Networked control and estimation under restrictions on channel capacity

Quentin Voortman

Networked Control and Estimation Under Restrictions on Channel Capacity

Quentin Voortman

The research reported in this thesis is part of the research programme of the Dutch Institute of Systems and Control (DISC). The author has successfully completed the educational program of the Graduate School DISC.

The work described in this thesis was carried out in the UCoCoS project which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675080.

A catalogue record is available from the Eindhoven University of Technology Library.
ISBN: 978-90-386-5328-0

Reproduction: ProefschriftMaken.

Networked Control and Estimation Under Restrictions on Channel Capacity

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op vrijdag 3 september 2021 om 13.30 uur

door

Quentin Voortman

geboren te Etterbeek, België

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	Prof.dr. L.P.H. de Goey
1^{e} promotor:	Prof.dr. H. Nijmeijer
2^{e} promotor:	Prof.dr.ir. J.-P. Richard (Centrale Lille)
copromotoren:	Dr.ir. A. Y. Pogromsky
	Dr.ir. D. Efimov (Inria)
leden:	Prof.dr.ir. W.P.M.H. Heemels
	Prof.dr.ir. S. Yüksel (Queen's University)
	Dr.ir. S. Tarbouriech (LAAS-CNRS Toulouse)

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

CENTRALE LILLE

THÈSE

Présentée en vue
d'obtenir le grade de

DOCTEUR

En

Spécialité : Automatique, Génie Informatique, Traitement du Signal et des Images

Par

Quentin VOORTMAN

DOCTORAT DELIVRÉ CONJOINTEMENT PAR CENTRALE LILLE ET TECHNISCHE UNIVERSTEIT EINDHOVEN

Titre de la thèse :
Contrôle et estimation sous contraintes de capacité de communication

Soutenue le 3 septembre 2021 devant le jury d'examen :

Président	Philip DE GOEY	Professeur, TU Eindhoven
Directeur de thèse	Jean-Pierre RICHARD	Professeur, Centrale Lille
Directeur de thèse	Henk NIJMEIJER	Professeur, TU Eindhoven
Codirecteur de thèse	Denis EFIMOV	Directeur de Recherche, Inria
Codirecteur de thèse	Alexander POGROMSKY	Professeur, TU Eindhoven
Rapporteur	Serdar YÜKSEL	Professeur, Queen's University
Rapporteur	Sophie TARBOURIECH	Directrice de Recherche, LAAS-CNRS
Examinateur	Maurice HEEMELS	Professeur, TU Eindhoven

Thèse préparée dans le Laboratoire CRIStAL
Ecole Doctorale SPI 072

"The lessons of mathematics are simple ones and there are no numbers in them: that there is structure in the world; that we can hope to understand some of it and not just gape at what our senses present to us; that our intuition is stronger with a formal exoskeleton than without one. And that mathematical certainty is one thing, the softer convictions we find attached to us in everyday life another, and we should keep track of the difference if we can."

J. Ellenberg, [Ellenberg, 2014]

Summary

Networked Control and Estimation Under Restrictions on Channel Capacity

Wireless communication technologies are omnipresent in the modern world and hence, there are plenty of examples from the control engineering field involving dynamical systems interacting via communication technologies. The fact that most current-day wireless communication technologies rely on packets implies that they suffer from several limitations/drawbacks: limited packet size, limited packet sending rate, and packet losses.

The typical structure of most problems pertaining to dynamical systems and communication technology interactions is as follows: one or several dynamical systems, or the components thereof are connected via communication channels. The systems are subject to a source of uncertainty (a source of uncertainty can be noise, parametric uncertainty, perturbations, sensitivity to initial conditions). This source of uncertainty generates information that needs to be transferred via the communication channel. In order to solve the underlying control/observation tasks, it is necessary to design specific communication strategies that deal with the limitations and drawbacks of communication technologies. This thesis provides several such communication strategies, each for a different combination of drawbacks and sources of uncertainty.

The first result is for the remote observation of a nonlinear dynamical system over a communication channel which is subject to losses. The goal is to reconstruct online estimates of the state of a system at a remote location whilst using as few bits per unit of time as possible. A solution, in the form of a communication protocol, consisting of several interacting devices, is designed. The communication protocol is designed such that it functions even in the presence of losses in the communication channel.

