathbf{L}}{\partial U^{(i)}} = 0, \tag{6.5.11b}$$

$$U^{(i)} - \mathfrak{J}(r_\beta^{(i)}) \geq 0, \; \forall \beta \in [1 : B^{(i)}], \tag{6.5.11c}$$

$$r_\beta^{(i)} - \varepsilon^{(i)} - \log_2 |\lambda_\beta^{(i)}| \geq 0, \; \forall \beta \in [1 : B^{(i)}], \tag{6.5.11d}$$

$$R^{(i)} - \delta^{(i)} - \sum_{\beta=1}^{B^{(i)}} r_\beta^{(i)} \geq 0, \tag{6.5.11e}$$

$$p_{1,\beta} \left(U^{(i)} - \mathfrak{J}(r_\beta^{(i)}) \right) = 0, \; \forall \beta \in [1 : B^{(i)}], \tag{6.5.11f}$$

$$p_{2,\beta} \left(r_\beta^{(i)} - \varepsilon^{(i)} - \log_2 |\lambda_\beta^{(i)}| \right) = 0, \; \forall \beta \in [1 : B^{(i)}], \tag{6.5.11g}$$

$$p_3 \left(R^{(i)} - \delta^{(i)} - \sum_{\beta=1}^{B^{(i)}} r_\beta^{(i)} \right) = 0, \tag{6.5.11h}$$

$$p_{1,\beta}, \; p_{2,\beta}, \; p_3 \geq 0, \; \forall \beta \in [1 : B^{(i)}]. \tag{6.5.11i}$$

Condition (6.5.11a) yields

$$\begin{aligned}\frac{\partial \mathbf{L}}{\partial r_\beta^{(i)}} &= p_{1,\beta}\frac{\partial \mathfrak{J}(r_\beta^{(i)})}{\partial r_\beta^{(i)}} - p_{2,\beta} + p_3 \\ &= \underbrace{-2n|\lambda_\beta^{(i)}|\ln(2)\bar{V}^{(i)}\frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}}_{:=C\left(\lambda_\beta^{(i)},n,\bar{V}^{(i)}\right)}\frac{2^{-nr_\beta^{(i)}}}{\left(1-|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}}\right)^2}p_{1,\beta} - p_{2,\beta} + p_3 \\ &\overset{!}{=} 0. \end{aligned} \tag{6.5.12}$$

Additionally, from condition (6.5.11b) we obtain the following relation

$$\frac{\partial \mathbf{L}}{\partial U^{(i)}} = 1 - \sum_{\beta=1}^{B^{(i)}} p_{1,\beta} \overset{!}{=} 0, \tag{6.5.13}$$

which means that

$$\sum_{\beta=1}^{B^{(i)}} p_{1,\beta} \overset{!}{=} 1. \tag{6.5.14}$$

We now consider the following different cases:

- **Case 1:** Suppose that $p_3 = 0$. In this case, (6.5.12) leads to the following:

$$p_{1,\beta} = \frac{1}{C\left(\lambda_\beta^{(i)},n,\bar{V}^{(i)}\right)}\frac{\left(1-|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}}\right)^2}{2^{-nr_\beta^{(i)}}}p_{2,\beta}. \tag{6.5.15}$$

  Note that $C(\lambda_\beta^{(i)},n,\bar{V}^{(i)}) < 0$ for $|\lambda_\beta^{(i)}| > 1$. Hence, $p_{1,\beta}$ and $p_{2,\beta}$ have opposite signs. However, condition (6.5.11i) tells that $p_{1,\beta} \geq 0$ and $p_{2,\beta} \geq 0$. This is, therefore, only possible when $p_{1,\beta} = p_{2,\beta} = 0$, for all $\beta \in [1 : B^{(i)}]$. Nevertheless, this conclusion contradicts with (6.5.14), and consequently, with condition (6.5.11b). Thus, we conclude that this case is impossible.

- **Case 2:** Suppose now that $p_3 > 0$ and $p_{2,\beta} = 0, \forall \beta \in [1 : B^{(i)}]$. Hence, (6.5.12) yields

$$\begin{aligned} p_{1,\beta} &= -\frac{1}{C\left(\lambda_\beta^{(i)},n,\bar{V}^{(i)}\right)}\frac{\left(1-|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}}\right)^2}{2^{-nr_\beta^{(i)}}}p_3 \\ &\geq 0. \end{aligned} \tag{6.5.16}$$

Furthermore, condition (6.5.11f) implies that either $p_{1,\beta} = 0$ or $U^{(i)} - \mathfrak{J}(r_\beta^{(i)}) = 0$.

Observe that, since $p_3 > 0$, $p_{1,\beta} = 0$ if and only if

$$r_\beta^{(i)} = \log|\lambda_\beta^{(i)}|. \tag{6.5.17}$$

However, this cannot be the case because of condition (6.5.11d). On the other, $U^{(i)} - \mathfrak{J}(r_\beta^{(i)}) = 0$ means that

$$U^{(i)} = \frac{1+|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}}}{1-|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}}} \cdot \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}, \ \ \forall \beta \in [1 : B^{(i)}]. \tag{6.5.18}$$

Consequently, we obtain

$$\begin{aligned}
&(1-|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}})U^{(i)} = (1+|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}})\frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)} \\
\Leftrightarrow &|\lambda_\beta^{(i)}|^n 2^{-nr_\beta^{(i)}} = \frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}} \\
\Leftrightarrow &-nr_\beta^{(i)} = \log_2\left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}}\right) - n\log_2|\lambda_\beta^{(i)}|,
\end{aligned} \tag{6.5.19}$$

where Step (6.5.19) is valid only if

$$\begin{aligned}
U^{(i)} &\geq \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}, \ \forall \beta \in [1 : B^{(i)}], \\
\Leftrightarrow U^{(i)} &\geq \frac{|\lambda_{\max}^{(i)}|^n - 1}{|\lambda_{\max}^{(i)}| - 1}\bar{V}^{(i)},
\end{aligned} \tag{6.5.20}$$

with $|\lambda_{\max}^{(i)}| = \max_{\beta \in [1:B^{(i)}]}\{|\lambda_\beta^{(i)}|\}$. Hence, we obtain the following expression of the rates $r_\beta^{(i)}$

$$r_\beta^{(i)} = \log_2|\lambda_\beta^{(i)}| - \frac{1}{n}\log_2\left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}}\right). \tag{6.5.21}$$

Because of condition (6.5.11d), (6.5.21) is valid only in the case where

$$\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}} \leq 1, \tag{6.5.22}$$

which is true for $|\lambda_\beta^{(i)}| > 1$, since then $\frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)} \geq 0$.

For condition (6.5.11h) to hold, we must have

$$R^{(i)} - \delta^{(i)} - \sum_{\beta=1}^{B^{(i)}} r_\beta^{(i)} = 0, \; \forall \beta \in [1 : B^{(i)}].$$

Furthermore, from (6.5.21) we know that

$$\begin{aligned}\sum_{\beta=1}^{B^{(i)}} r_\beta^{(i)} &= \sum_{\beta=1}^{B^{(i)}} \left(\log_2 |\lambda_\beta^{(i)}| - \frac{1}{n} \log_2 \left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}} \right) \right) \\ &= H^{(i)} - \frac{1}{n} \sum_{\beta=1}^{B^{(i)}} \underbrace{\log_2 \left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}} \right)}_{\leq 0}, \end{aligned} \tag{6.5.23}$$

where Step (6.5.23) follows from the definition of $H^{(i)}$ in (6.1.8) and that the considered eigenvalues in this setup have all a magnitude larger than 1 (unstable states). Therefore, we obtain

$$H^{(i)} - R^{(i)} + \delta^{(i)} = \frac{1}{n} \sum_{\beta=1}^{B^{(i)}} \log_2 \left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1}\bar{V}^{(i)}} \right). \tag{6.5.24}$$

The problem parameters and their domains of definition are summarized in Table 6.1 and Table 6.2, respectively.