Consensus for a network of agents, which communicate over data-rate constrained communication channels is the second result. Each agent consists of a nonlinear discrete-time dynamical system, which determines its dynamics, and is equipped with a smart sensor (a device capable of measuring the state and performing some computations) and a controller. All agents are interconnected through data-rate constrained channels. The smart sensor and controller of each agent are placed at locations remote from one another such that the smart sensor and controller need to use the communication channels as well. By exchanging messages, the sensors and controllers should steer the agents so that they achieve a particular type of consensus. Three different designs of smart sensors, controllers, and communication protocols that achieve this feature are presented, each with an increasing degree of interaction between the agents. For each protocol, a theorem providing conditions on the sufficient minimal data rates to implement them is presented. The protocols are tested on various example networks of dynamical systems, for which the theoretical bounds on the rate are compared to the rates observed in simulations.

The third result is an event-triggered data-rate constrained observation scheme for a perturbed continuous-time Lipschitz-nonlinear dynamical system. The system is connected to a remote location by means of a communication channel that can only send a limited number of bits per unit of time. The goal is to provide estimates of the state of the system at the remote location whilst respecting the communication channel capacity constraint. The developed solution uses an event-triggered mechanism to reduce the average data rate. For this solution, abound on the minimum channel capacity is provided. This actual rate resulting from the implementation is tested through simulations.

The fourth result consists of a remote data-rate constrained observer for a single discrete-time Lipschitz-nonlinear dynamical system which is governed by an external signal and subjected to bounded state perturbations and measurement error. The objective is to provide estimates of the state at the remote location by sending messages via a data-rate communication channel and to limit the bandwidth usage of the communication. A solution in the form of several interacting agents is proposed. This solution makes use of an event-triggering mechanism to reduce bandwidth usage. A theoretical maximum communication rate is computed. This theoretical rate is then compared to the actual communication rate by means of simulations on several dynamical systems.

The final result of the thesis is an observer that is experimentally validated

on unicycle-type TurtleBot mobile robots. A specific event-triggered data-rate constrained observer is designed. The theoretical maximum rate is computed and then compared to the actual rate in experiments, which confirms the effectiveness of the designed communication protocol.

Résumé

Contrôle et estimation sous contraintes de capacité de communication

Les technologies de communication sans fil sont omniprésentes dans le monde moderne et, de fait, il existe pléthore d'exemples du domaine de l'automatique impliquant des systèmes dynamiques qui interagissent via des technologies de communication. Le fait que la plupart des technologies de communication actuelles reposent sur des envois d'informations regroupées en paquets implique qu'elles souffrent de plusieurs limitations et revers : taille de paquets limitée, vitesse d'envoi de paquets limitée et pertes de paquets.

La structure typique de la plupart des problèmes ayant attrait aux interactions entre les systèmes dynamiques et les technologies de communication est la suivante : un ou plusieurs systèmes dynamiques, ou leurs composantes, sont connectés par le biais de canaux de communication. Les systèmes sont affectés par une source d'incertitude (une source d'incertitude peut être du bruit, de l'incertitude paramétrique, des perturbations ou bien de la sensibilité aux conditions initiales). Cette source d'incertitude génère de l'information qui doit être transmise via le canal de communication. Afin de résoudre les tâches de contrôle/observation sous-jacentes, il est nécessaire de développer des stratégies de communication spécifiques qui gèrent les limitations et revers des technologies de communication. Cette thèse propose plusieurs de ces stratégies, chacune pour une combinaison différente de limitations et sources d'incertitude.

Le premier résultat concerne l'observation à distance d'un système dynamique non linéaire via un canal de communication qui est sujet à des pertes. Le but est de produire, en temps réel, des estimations de l'état d'un système distant, en envoyant aussi peu de bits par unité de temps que possible. Une solution est développée sous la forme d'un protocole de communication constitué de plusieurs appareils interagissants. Ce protocole de communication est ima-

giné de façon qu'il fonctionne même en cas de pertes de paquets dans le canal de communication.

Le deuxième résultat est le consensus d'un réseau d'agents qui communiquent via des canaux de communication avec contraintes de données. Chaque agent est constitué d'un système dynamique, qui détermine son comportement, et est équipé d'un capteur intelligent (un appareil capable de mesurer l'état du système ainsi que d'effectuer des calculs simples), ainsi que d'un contrôleur. Les agents sont interconnectés via des canaux de communication limités en termes de données transmissibles. Les capteurs intelligents et contrôleurs de chaque agent sont placés à des endroits éloignés l'un de l'autre et ils doivent donc également utiliser les canaux de communication pour communiquer entre eux. En échangeant des messages, les capteurs et contrôleurs doivent mener les systèmes à une forme de consensus particulière. Trois conceptions différentes de capteurs, contrôleurs et protocoles de communication sont développées, chacune avec un niveau croissant d'interaction entre les agents. Pour chaque protocole, le taux de transmission minimal suffisant est calculé. Les protocoles sont testés sur divers réseaux de systèmes dynamiques, pour lesquels le taux de transmission minimal est comparé au taux de communication observé pendant des simulations.

Le troisième résultat est un observateur à évènements discrets avec contraintes de données pour un système en temps continu non-linéarités lipschitziennes. Le système est relié à un lieu distant par le biais d'un canal de communication qui ne peut qu'envoyer des quantités limitées de bits par unités de temps. Le but est de produire des estimations de l'état du système à distance tout en respectant la contrainte de transmission du canal de communication. La solution développée utilise un mécanisme d'évènements discrets afin de réduire le nombre moyen de communications. Pour cette solution, une borne sur la capacité minimale de transmission nécessaire est calculée. Le réel taux de transmission résultant de l'implémentation de ce protocole de communication est testé via des simulations.

Le quatrième résultat consiste en un observateur pour système en temps discret avec non-linéarités lipschitziennes, perturbations et signal de commande. L'objectif est de produire des estimations de l'état en envoyant des messages via un canal de communication tout en limitant l'utilisation de bande-passante. Une solution est proposée sous la forme de plusieurs appareils interagissants. Cette solution fait usage d'un mécanisme d'évènements discrets pour réduire l'usage de la bande passante. Le maximum de bande passante théorique nécessaire est cal-

culé. Ce maximum est alors comparé, via des simulations sur plusieurs systèmes dynamiques, au taux résultant de l'implémentation du protocole de communication.

Le dernier résultat est un observateur qui est validé expérimentalement sur des TurtleBots, qui sont des robots mobiles de type unicycle. Un observateur avec contraintes de données est développé spécifiquement pour ce type de robots. Le taux de communication maximum est calculé et ensuite comparé au taux réel via des expériences, qui confirment l'efficacité du protocole de communication développé.

Contents

Chapter 1

Introduction

1.1 General Introduction

1.1.1 Historical Context

Communication plays a central role in what makes us humans. Sending messages over long distances has always been challenging and throughout history, humanity has used many different ways to achieve this: fires, smoke signals, beacons, drums, mail, telegraph, or even pigeon posts. On June 21st, 1880, two scientists made an invention that would later revolutionize the modern world. Alexander Graham Bell and Charles Sumner Tainter transmitted a voice telephone message wirelessly over 213 meters by converting an audio signal into a light signal which was then received by another light to audio converter ([Hutt et al., 1993]). Thus was born the photophone. At that time, they could not possibly imagine the importance that wireless communications would have nowadays. Yet, 150 years later, it is virtually impossible to imagine what the world would look like without wireless communications. Wireless communications are everywhere and used every day. Whether it be through radio waves, Bluetooth, Wi-Fi, 4G, 5G, or some other technologies, most people use multiple devices that rely on wireless technologies several times per day.