6.5.3 Analysis of the Large n Regime

In this part, we study the behavior of the solutions of the rate allocation problem for asymptotically large code block-lengths.

Fixed Parameters	Optimization Variables
$H^{(i)}$	$U^{(i)}$ (6.5.24)
$R^{(i)}$	$r_\beta^{(i)}$ (6.5.21)
$\delta^{(i)}$	-
$\lambda_\beta^{(i)}$	-
$\bar{V}^{(i)}$	-
n	-
$B^{(i)}$	-

Table 6.1: Summary of the problem parameters and their expressions, with $\beta \in [1 : B^{(i)}]$ and $i \in \{0,1,2\}$.

Parameter	Domain
$U^{(i)}$	$\left[\frac{\lvert\lambda_{\max}^{(i)}\rvert^n - 1}{\lvert\lambda_{\max}^{(i)}\rvert - 1}\bar{V}^{(i)}, \infty\right)$
$r_\beta^{(i)}$	$\left(\log_2 \lvert\lambda_\beta^{(i)}\rvert, R^{(i)} - \delta^{(i)}\right]$

Table 6.2: Domains of definition of $U^{(i)}$ and $r_\beta^{(i)}$, with $\beta \in [1 : B^{(i)}]$ and $i \in \{0,1,2\}$.

Lemma 6. *Consider a system $\mathbf{P}^{(i)}$ with state matrix $A^{(i)} = \operatorname{diag}(\lambda_1^{(i)}, \cdots, \lambda_{B^{(i)}}^{(i)})$, topological entropy $H^{(i)}$ in (6.1.8), disturbed by process noise $V^{(i)}$, and $\bar{V}^{(i)}$ is defined in (6.2.4). Furthermore, assume that the β-th state $X_\beta^{(i)}$, with $\beta \in [1 : B^{(i)}]$, is processed by means of an adaptive uniform quantizer and estimated at a rate $r_\beta^{(i)}$. Let $R^{(i)}$ be the total available rate and $\delta^{(i)}$ a non-negative pull-back factor such that*

$$R^{(i)} - \delta^{(i)} > H^{(i)}. \tag{6.5.25}$$

Moreover, let σ be

$$\sigma := -(H^{(i)} - R^{(i)} + \delta^{(i)}). \tag{6.5.26}$$

For asymptotically large code block-lengths, i.e., $n \to \infty$, and when $n\sigma \gg 1$, the cost variable $U^{(i)}$ (6.5.2) is given by

$$U^{(i)} = \frac{\lvert\lambda_{\max}^{(i)}\rvert^n}{\lvert\lambda_{\max}^{(i)}\rvert - 1}\bar{V}^{(i)} + O\left(\frac{\lvert\lambda_{\max}^{(i)}\rvert^n}{2^{n\sigma/\kappa_{\max}}}\right), \tag{6.5.27}$$

and the rates in this regime are allocated as follows

$$r_\beta^{(i)} = \begin{cases} \log_2 |\lambda_{\max}^{(i)}| + \frac{\sigma}{\kappa_{\max}} + o\left(\frac{1}{n}\right), & \text{if } |\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|, \\ \log_2 |\lambda_\beta^{(i)}| + o\left(\frac{1}{n}\right), & \text{if } |\lambda_\beta^{(i)}| < |\lambda_{\max}^{(i)}|, \end{cases} \tag{6.5.28}$$

with $\kappa_{\max}$ *the total number of* $\lambda_\beta^{(i)}$ *such that* $|\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|$.

Remark 17. *The result obtained in Lemma 6 provides a set of guidelines for the system designer to allocate the available data resources in an efficient manner, while maintaining the estimation error at the receiver side bounded, and hence, ensuring a reliable state estimation. The rate allocation strategy suggests the following:*

- *To reliably estimate states associated with the dominant eigenvalue, the dedicated rate must be slightly larger, i.e., by* $o(1/n)$*, than the sum of the eigenvalue's log-magnitude and the back-off factor* σ *equally divided among the corresponding states, and thus, these states are estimated using equal rates.*

- *Additionally, in order to estimate the rest of the states, it is enough to allocate a rate strictly larger than the log-magnitude of the associated eigenvalue.*

Proposition 1. *If* $n\sigma \gg 1$ *and the process noise* $V^{(i)}$ *has a hypercuboidal structure, then the lower and upper bounds on the max-norm of the estimation error given by (5.6.19) and (6.5.27) respectively, coincide. Hence, we have*

$$\limsup_{k\to\infty} [\![\|E^{(i)}(nk)\|]\!] = \frac{|\lambda_{\max}^{(i)}|^n}{|\lambda_{\max}^{(i)}| - 1} \bar{V}^{(i)} + o(|\lambda_{\max}^{(i)}|^n), \tag{6.5.29}$$

where $\bar{V}^{(i)}$ *is defined in the sense of (6.2.4).*

We devote the rest of this section to prove Lemma 6. As $n \to \infty$, then by (6.5.24)

$$\frac{1}{n} \sum_{\beta=1}^{B^{(i)}} \log_2 \left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}} \right) = \underbrace{H^{(i)} - R^{(i)} + \delta^{(i)}}_{\text{const.}} < 0. \tag{6.5.30}$$

Let σ be a fixed strictly positive real number such that

$$\sigma := -(H^{(i)} - R^{(i)} + \delta^{(i)}). \tag{6.5.31}$$

Hence, the following must hold

$$\sum_{\beta=1}^{B^{(i)}} \log_2 \left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}} \right) = -n\sigma \overset{n \to \infty}{\longrightarrow} -\infty. \tag{6.5.32}$$

For (6.5.32) to be satisfied, there exists at least one $\beta \in [1 : B^{(i)}]$ for which it holds

$$U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)} \ll U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}, \text{ for large } n. \tag{6.5.33}$$

On the other hand, because of (6.5.20), (6.5.33) can be valid only for β such that $|\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|$, where $\lambda_{\max}^{(i)}$ is the dominant eigenvalue, i.e., with the largest magnitude. Subsequently, in the large n regime, we can approximate $U^{(i)}$ as follows

$$U^{(i)} = \frac{|\lambda_{\max}^{(i)}|^n - 1}{|\lambda_{\max}^{(i)}| - 1} \bar{V}^{(i)} + \psi_n, \tag{6.5.34}$$

where ψ_n is a residual error term that depends on n. Then, (6.5.32) becomes

$$\begin{aligned}
-n\sigma &= \sum_{\beta=1}^{B^{(i)}} \log_2 \left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}} \right) \\
&\overset{(6.5.34)}{=} \sum_{\beta : |\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|} \log_2 \left(\frac{\psi_n}{2 \left(\frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)} \right) + \psi_n} \right) + \\
&\qquad \sum_{\beta : |\lambda_\beta^{(i)}| \neq |\lambda_{\max}^{(i)}|} \log_2 \left(\frac{\left(\frac{|\lambda_{\max}^{(i)}|^n - 1}{|\lambda_{\max}^{(i)}| - 1} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \right) \bar{V}^{(i)} + \psi_n}{\left(\frac{|\lambda_{\max}^{(i)}|^n - 1}{|\lambda_{\max}^{(i)}| - 1} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \right) \bar{V}^{(i)} + \psi_n} \right) \\
&= \sum_{\beta : |\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|} \log_2 \left(\frac{\psi_n}{2 \left(|\lambda_\beta^{(i)}|^n - 1 \right) \bar{V}^{(i)} + \psi_n} \right) + \underbrace{\sum_{\beta : |\lambda_\beta^{(i)}| \neq |\lambda_{\max}^{(i)}|} \log_2 \left(\frac{\left(1 - \frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n} \right) \bar{V}^{(i)} + \psi_n}{\left(1 + \frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n} \right) \bar{V}^{(i)} + \psi_n} \right)}_{\approx O\left(\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n} \right) = o(1)}
\end{aligned}$$

$$= \sum_{\beta:|\lambda_\beta^{(i)}|=|\lambda_{\max}^{(i)}|} \log_2\left(\frac{\psi_n}{2|\lambda_\beta^{(i)}|^n \bar{V}^{(i)} + \psi_n}\right) + o(1). \tag{6.5.35}$$