Since the mathematical field of control is a field that finds applications in many different areas of the modern world, it was massively impacted by the technological revolution that wireless communication has brought. Twenty years ago, Richard Murray and his co-authors ([Murray et al., 2003]) already observed that:

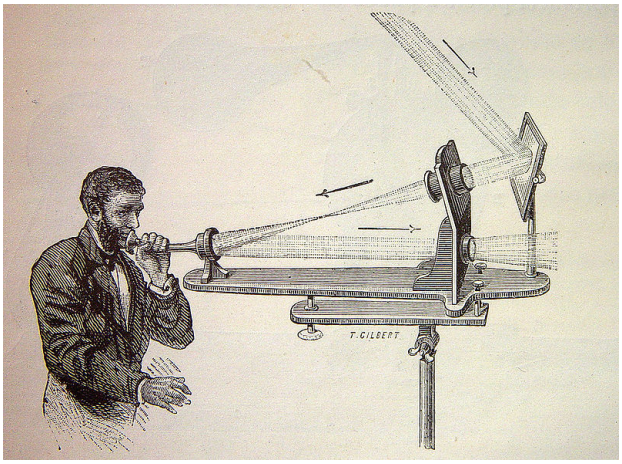

Figure 1.1: Illustration of the photophone's transmitter, as invented by Bell and Tainer in 1880. Source: [Guillemin, 1885]

> *"The advent of ubiquitous, distributed computation, communication, and sensing systems has begun to create an environment in which we have access to enormous amounts of data and the ability to process and communicate that data in ways that were unimagined 20 years ago."*

and nowadays, systems and control and communication technologies cannot be separated from one another, as was demonstrated in [Lamnabhi-Lagarrigue et al., 2017]:

> *"...Systems & Control is at the heart of the Information and Communication Technologies to most application domains"*

There are many examples of applications where dynamical systems and communication technologies are mixed, among them, we note: lunar base construction ([Brooks et al., 1990]), cooperation of mobile robots ([Eustace et al., 1993], [Arai and Ota, 1996]), payload transportation ([Johnson and Bay, 1995]), distributed sensor networks ([Sinopoli et al., 2003]), formation control ([Olfati-Saber and Murray, 2002, Eren et al., 2002, Vidal et al., 2003]), flocking ([Vicsek et al., 1995, Reynolds, 1987, Toner and Tu, 1998]), communication of underwater vehicles ([Awan et al., 2019, Hamilton et al., 2020]),...

When faced with a situation where a dynamical system is mixed with communication technology, one might think that both problems can be treated separately: leave the control to the control engineer and the communication to the telecommunication engineer, but as [Nair et al., 2007] emphasised:

> *"In engineering systems with large communication bandwidth, it makes sense to treat communication and control as independent functions, since the analysis and design of the overall system are simplified. However, recent emerging applications ... have begun to challenge the validity of this modular approach."*

To obtain the best results, it is thus necessary to consider situations where control problems are intertwined with communication problems and to tackle the particular challenges due to these interactions ([Richard and Divoux, 2007]).

1.1.2 Drawbacks of Communication Technologies

Since the sixties, most communication technologies rely on packets to send information, as opposed to relying on a continuous flow of bits ([Davies, 2001]). The reason for the usage of packets is that it allows several devices to use the same communication channel at the same time, which is a valuable property. There are however some inherent drawbacks due to package-based communication. These are ([Bemporad et al., 2010, Heemels et al., 2010]):

1. *Limited packet size:* all technologies rely on packet sizes to be limited, this implies that every message only contains a finite number of bits. The number of different messages that can be sent at any given time is thus finite;

2. *Limited transmission rate:* the number of packets that can be sent per unit of time is restricted, this can be due to the design of the communication technology, or the fact that several devices share the same communication medium;

3. *Packet losses:* some packets might be sent but not received on the other side of the communication channel, because of faulty communication devices, or interferences;

4. *Packet corruption:* due to technical problems, or interference, the content that is received by the receiving device might not have the same information content as the one that was sent. In the case of interference, if the intent is malicious, the issue is related to cyber-physical security issues;

5. *Latency:* there is a delay between the time when the packet is sent, and it is received. The latency may be due to scheduling protocols, which alternate between the different users of the communication medium, or simply due to physical reasons (e.g., when the communication agents are moving). This may even result in packets arriving in a different order than they were sent in.

The drawbacks can be split into two parts: limitations on rate/network congestion (points 1, 2, and 5), and disturbances (points 3 and 4). It has been observed that network congestion causes perturbations, in the sense that it increases the risk of packet losses, packet corruption, and transmission delays ([Granelli et al., 2007]). By reducing the network usage, one also reduces the chances of perturbations. Limiting the packet transmission rate is also important for all wireless, battery-powered objects (sensors, actuators) for which the energy consumption is an issue. For this reason, focusing on either limited packet sizes, or limited packet sending rates, or both should be the main focus when tackling the problems related to mixing dynamical systems and communication technologies.