When σ is kept constant and $n \to \infty$, then $n\sigma \gg 1$. From (6.5.35), one can see that this regime implies that $2|\lambda_\beta^{(i)}|^n \bar{V}^{(i)}$ is the dominant term in the denominator of the fraction $\frac{\psi_n}{2|\lambda_\beta^{(i)}|^n \bar{V}^{(i)} + \psi_n}$. Hence, $\psi_n \ll |\lambda_{\max}^{(i)}|^n$. Suppose that ψ_n can be written as

$$\psi_n := c(\rho_n)^n, \tag{6.5.36}$$

where ρ_n is a real number that might depend on n. Thus, we obtain the following

$$\begin{aligned} -n\sigma &\overset{(6.5.35)}{=} \sum_{\beta:|\lambda_\beta^{(i)}|=|\lambda_{\max}^{(i)}|} \log_2\left(\frac{\psi_n}{2|\lambda_\beta^{(i)}|^n \bar{V}^{(i)}}\right) + o(1) \\ &\overset{(6.5.36)}{=} \sum_{\beta:|\lambda_\beta^{(i)}|=|\lambda_{\max}^{(i)}|} \left(n \log_2\left(\frac{\rho_n}{|\lambda_\beta^{(i)}|}\right) - \log_2\left(\frac{2\bar{V}^{(i)}}{c}\right)\right) + o(1). \end{aligned}$$

Therefore, the following result holds

$$-n\sigma = n\kappa_{\max} \log_2\left(\frac{\rho_n}{|\lambda_{\max}^{(i)}|}\right) - \kappa_{\max} \log_2\left(\frac{2\bar{V}^{(i)}}{c}\right) + o(1), \tag{6.5.37}$$

where $\kappa_{\max}$ is the sum of algebraic multiplicities of the eigenvalues $\lambda_\beta^{(i)}$ such that $|\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|$. This yields

$$\begin{aligned} \kappa_{\max} \log_2\left(\frac{\rho_n}{|\lambda_{\max}^{(i)}|}\right) &= -\sigma + \frac{\kappa_{\max}}{n} \log_2\left(\frac{2\bar{V}^{(i)}}{c}\right) + o\left(\frac{1}{n}\right) \\ \Leftrightarrow\ \rho_n &= |\lambda_{\max}^{(i)}| 2^{-\sigma/\kappa_{\max}} \cdot 2^{\log_2\left(\frac{2\bar{V}^{(i)}}{c}\right)/n} \cdot 2^{o(1/n)}. \end{aligned} \tag{6.5.38}$$

By expressing ψ_n (6.5.36) in terms of ρ_n (6.5.38) and substituting it in (6.5.34), we get

$$\begin{aligned} U^{(i)} &= \frac{|\lambda_{\max}^{(i)}|^n - 1}{|\lambda_{\max}^{(i)}| - 1} \bar{V}^{(i)} + 2^{-n\sigma/\kappa_{\max}} |\lambda_{\max}^{(i)}|^n \cdot \frac{2\bar{V}^{(i)}}{c} \cdot 2^{o(1)} \\ &= \frac{|\lambda_{\max}^{(i)}|^n - 1}{|\lambda_{\max}^{(i)}| - 1} \bar{V}^{(i)} + \left(2^{-n\sigma/\kappa_{\max}} |\lambda_{\max}^{(i)}|^n \cdot \frac{2\bar{V}^{(i)}}{c}\right)(1 + o(1)) \end{aligned}$$

$$= \frac{|\lambda_{\max}^{(i)}|^n - 1}{|\lambda_{\max}^{(i)}| - 1}\bar{V}^{(i)} + 2^{-n\sigma/\kappa_{\max}}|\lambda_{\max}^{(i)}|^n \cdot \frac{2\bar{V}^{(i)}}{c} + o\left(\frac{|\lambda_{\max}^{(i)}|^n}{2^{n\sigma/\kappa_{\max}}}\right) \tag{6.5.39}$$

$$= \frac{|\lambda_{\max}^{(i)}|^n}{|\lambda_{\max}^{(i)}| - 1}\bar{V}^{(i)} + O\left(\frac{|\lambda_{\max}^{(i)}|^n}{2^{n\sigma/\kappa_{\max}}}\right). \tag{6.5.40}$$

In this regime, where $n \to \infty$ and $n\sigma \gg 1$, the dominant term in (6.5.40) is $\frac{|\lambda_{\max}^{(i)}|^n}{|\lambda_{\max}^{(i)}|-1}\bar{V}^{(i)}$. We then use this result to develop an optimum rate allocation policy for this regime.

Rate Allocation Rule. Subsequently, by virtue of (6.5.21), we obtain the following rate allocations in this regime, with $i \in \{0,1,2\}$ and $\forall \beta \in [1 : B^{(i)}]$

$$r_\beta^{(i)} = \log_2|\lambda_\beta^{(i)}| - \frac{1}{n}\log_2\left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}|-1}\bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}|-1}\bar{V}^{(i)}}\right)$$

$$\overset{(6.5.39)}{=} \log_2|\lambda_\beta^{(i)}| - \frac{1}{n}\log_2\left(\frac{\left(\frac{|\lambda_{\max}^{(i)}|^n}{|\lambda_{\max}^{(i)}|-1} - \frac{|\lambda_\beta^{(i)}|^n}{|\lambda_\beta^{(i)}|-1}\right)\bar{V}^{(i)} + 2^{-n\sigma/\kappa_{\max}}|\lambda_{\max}^{(i)}|^n \cdot 2\bar{V}^{(i)} + o\left(\frac{|\lambda_{\max}^{(i)}|^n}{2^{n\sigma/\kappa_{\max}}}\right)}{\left(\frac{|\lambda_{\max}^{(i)}|^n}{|\lambda_{\max}^{(i)}|-1} + \frac{|\lambda_\beta^{(i)}|^n}{|\lambda_\beta^{(i)}|-1}\right)\bar{V}^{(i)} + 2^{-n\sigma/\kappa_{\max}}|\lambda_{\max}^{(i)}|^n \cdot 2\bar{V}^{(i)} + o\left(\frac{|\lambda_{\max}^{(i)}|^n}{2^{n\sigma/\kappa_{\max}}}\right)}\right). \tag{6.5.41}$$

We have now to distinguish between two cases, namely when **(i)** $|\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|$ and **(ii)** $|\lambda_\beta^{(i)}| < |\lambda_{\max}^{(i)}|$:

(i) The rate allocated to estimate states with $|\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|$ is computed using the second-order approximation of $U^{(i)}$, and corresponds to

$$r_\beta^{(i)} = \log_2|\lambda_{\max}^{(i)}| - \frac{1}{n}\log_2\left(\frac{2^{-n\sigma/\kappa_{\max}}|\lambda_{\max}^{(i)}|^n \cdot 2\bar{V}^{(i)} + o\left(\frac{|\lambda_{\max}^{(i)}|^n}{2^{n\sigma/\kappa_{\max}}}\right)}{2\bar{V}^{(i)}\frac{|\lambda_{\max}^{(i)}|^n}{|\lambda_{\max}^{(i)}|-1} + 2^{-n\sigma/\kappa_{\max}}|\lambda_{\max}^{(i)}|^n \cdot 2\bar{V}^{(i)} + o\left(\frac{|\lambda_{\max}^{(i)}|^n}{2^{n\sigma/\kappa_{\max}}}\right)}\right)$$

$$\approx \log_2|\lambda_{\max}^{(i)}| - \frac{1}{n}\log_2\left(\frac{2^{-n\sigma/\kappa_{\max}}|\lambda_{\max}^{(i)}|^n}{\frac{|\lambda_{\max}^{(i)}|^n}{|\lambda_{\max}^{(i)}|-1} + 2^{-n\sigma/\kappa_{\max}}|\lambda_{\max}^{(i)}|^n}\right)$$

$$\approx \log_2|\lambda_{\max}^{(i)}| - \frac{1}{n}\log_2\left(2^{-n\sigma/\kappa_{\max}}\left(|\lambda_{\max}^{(i)}| - 1\right)\right)$$

$$= \log_2|\lambda_{\max}^{(i)}| + \frac{\sigma}{\kappa_{\max}} + o\left(\frac{1}{n}\right), \tag{6.5.42}$$

with $\beta \in [1 : B^{(i)}]$ s.t. $|\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|$.

(ii) For $|\lambda_\beta^{(i)}| < |\lambda_{\max}^{(i)}|$, the respective rates can be obtained using the first-order approximation of $U^{(i)}$ (6.5.40) as follows

$$
\begin{aligned}
r_\beta^{(i)} &\approx \log_2|\lambda_\beta^{(i)}| - \frac{1}{n}\log_2\left(\frac{1-\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n}}{1+\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n}}\right) \\
&= \log_2|\lambda_\beta^{(i)}| - \frac{1}{n\ln 2}\left(\ln\left(1-\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n}\right) - \ln\left(1+\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n}\right)\right) \\
&= \log_2|\lambda_\beta^{(i)}| - \frac{1}{n\ln 2}\left(-\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n} - \frac{1}{2}\left(\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n}\right)^2\right. \\
&\quad \left. -\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n} + \frac{1}{2}\left(\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n}\right)^2 + O\left(\left(\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n}\right)^3\right)\right) \\
&= \log_2|\lambda_\beta^{(i)}| - \frac{1}{n\ln 2}\left(-2\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n} + \underbrace{O\left(\left(\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n}\right)^3\right)}_{=o(1)}\right).
\end{aligned}
$$

Consequently, the rates $r_\beta^{(i)}$ are

$$
\begin{aligned}
r_\beta^{(i)} &= \log_2|\lambda_\beta^{(i)}| + \underbrace{\frac{2}{n\ln 2}\frac{|\lambda_{\max}^{(i)}|-1}{|\lambda_\beta^{(i)}|-1}\frac{|\lambda_\beta^{(i)}|^n}{|\lambda_{\max}^{(i)}|^n}}_{=o(\frac{1}{n})} + o\left(\frac{1}{n}\right) \\
&= \log_2|\lambda_\beta^{(i)}| + o\left(\frac{1}{n}\right),
\end{aligned} \tag{6.5.43}
$$

with $\beta \in [1 : B^{(i)}]$ s.t. $|\lambda_\beta^{(i)}| < |\lambda_{\max}^{(i)}|$.

This analysis shows that the states associated with eigenvalues $\lambda_\beta^{(i)}$, such that $|\lambda_\beta^{(i)}| = |\lambda_{\max}^{(i)}|$, should be estimated using equal rates (6.5.42) that are mainly the sum of the log-magnitude of the eigenvalue and the term $R^{(i)} - \delta^{(i)} - H^{(i)}$ equally divided among these states. On the other hand, the estimation of all other states can be performed using rates that are slightly larger than the log-magnitude of the associated eigenvalue, as it can be seen in (6.5.43).

6.6 Numerical Example

We now explore the following numerical example: Consider two unrelated plants $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}$ whose state matrices $A^{(1)}$ and $A^{(2)}$ have the following structure

$$A^{(i)} = \begin{pmatrix} \lambda_1^{(i)} & \\ & \lambda_2^{(i)} \end{pmatrix}, \text{ for } i \in \{1,2\}. \tag{6.6.1}$$

Note that, in this setting, we have for each dynamical system two Jordan blocks, each of dimension one (scalar), i.e, for $i \in \{1,2\}$

$$B^{(i)} = 2, \tag{6.6.2}$$

$$b_\beta^{(i)} = 1, \ \forall \beta \in [1:2]. \tag{6.6.3}$$

Recall that the topological entropy for each of the plants is defined as follows

$$H^{(i)} = \sum_{\beta=1}^{2} \log_2 |\lambda_\beta^{(i)}|. \tag{6.6.4}$$

The communication is assumed to occur over a BAC such that $H^{(i)} < R^{(i)}$, for $i \in \{1,2\}$, and $(R^{(1)}, R^{(2)}) \in \mathscr{C}_0$. Furthermore, unless otherwise stated, we suppose that the noise ranges are

$$[\![V^{(i)}]\!] = [\![V]\!] = [-1,1] \times [-1,1], \text{ for } i \in \{1,2\}. \tag{6.6.5}$$

For a fixed code block-length n, we have

$$\begin{aligned} -\sigma = H^{(i)} - R^{(i)} + \delta^{(i)} &= \frac{1}{n} \sum_{\beta=1}^{2} \log_2 \left(\frac{U^{(i)} - \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}}{U^{(i)} + \frac{|\lambda_\beta^{(i)}|^n - 1}{|\lambda_\beta^{(i)}| - 1} \bar{V}^{(i)}} \right) \\ \Leftrightarrow \ 2^{-n\sigma} &= \frac{(U^{(i)})^2 - \left(\frac{|\lambda_1^{(i)}|^n - 1}{|\lambda_1^{(i)}| - 1} + \frac{|\lambda_2^{(i)}|^n - 1}{|\lambda_2^{(i)}| - 1} \right) \bar{V}^{(i)} U^{(i)} + \left(\frac{|\lambda_1^{(i)}|^n - 1}{|\lambda_1^{(i)}| - 1} \cdot \frac{|\lambda_2^{(i)}|^n - 1}{|\lambda_2^{(i)}| - 1} \right) (\bar{V}^{(i)})^2}{(U^{(i)})^2 + \left(\frac{|\lambda_1^{(i)}|^n - 1}{|\lambda_1^{(i)}| - 1} + \frac{|\lambda_2^{(i)}|^n - 1}{|\lambda_2^{(i)}| - 1} \right) \bar{V}^{(i)} U^{(i)} + \left(\frac{|\lambda_1^{(i)}|^n - 1}{|\lambda_1^{(i)}| - 1} \cdot \frac{|\lambda_2^{(i)}|^n - 1}{|\lambda_2^{(i)}| - 1} \right) (\bar{V}^{(i)})^2} \end{aligned} \tag{6.6.6}$$