1.1.3 Sources of Uncertainty

If there are drawbacks to using communication technologies, one might ask: "Why is it necessary to communicate?" and "What information needs to be transmitted?". The answer lies in the structure of the problems that involve the technologies. The first component of these problems is distance (as in distance that separates two locations) which cannot be covered other than by using communication technologies. The second component has to do with some missing information and pertains to the concept of a source of uncertainty.

Regarding the first component: in some applications, (mobile robots, aircraft, platooning/automated driving, underwater vehicles, drones, Mars rovers,...), due to the nature of the applications, a physical connection is not possible. If communication is necessary, then the only solution is to use wireless communication devices. The distance component is not sufficient to require communication, as there should also be a source of uncertainty.

A source of uncertainty is an abstract concept that is used to represent something that generates uncertainty. If there no uncertainty, meaning that everything about the connected agents is either known in advance, or can be predicted, there is no need to communicate, even if there is distance (uncertainty is information, see [Shannon, 1948], more on that later). The reason is that in the absence of uncertainty, one can entirely predict the behaviour of the connected agents by simply using the known information. In the case of dynamical systems, uncertainty manifests itself in several ways. Sources of uncertainty mainly come from the absence of an accurate, deterministic model. Mathematical constructions are always a simplification of reality and in control science, the model one has in mind is an interpretation, constructed from some (real or virtual) data for the purpose of solving some control problem. This means that, despite its necessity for the design step, every model contains uncertainties, in an irreducible way. For a given structure of a dynamical system, those uncertainties can be mainly regrouped in the following classes, each illustrated with

a basic example:

1. *Parametric uncertainty*: the dynamical system has some parameters which may vary over time and/or be unknown. Example ([Hill, 1886]):

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -f(t) & 0 \end{bmatrix} x(t),$$

   where $f(t)$ is a time-varying parameter. Depending on $f(t)$, the system may show periodic and bounded solutions, or it might be unstable ([Teschl, 2004]). If $f(t)$ is not known in advance, predicting the state is not possible, and hence, this forms a source of uncertainty.

2. *Perturbations/Noise*: the dynamical system's state/output is affected by noise or perturbations. Example ([Meier et al., 1967]):

$$\begin{aligned} x[k+1] &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x[k] + \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} u[k] + w[k], \\ y[k] &= x[k] + v[k], \end{aligned}$$

   where $w[k]$ and $v[k]$ are Gaussian white noises. Because the state is perturbed by a perturbation and the measurement is corrupted by a noise, it is not possible to reconstruct the state exactly from the measurements. Since there is no damping in the state-transition matrix (it has two eigenvalues equal to 1), the state perturbations can be accumulated over time, which forms a source of uncertainty.

3. *Sensitivity to initial conditions*: The system present a high sensitivity to initial conditions (e.g., the system has multiple equilibria or chaotic behavior), while its initial state is uncertain. Example: the Hénon map[1],([Henon, 1976]):

$$\begin{aligned} x_1[k+1] &= 1.4 + 0.3x_2[k] - x_1[k]^2 \\ x_2[k+1] &= x_1[k] \end{aligned}$$

   with $x[0] = x_0$ and $\|x_0\| \leq 1$. This dynamical system exhibits chaotic behaviour ([Strogatz, 1994]). A small difference in initial conditions can result in large differences in state-space trajectories, simply due to the chaotic nature of the system's dynamics. If the initial condition is unknown, it is not possible to predict the state-space trajectories of the system.

Sources of uncertainty all share the common feature that there is additional information necessary to be able to determine the state of the system in real-time.

[1] This example will be studied in Chapter 3, where the data rate constrained consensus of a network of Hénon maps is considered.