Solving (6.6.6) with respect to $U^{(i)}$ and letting $\Lambda^{(i)}_{\beta,n} := \frac{|\lambda^{(i)}_\beta|^n - 1}{|\lambda^{(i)}_\beta| - 1}\bar{V}^{(i)}$ yields the following solutions

$$U^{(i)}_- = \frac{1}{-2(1-2^{-n\sigma})}\Big(-(\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})(1+2^{-n\sigma}) \\ -\sqrt{(\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})^2(1+2^{-n\sigma})^2 - 4(1-2^{-n\sigma})^2\Lambda^{(i)}_{1,n}\Lambda^{(i)}_{2,n}}\Big), \tag{6.6.7}$$

and

$$U^{(i)}_+ = \frac{1}{-2(1-2^{-n\sigma})}\Big(-(\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})(1+2^{-n\sigma}) \\ +\sqrt{(\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})^2(1+2^{-n\sigma})^2 - 4(1-2^{-n\sigma})^2\Lambda^{(i)}_{1,n}\Lambda^{(i)}_{2,n}}\Big), \tag{6.6.8}$$

We now prove that $U^{(i)}_+$ does not fall into the domain of $U^{(i)}$ (Table 6.2), and hence it is not the desired solution. Without loss of generality, suppose that $|\lambda^{(i)}_2| \leq |\lambda^{(i)}_1|$, and consider the following expression

$$\begin{aligned} &-2(1-2^{-n\sigma})(U^{(i)}_+ - \Lambda^{(i)}_{1,n}) \\ &= \underbrace{-(\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})(1+2^{-n\sigma})}_{:=A>0} + \\ &\underbrace{\sqrt{(\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})^2(1+2^{-n\sigma})^2 - 4(1-2^{-n\sigma})^2\Lambda^{(i)}_{1,n}\Lambda^{(i)}_{2,n}} + 2(1-2^{-n\sigma})\Lambda^{(i)}_{1,n}}_{:=B>0} \\ &= B - A. \end{aligned} \tag{6.6.9}$$

Our objective now is to examine the sign of $B-A$, i.e., determine $B \overset{?}{\gtrless} A$. Since $A, B > 0$, we study the sign of $B^2 - A^2$, where

$$\begin{aligned} A^2 &= (\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})^2(1+2^{-n\sigma})^2, \\ B^2 &= (\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})^2(1+2^{-n\sigma})^2 - 4(1-2^{-n\sigma})^2\Lambda^{(i)}_{1,n}\Lambda^{(i)}_{2,n} + 4(1-2^{-n\sigma})^2(\Lambda^{(i)}_{1,n})^2 + \\ &\quad 4(1-2^{-n\sigma})\Lambda^{(i)}_{1,n}\sqrt{(\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})^2(1+2^{-n\sigma})^2 - 4(1-2^{-n\sigma})^2\Lambda^{(i)}_{1,n}\Lambda^{(i)}_{2,n}}. \end{aligned}$$

Hence, the difference between B^2 and A^2 yields the following

$$B^2 - A^2 = -4(1-2^{-n\sigma})^2\Lambda^{(i)}_{1,n}\Lambda^{(i)}_{2,n} + 4(1-2^{-n\sigma})^2(\Lambda^{(i)}_{1,n})^2 \\ +4(1-2^{-n\sigma})\Lambda^{(i)}_{1,n}\sqrt{(\Lambda^{(i)}_{1,n}+\Lambda^{(i)}_{2,n})^2(1+2^{-n\sigma})^2 - 4(1-2^{-n\sigma})^2\Lambda^{(i)}_{1,n}\Lambda^{(i)}_{2,n}}. \tag{6.6.10}$$

Since $4(1-2^{-n\sigma})\Lambda_{1,n}^{(i)} > 0$, we have

$$\frac{B^2-A^2}{4(1-2^{-n\sigma})\Lambda_{1,n}^{(i)}} = (1-2^{-n\sigma})(\Lambda_{1,n}^{(i)}-\Lambda_{2,n}^{(i)}) + \sqrt{(\Lambda_{1,n}^{(i)}+\Lambda_{2,n}^{(i)})^2(1+2^{-n\sigma})^2 - 4(1-2^{-n\sigma})^2\Lambda_{1,n}^{(i)}\Lambda_{2,n}^{(i)}}. \tag{6.6.11}$$

It remains to show that $\Lambda_{1,n}^{(i)}-\Lambda_{2,n}^{(i)} \geq 0$. To do this, we study the behavior of the function $\phi(x)$ defined as

$$\phi(x) := \frac{x^n-1}{x-1}, \text{ with } x>1,\ n \in \mathbb{Z}_{\geq 1}. \tag{6.6.12}$$

Observe that

$$\Lambda_{\beta,n}^{(i)} = \phi\left(|\lambda_\beta^{(i)}|\right)\bar{V}^{(i)}. \tag{6.6.13}$$

The first derivative $\phi^{'}(x)$ is then

$$\begin{aligned}\phi^{'}(x) &= \frac{(n-1)x^n - nx^{n-1}+1}{(x-1)^2} \\ &= \frac{x^n}{(x-1)^2}\left((n-1)-nx^{-1}+x^{-n}\right) \\ &= \frac{x^n}{(x-1)^2}\left((n+x^{-n})-(1+x^{-1})\right).\end{aligned} \tag{6.6.14}$$

Since $\frac{x^n}{(x-1)^2} > 0$, then $\text{sign}(\phi^{'}(x)) = \text{sign}((n+x^{-n})-(1+x^{-1}))$. Now, we distinguish between two cases as follows

- $n=1$: It is clear in this case that $\phi(x)=1$ for any $x>1$, and hence $\phi(|\lambda_1^{(i)}|)-\phi(|\lambda_2^{(i)}|)=0$, for any $|\lambda_1^{(i)}|, |\lambda_2^{(i)}| > 1$.

- $n \geq 2$: As $x>1$, then it holds that

$$1+\frac{1}{x} < 2 \leq n < n+\frac{1}{x^n}. \tag{6.6.15}$$

  Hence, $\phi^{'}(x) > 0$.

From the analysis above, we conclude that $\phi(x)$ is non-decreasing in x, for $x>1$ and $n \in \mathbb{Z}_{\geq 1}$. Hence, since by assumption $|\lambda_2^{(i)}| \leq |\lambda_1^{(i)}|$, then

$$\Lambda_{1,n}^{(i)} - \Lambda_{2,n}^{(i)} \geq 0. \tag{6.6.16}$$

Therefore, by using this result in (6.6.11), we conclude that

$$B^2 - A^2 \geq 0. \tag{6.6.17}$$

And as $A, B > 0$, this means that $B \geq A$. Subsequently, from (6.6.9) we obtain

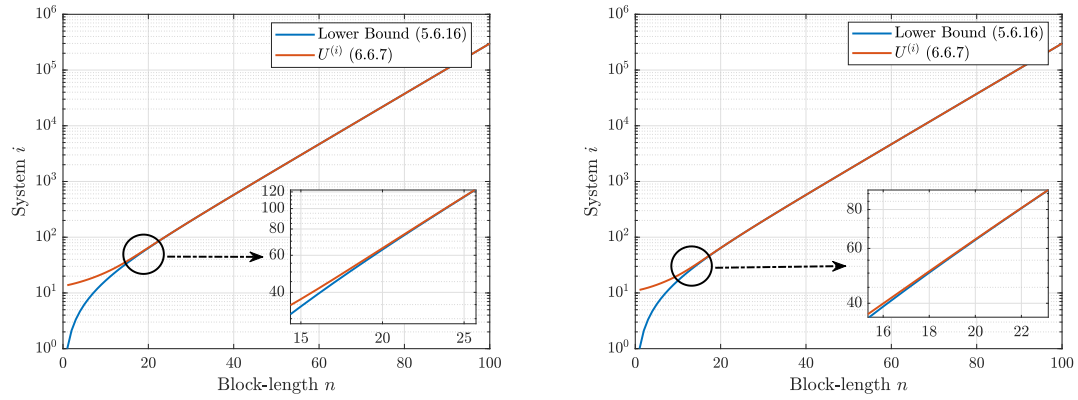
$$\begin{aligned} -2(1-2^{-n\sigma})(U_+^{(i)} - \Lambda_{1,n}^{(i)}) &\geq 0 \\ \iff U_+^{(i)} &\leq \Lambda_{1,n}^{(i)}. \end{aligned} \tag{6.6.18}$$

Since the domain of $U^{(i)}$ is $\left[\frac{|\lambda_{\max}^{(i)}|^n - 1}{|\lambda_{\max}^{(i)}| - 1}\bar{V}^{(i)}, \infty\right)$, we conclude from this analysis that $U_+^{(i)}$ is not the desired solution (and $U_-^{(i)}$ is consequently the unique solution).