Sources of uncertainty are particularly relevant in the context of communication systems and that is why they have been immediately discussed in the 1948 paper: A Mathematical Theory of Communication by Claude Shannon ([Shannon, 1948]). In this paper, the author very appropriately depicts the fundamental problem of communication as "*reproducing at one point either exactly or approximately a message selected at another point*". He then observes that the meaning a message possesses very much depends on the physical or conceptual entity that the message is related to. The messages are always selected from a set of possible messages, each having a separate meaning, which can be inferred from the particular situation. The number of different possible messages is of course related to the complexity of the physical or conceptual entity that the messages are related to. If the number of messages is finite, then "*this number or any monotonic function of this number*" can be used as a measurement of the amount of information that is sent, every time one message is issued.

As was previously exposed, sources of uncertainty make the system's behaviour unpredictable in some situations, which means that more information is required to accurately determine its state. The variety of messages to accurately describe the state of the system is thus greater. Also, the "larger" the uncertainty, the "larger" the set of all possible messages to describe the system affected by the source of uncertainty. Sources of uncertainty are thus directly related to the amount of information that is sent in problems that mix both dynamical systems and communication technologies.

In conclusion, here is how one could describe the common features of all challenges arising from mixing dynamical systems and communication technologies:

> One or several devices, which are modelled as dynamical systems or components thereof, are placed at locations that are remote from one another. These locations are connected only by means of communication technologies. One or several sources of uncertainty generate information in the sense of [Shannon, 1948]. The task at hand is to find efficient communication strategies to deal with both the drawbacks of the underlying communication technologies and the sources of uncertainty.

1.1.4 Time Domains and Measurement Devices

Before we provide an overview of the existing results in the literature, it is necessary to briefly discuss the matter of the time domain over which the underlying dynamical system operates and how this is related to measurement devices.

Based on the equations that describe them, plenty of dynamical systems can be put into one of the following categories: continuous-time systems, discrete-

time systems, and hybrid systems (Hybrid systems are "*dynamical systems whose evolution depends on a coupling between variables that take values in a continuum and variables that take values in a finite or countable set*", [van der Schaft and Schumacher, 2000]). Generally speaking, what equation is used to describe a process depends on several factors including: the nature of process that the system describes, which part of the process the system should replicate, what the system's equation will be employed for (simulation, control, ...), and what classes of systems are widely used. Examples for each category of systems include:

- *(Continuous-time system):* The causal relation of input (electrical voltage) and the output (angular velocity) of a DC motor can be described by the following electro-mechanical system ([Zaccarian, 2005]):

$$\begin{aligned} L\dot{x}_1(t) &= -Rx_1(t) - k_e x_2 + u(t) \\ J\dot{x}_2(t) &= k_M x_1(t) - b_M x_2(t), \end{aligned}$$

  where x_1 is the current, x_2 the angular velocity of the rotor, u the input voltage, and L, R, k_e, J, k_M, and b_M are positive motor constants. This system operates over a continuous time domain.

- *(Discrete-time system):* A Verhulst process is a discrete-time system which describes the evolution of a population that has access to a limited resource ([Murray, 2002]):

$$x(k+1) = rx(k)\left(1 - \frac{x(k)}{K}\right),$$

  where $x(k)$ is the population at time k and r and K are positive constants. The evolution of a population is the change over a continuous time frame of a discrete variable (number of individuals). However, because of the lack of a system paradigm operating over such domains (discrete state, continuous time domain), the Verhulst process is modelled as a continuous variable operating over a discrete time domain (discrete-time system).

- *(Hybrid system):* A bouncing ball can be modelled as a hybrid system consisting of a differential equation which describes the free movement of the ball in between impacts and an impact rule which describes the impact of the ball ([van der Schaft and Schumacher, 2000]):

$$\begin{aligned} \ddot{x}(t) &= -1, \\ \dot{x}(\tau^+) &= -e\dot{x}(\tau^-), \end{aligned}$$

  where t is the time τ^- and τ^+ are the jump instants, $x(t)$ is the position of the bouncing ball and e is positive constant. This system operates both over a a continuous time domain and a discrete-time domain.

Note that some processes, such as the aforementioned bouncing ball, can be modelled in several ways (the bouncing ball system is sometimes also modelled as a discrete-time system, see [Tufillaro et al., 1992] for example).

Although it is possible to discretize continuous-time systems ([Lewis, 1992]) to transform them into discrete-time systems, it is often not advisable to do so as this can result in losses of precision (depending on the discretization method that is employed). Because the equations that describe discrete-time systems, continuous-time systems, and hybrid systems are very different, results that hold for one type of system generally need thorough reformulation to apply to the other type of systems and can sometimes require much more mathematical efforts to hold. In this thesis, results are developed for continuous-time systems and discrete-time systems.

As was mentioned earlier, modern communication technologies are intrinsically discrete-time processes: packets are sent with information at specific instants and in between packets nothing happens. When a continuous-time system or hybrid system is connected with a communication channel, there is thus automatically a problem of incompatibility of the time domain on which both entities operate (continuous time domain versus discrete communication instants). For discrete-time systems, another problem can occur: "what if the sampling times of the system are different from the communication instants?"

For all three types of systems, there is the need for a tool that operates the transition between different time domains. Such tools are often referred to as samplers. Samplers measure the full state of a system at specific discrete time instants. The simplest example of a sampler is an analog-to-digital converter ([Lathi and Ding, 2018]). Initially, samplers were designed such that the sampling occurs at equally spaced out instants of time but in more recent times, samplers have been based on events occurring, rather than based on time passing ([Tarbouriech et al., 2017]). More on this event-based approach in a subsequent section. The sampling tool can sometimes be a part of a device that measures the state of the system, which is called a sensor. In that case, the output of the system is possibly not the full state but rather some mapping of part of the states taken at discrete time instants. In classic (unconstrained) control theory, an observer is then used to reconstruct the state based on this output (on the condition that the system is observable). Another possibility with sensors is that they measure the full state corrupted by some measurement noise. In that case, a state estimator is generally used to obtain estimates of the uncorrupted state.

Problems involving dynamical systems and communication technologies sometimes assume full state measurement ([Liberzon, 2003a, Savkin, 2006]), sometimes output measurement ([Fradkov et al., 2006, Postoyan and Nesic,

2012]). Noise is sometimes also considered ([Matveev, 2008]). In the following section, many different results will be discussed. The terms remote observer and remote state estimator are sometimes used, even in the case of noiseless full state measurement ([Matveev and Pogromsky, 2016]). This terminological inaccuracy is due to the fact that even with noiseless full state measurement, it is not always possible to reconstruct estimates of the state at a remote location, to which the system is only connected via a communication channel. Regarding the problem of reconstructing remote estimates, it should be noted that in the literature, the terms observer and state estimator are sometimes used despite the absence in the developed solutions of what classically understood as an observer or a state estimator, simply because of the absence of better terminology.

1.1.5 Brief Overview of Solutions to Communication Limitations for Dynamical Systems

This section is dedicated to exploring the solutions that have been developed in the literature so far. As was previously mentioned, when handling problems related to combining dynamical systems and communication technologies, the main focus is on dealing with the limiting properties of the communication technology: the limit on the package transmission rate, the limited size of the packets, and latency. Although the latency problem is extremely relevant in current days (these lines are written only a few days after the landing of the Mars Rover: Perseverance), discussions on latency are outside of the scope of this thesis. The focus shall hence be put on the transmission rate and package size problems. From these two issues, it is possible to sort the works in this field into three categories: works that deal with the limit on the transmission rate, works that deal with the limit on the size of the packets, and works that tackle both problems at the same time.

1.1.5.1 Limit on the Transmission Rate

The fact that packets are used to communicate necessarily implies that in between two packets, no new data is available. The rate at which packets arrive thus determines how often new information is available about the dynamical system. For control engineers, this limitation gives rise to a phenomenon called sampling. Consider for example the problem depicted in Figure 1.2. One should design a controller for a plant where instead of having continuous-time measurements of the state $x(t)$, only measurements $x(t_k)$, sampled at instants t_k are available. Depending on the sampling rate (which, in the periodic case, is the frequency at which the sampling instants t_k occur), the design of a controller might simply be impossible: for example if the sampling frequency is too low, then the measured signal might not capture the dynamics of the system properly,

which makes it impossible to design a stabilizing controller.

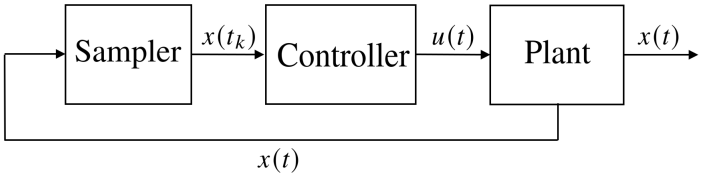

Figure 1.2: Example of a setup of a control problem involving sampling.