To compare between the lower and upper bounds on the max-norm of the time-asymptotic estimation error, we plot the results obtained in (5.6.16) and (6.6.7) in Fig. 6.1 for different combinations of $(h_1^{(i)}, h_2^{(i)})^T$ and $\boldsymbol{\delta}^{(i)}$. In fact, each row in Fig. 6.1 shows the behaviour of the lower bound and $U^{(i)}$ using the same vector of topological entropies but different $\boldsymbol{\delta}^{(i)}$. Clearly, for all of the six scenarios, $U^{(i)}$ (6.6.7) and the lower bound (5.6.16) converge to the same limit, as it was discussed in Proposition 1. A further interesting observation here is that the increase of the pull-back factor $\boldsymbol{\delta}^{(i)}$ (from 0.01 to 0.1) causes the convergence of $U^{(i)}$ and the lower bound to occur at a larger block-length n (see plots **(a)** vs **(b)**, **(c)** vs **(d)** and **(e)** vs **(f)**). Additionally, as the sum of topological entropies, i.e., $H^{(i)} = \sum_{j\in\{1,2\}} h_j^{(i)}$, grows, it takes also larger block-lengths n for $U^{(i)}$ and the lower bound to converge. For instance, take the example of $\boldsymbol{\delta}^{(i)} = 0.1$ and $H^{(i)} = 0.2$ (Fig. 6.1(a)), then the convergence happens at a block-length n of ~ 18, whereas for $H^{(i)} = 0.5$ (Fig. 6.1(c)), n needs to be approximately 45 for $U^{(i)}$ and the lower bound to approximately concur.

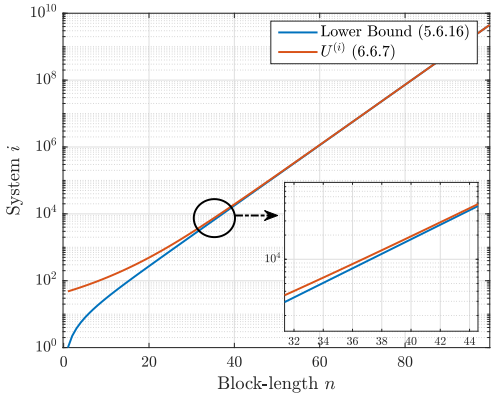
6.7 Summary

In this chapter, we started by deriving an upper bound on the time-asymptotic max-norm of the estimation error when the measurements are uniformly quantized. In the context of this analysis, we studied three cases, namely one-dimensional Jordan blocks (Theorem 10), non-scalar Jordan blocks with real eigenvalues (Theorem 11), and non-scalar Jordan blocks with complex eigenvalues (Theorem 12). Afterwards, we focused mainly on scalar Jordan blocks, and formulated the upper bound as a convex optimization problem in Section 6.5. This led us to derive the optimal rate allocation strategy that minimizes the previously obtained upper bound (Theorem 13), i.e., in other words, that results in the tightest upper bound. Subsequently, by conducting an analysis for the large block-length regime in Subsection 6.5.3, we showed that when noise has a hypercuboidal structure, then the upper and lower bounds converge to the same expression in Proposition 1. This result was finally illustrated using a numerical example in Section 6.6.

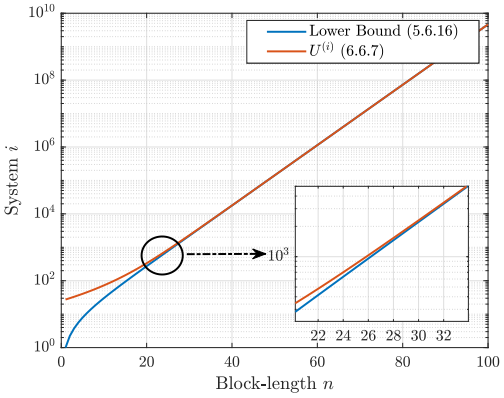


(a) $(h_1^{(i)}, h_2^{(i)})^T = (0.15, 0.05)$ and $\delta^{(i)} = 0.1$.

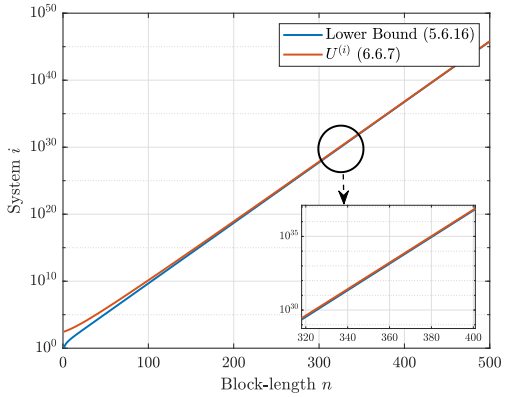
(b) $(h_1^{(i)}, h_2^{(i)})^T = (0.15, 0.05)$ and $\delta^{(i)} = 0.01$.



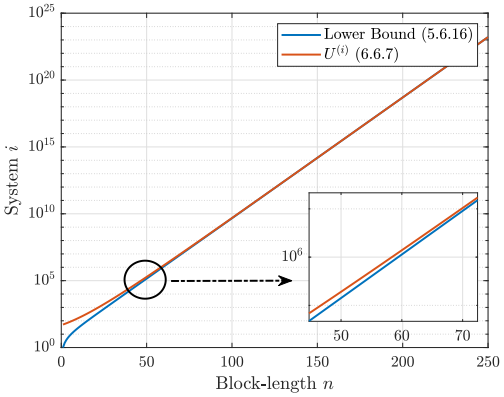
(c) $(h_1^{(i)}, h_2^{(i)})^T = (0.3, 0.2)$ and $\delta^{(i)} = 0.1$.



(d) $(h_1^{(i)}, h_2^{(i)})^T = (0.3, 0.2)$ and $\delta^{(i)} = 0.01$.



(e) $(h_1^{(i)}, h_2^{(i)})^T = (0.3, 0.3)$ and $\delta^{(i)} = 0.1$.



(f) $(h_1^{(i)}, h_2^{(i)})^T = (0.3, 0.3)$ and $\delta^{(i)} = 0.01$.

Figure 6.1: Behaviour of $U^{(i)}$ (6.6.7) and the lower bound (5.6.16) on the state estimation error of the i-th system for different configurations of the topological entropy vector $(h_1^{(i)}, h_2^{(i)})^T$ and pull-back factor $\delta^{(i)}$. The rate $R^{(i)}$ is set to 0.7203 bits/use. Note that the results on the LHS are plotted for $\delta^{(i)} = 0.1$, whereas the ones on the RHS are obtained for $\delta^{(i)} = 0.01$.

"I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me."

— Sir Isaac Newton (1643–1727)

7 Conclusion

In this chapter of the dissertation, we conclude with a summary of the contributions and the insights developed throughout the thesis, followed by a discussion of possible future research directions and remaining open problems of interest.

7.1 Thesis Summary

The underlying motivation behind this work was to investigate the problem of state estimation in distributed systems. More specifically, the states of an unstable system are observed by means of different sensors whose measurements are simultaneously communicated over a MAC. The objective at the receiver side is to reliably estimate the system states, i.e., to maintain bounded estimation error. To this end, we started by investigating the zero-error capacity region $\mathscr{C}_0$ of MACs in Chapter 3. Firstly, we rigorously studied zero-error communication over two-user MAC with common message. An initial result in this chapter was proving the convexity of $\mathscr{C}_0$ (Theorem 1), followed by providing the required conditions that guarantee the non-emptiness of the interior of $\mathscr{C}_0$ (Lemma 1). Afterwards, we derived a characterization of $\mathscr{C}_0$ of the two-user MAC model using the notions of nonstochastic information as well as nonstochastic conditional information (Theorem 2). Subsequently, we extended this result to both the $\boldsymbol{M}$-user MAC with one common message (Theorem 3) as well as the $\boldsymbol{M}$-user MAC with both a common message among all users and pairwise shared messages (Theorem 4) .

Next, the state estimation problem over two-user MAC with a common message was addressed in Chapter 4. The setup studied in this chapter consists of three mutually unrelated LTI systems whose outputs are detected by two separate sensors. The output signals of two of these dynamical systems

are privately measured, whereas the remaining plant's output is available to both sensors. We derived both necessary and sufficient conditions under uniformly bounded error criterion (Theorem 5). This result allowed us to connect the intrinsic properties of the dynamical systems, namely their topological entropies, to the zero-error capacity region of the communication channel. In particular, it turns out that if the vector of topological entropies $\boldsymbol{h}$ lies within the interior region of $\mathscr{C}_0$, it is possible to construct an encoder-estimator tuple yielding uniformly bounded estimation errors. Additionally, we have proven that if there exists an encoder-estimator tuple achieving the desired criterion, then $\boldsymbol{h} \in \mathscr{C}_0$. Subsequently, we showed that this setup is indeed general in the case of noiseless LTI systems, and provided a numerical example to illustrate the result using the BAC model.

In Chapter 5, we focused on studying the system performance by deriving universal lower bounds on the estimation errors. We used two approaches, namely centralized and decentralized, to obtain a set of lower bounds that characterized the fundamental trade-off between the communication data rate, code block-length, system dynamics and the state estimation performance (Theorem 7 and Theorem 9). Furthermore, we proved the tightness of the decentralized lower bounds in the case of one-dimensional systems. Finally, the centralized bounds are generalized to accommodate the case of a system with an arbitrary number of decoupled states in Lemma 4, and the behaviour of the obtained lower bound was studied when the code block-length is asymptotically large (Lemma 5). In such regime, we showed that the system performance is mainly driven by the dominant mode and the process noise affecting the respective state.

Finally, in Chapter 6, we turned our attention to deriving upper bounds on the max norm of the estimation error, as this would give us a guarantee on the system performance. The same setup was considered in this chapter along with the assumption that the plants' measurements are uniformly quantized, and the states are estimated at a share of the total rate at which the channel operates. By focusing on the case where the state matrix consists of one-dimensional Jordan blocks, we formulated the upper bound on the time asymptotic max norm of the estimation error as a convex optimization problem, whose solution corresponds to the optimal rate allocation strategy. This result allowed us to gain better insights on how to efficiently allocate the available data resources while maintaining the estimation error bounded. For instance, for asymptotically large block-lengths the rates allocated to estimate the states with dominant eigenvalues must be strictly larger than the sum of the eigenvalue's log-magnitude and the back-off term equally distributed among these states, whilst the rest of the states can be reliably estimated using rates strictly larger then the log-magnitude of the corresponding eigenvalues. Additionally, we devised a process to derive the upper bound, and a closed-form expression when the code block-length $n \to \infty$. In this regime, it turns out that, for hypercuboidal process noise, both upper and lower bounds converge to the same limit.

7.2 Future Research

The obtained results in this work provide preliminary steps towards understanding information flow in complex networks that will allow us to address worst-case distributed estimation and control problems. Some of the interesting research questions that can be can be investigated in the future are discussed below.

- So far, we have only considered a communication system operating over a MAC without feedback, i.e., the current channel output does not influence the future output. Therefore, a natural extension of this scenario is to find a characterization of the zero-error feedback capacity region $\mathscr{C}_{0,f}$. In a paper published in 1985 [46], Dueck determined $\mathscr{C}_{0,f}$ in the classical information-theoretic framework for a special class of MACs with feedback, namely where each input is a function of the output and the other input. It is, hence, interesting to address such model using the tools offered by nonstochastic information theory and use this as a starting point to study the zero-error capacity region of more general MACs with feedback.

- Once the characterization of zero-error feedback capacity $\mathscr{C}_{0,f}$ of MAC with feedback is obtained, we will be in a position to study stabilizability of LTI systems across a shared MAC. In fact, it was shown in the literature for point-to-point channels that the necessary and sufficient condition to achieve uniformly bounded stability, is that $\mathscr{C}_{0,f}$ being strictly larger than the plant's topological entropy (see [2] for treating DMCs and [47] studies point-to-point channels with memory). It is therefore interesting to see whether a similar condition arises when control occurs over MACs.

- In our study, we have mainly focused on different classes of MACs. A further interesting direction is to investigate other channel models in multi-user communication such as *broadcast channels* (BC), *interference channels* (IC), and *relay channels* in the nonstochastic framework. It is, however, worthwhile noting, that the problem of characterizing the classical capacity region for these classes of multi-user channels is quite difficult. The capacity region of the IC, for instance, is still unknown in general [36]. It is possible, nonetheless, that in the context of the nonstochastic information theory, some inner and outer bounds on the zero-error capacity region could be obtained. Some promising preliminary results in this direction were obtained in [48]. Using graph theory, the author derived necessary and sufficient conditions for the zero-error capacity of some multi-user channels to be strictly positive.

  - An example of broadcast channels (BCs) is depicted in Fig. 7.1. There are three sources, one encoder, and two decoders. The broadcast channel is characterized therefore by coordination at transmitter, whilst the decoders act independently from each other.

  - In Fig. 7.2, an example of interference channels (ICs) is illustrated. It consists of two transmitters and two receivers. Each of the unrelated messages $W^{(1)}$ and $W^{(2)}$ is destined

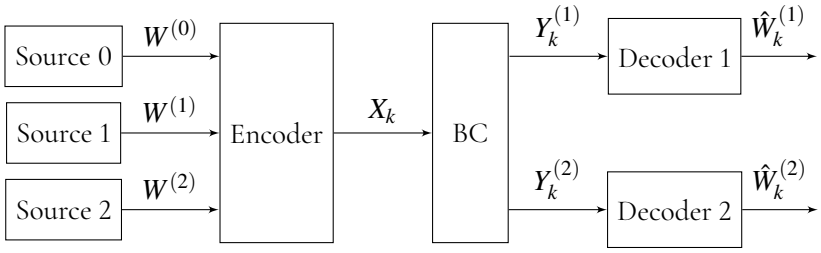


Figure 7.1: A high-level model of the two-receiver BC operating at time instant k.

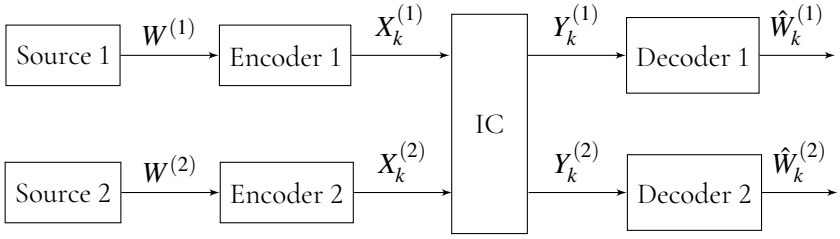


Figure 7.2: A high-level model of the two-receiver IC operating at time instant k.

for the corresponding receiver 1 and 2. In this setup, there is coordination between the transmitters as well as the receivers.

- As outlined in the course of this dissertation, an exact $\mathscr{C}_0$ characterization for many MACs, such as the BAC, remains unknown. Hence, it is also of interest to develop efficient algorithms to compute nonstochastic information and to ultimately characterize the zero-error capacity region of any MAC, or at least to reduce the gap between the upper and lower bounds characterizing the region.

- In our analysis of the state estimation problem, we considered only LTI systems. It is, nonetheless, also interesting to extend this approach to time-varying and nonlinear systems.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 7 1948.

[2] G. N. Nair, "Nonstochastic information concepts for estimation and control," in *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 45–56, Dec 2015.

[3] G. N. Nair, "A nonstochastic information theory for communication and state estimation," *IEEE Transactions on Automatic Control*, vol. 58, pp. 1497–1510, June 2013.

[4] A. Rangi and M. Franceschetti, "Channel coding theorems in non-stochastic information theory," in *2021 IEEE International Symposium on Information Theory (ISIT)*, pp. 1790–1795, 2021.

[5] M. Gagrani, Y. Ouyang, M. Rasouli, and A. Nayyar, "Worst-case guarantees for remote estimation of an uncertain source," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1794–1801, 2021.

[6] M. Tabbara, D. Nesic, and A. R. Teel, "Stability of wireless and wireline networked control systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1615–1630, 2007.

[7] W. P. M. H. Heemels and N. van de Wouw, *Stability and Stabilization of Networked Control Systems*, pp. 203–253. London: Springer London, 2010.

[8] C. E. Shannon, "Two-way communication channels," 1961.

[9] P. Sethi and S. R. Sarangi, "Internet of things: Architectures, protocols, and applications," 01 2017.

[10] F. F. C. Rego, A. M. Pascoal, A. P. Aguiar, and C. N. Jones, "Distributed state estimation for discrete-time linear time invariant systems: A survey," *Annual Reviews In Control*, vol. 48, pp. 36–56, 2019.

[11] M. Franceschetti and P. Minero, "Elements of information theory for networked control systems," *Lecture Notes in Control and Information Sciences*, vol. 450, pp. 3–37, 01 2014.

[12] A. S. Matveev and A. V. Savkin, "Shannon zero error capacity in the problems of state estimation and stabilization via noisy communication channels," *International Journal of Control*, vol. 80, no. 2, pp. 241–255, 2007.

[13] M. Wiese, T. J. Oechtering, K. H. Johansson, P. Papadimitratos, H. Sandberg, and M. Skoglund, "Secure estimation and zero-error secrecy capacity," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1047–1062, 2018.

[14] A. Saberi, F. Farokhi, and G. N. Nair, "Bounded state estimation over finite-state channels: Relating topological entropy and zero-error capacity," 2021.

[15] M. Rich and N. Elia, "Optimal mean-square performance for MIMO networked systems," in *2015 American Control Conference (ACC)*, pp. 6040–6045, IEEE, 2015.

[16] E. Garone, B. Sinopoli, A. Goldsmith, and A. Casavola, "LQG control for MIMO systems over multiple erasure channels with perfect acknowledgment," *IEEE Transactions on Automatic Control*, vol. 57, no. 2, pp. 450–456, 2011.

[17] A. A. Zaidi, T. J. Oechtering, and M. Skoglund, "Sufficient conditions for closed-loop control over multiple-access and broadcast channels," in *49th IEEE Conference on Decision and Control (CDC)*, pp. 4771–4776, IEEE, 2010.

[18] J. Liu and V. Gupta, "Stabilizability conditions for linear time invariant systems across a Gaussian MAC channel," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2310–2323, 2018.

[19] R. Ahlswede, "Multi-way communication channels," *Second International Symposium on Information Theory: Tsahkadsor, Armenia, USSR, Sept. 2 - 8, 1971*, vol. 2, pp. 23–51, Sep.

[20] H.-J. Liao, *A coding theorem for multiple access communications*. Ph.D. Dissertation, U. Hawaii, 1972.

[21] P. Austrin, P. Kaski, M. Koivisto, and J. Nederlof, "Sharper upper bounds for unbalanced uniquely decodable code pairs," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1368–1373, 2017.

[22] O. Ordentlich and O. Shayevitz, "A VC-dimension-based outer bound on the zero-error capacity of the binary adder channel," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 2366–2370, June 2015.

[23] M. Mattas and P. Ostergard, "A new bound for the zero-error capacity region of the two-user binary adder channel," *IEEE transactions on information theory*, vol. 51, no. 9, pp. 3289–3291, 2005.

[24] D. Slepian and J. K. Wolf†, "A coding theorem for multiple access channels with correlated sources," *The Bell System Technical Journal*, vol. 52, pp. 1037–1076, Sep. 1973.

[25] T. S. Han, "The capacity region of general multiple-access channel with certain correlated sources," *Information and Control*, vol. 40, no. 1, pp. 37–60, 1979.

[26] D. Gündüz and O. Simeone, "On the capacity region of a multiple access channel with common messages," in *2010 IEEE International Symposium on Information Theory*, pp. 470–474, 2010.

[27] A. Rényi, *Foundations of Probability*. Holden-Day, 1970.

[28] C. Shannon, "The lattice theory of information," *Transactions of the IRE Professional Group on Information Theory*, vol. 1, pp. 105–107, Feb 1953.

[29] S. Wolf and J. Wultschleger, "Zero-error information and applications in cryptography," in *Information Theory Workshop*, pp. 1–6, IEEE, 2004.

[30] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, Jan 1973.

[31] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975.

[32] L. Yu and V. Y. F. Tan, "Wyner's common information under Rényi divergence measures," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3616–3632, 2018.

[33] L. Yu and V. Y. F. Tan, "On exact and ∞-Rényi common informations," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3366–3406, 2020.

[34] A. S. Avestimehr, S. N. Diggavi, and N. David, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 1872–1905, 2011.

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, July 2006.

[36] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. USA: Cambridge University Press, 2012.

[37] C. Shannon, "The zero-error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, pp. 8–19, Sep. 1956.

[38] L. Lovasz, "On the Shannon capacity of a graph," *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 1–7, 1979.

[39] P. J. Antsaklis and A. Michel, *Linear Systems*. McGraw-Hill Higher Education, 1997.

[40] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1056–1068, 2004.

[41] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[42] R. C. Dorf and R. H. Bishop, *Modern Control Systems*. USA: Prentice-Hall, Inc., 9th ed., 2000.

[43] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," *Proceedings of the IEEE*, vol. 95, pp. 108–137, Jan 2007.

[44] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Journal on Control and Optimization*, vol. 43, no. 2, pp. 413–436, 2004.

[45] S. Axler, *Measure, Integration & Real Analysis*. Springer Cham, 2020.

[46] G. Dueck, "The zero error feedback capacity region of a certain class of multiple-access channels," *Problems of Control and Information Theory*, vol. 14, no. 2, pp. 89–103, 1985.

[47] A. Saberi, F. Farokhi, and G. N. Nair, "Bounded estimation over finite-state channels: Relating topological entropy and zero-error capacity," *IEEE Transactions on Automatic Control*, pp. 1–16, 2021.

[48] N. Devroye, "When is the zero-error capacity positive in the relay, multiple-access, broadcast and interference channels?," *2016 54th Annual Allerton Conference on Communication, Control, and Computing*.