Such sampling problems are extremely common for digital control systems. One can find a wide array of results that were already obtained in the eighties in [Isermann, 1981]. A solution that naturally occurs when dealing with sampling problems is simply to impose that the sampling frequency is "high enough." In the case of communication technologies, however, this strategy is not always viable. What happens when the maximum sampling capacity of the system is reached and it is still not enough to achieve the control objective? What about network congestion? As was exposed earlier, high utilization of the communication technologies leads to higher network congestion, which in turn leads to perturbations. A solution to this problem appeared at the end of the nineties, in the form of event-triggered sampled-data control, which is also referred to as event-driven and event-based control.

The simplest description of event-driven/event-based sampled-data control is ([Åarzén, 1999]):

> *"In an event-based system it is the occurrence of an event rather than the passing of time, that decides when a sample should be taken."*

At its essence, event-based sampled-data control is a technique where processes that normally occur after a certain amount of time has passed, instead only occur when certain events happen. Various alternatives have been inspired by this idea of managing the sampling instant in an event-based, non-periodic way. In [Fiter et al., 2015], the authors list the following four types of "*dynamic sampling control*":

- event-triggering;
- self-triggering;
- periodic event-triggering;
- state-dependent sampling.

In all these cases, the continuous state system is subject to an event-based sampled-data control. For a general introduction to event-triggered control,

one can refer to [Heemels et al., 2012]. Early notable results in event-triggered control include [Åarzén, 1999], which developed an event-triggered PID controller, with the objective in mind to reduce the CPU utilization of the PID controller, by using an event-driven approach and [Åström and Bernhardsson, 1999] which compared time-driven sampling with event-driven sampling for various configurations involving linear systems (it was later proved in [Asadi Khashooei et al., 2018] that (periodic) event-triggered control can strictly outperform time-triggered control in this case). [Heemels et al., 1999] is also relevant for its development of a solution for master/slave synchronization of two motors by using time-driven and event-driven strategies. The common point of all of these papers is that they proved that event-driven control offered a promising alternative to time-driven control.

In [Yook et al., 2002], bandwidth is mentioned as the main motivation for the implementation of an event-triggered solution, which makes this paper particularly relevant in the context of this thesis. A framework for distributed control systems is developed where estimators are used at each node to estimate the values of the outputs and these outputs are then used in an event-triggered fashion: if the difference between the actual state and the estimated state exceeds a certain threshold value, a new estimate is broadcast over the network to all other nodes. In that work, local computation capacity is traded for bandwidth: by equipping each node with an estimator capable of performing computations, the overall required communication capacity is increased, but the benefit is that the network usage is reduced. Another paper, [Tabuada, 2007] develops event-triggered control and stabilization of nonlinear systems for which a continuous-time control law already exists (this is the principle of the emulation technique, see also [Omran et al., 2016]). This work is mentioned because the authors provide conditions such that so-called Zeno behaviour is avoided. Zeno behaviour ([Ames et al., 2005]), named after the Greek philosopher Zeno of Elea's Zeno's paradox, is a problem that occurs with event-driven control, where there exists an infinite number of triggering instants in a finite time interval. Regarding [Tabuada, 2007] and the conditions to avoid Zeno behaviour, in [Donkers and Heemels, 2012], the authors prove that the conditions dot not hold for arbitrarily small perturbations or in the case of output feedback control (see also [Borgers and Heemels, 2014] for more details). In [Fiter et al., 2012], the authors develop a state-dependent sampling control with an offline computation of regions of the state-space to trigger the control, which participates in alleviating the computational load of the online algorithm. We finally mention [Henningsson et al., 2008] because it presents an event-triggered controller for linear first-order stochastic systems, which is an example of a work that tackles noise, one of the three main sources of uncertainty for dynamical systems and [Fiter et al., 2015, Omran et al., 2014, 2016, Dolk et al., 2017] for the development of robustness conditions for perturbed systems (both linear and nonlinear cases are treated in

these works), again relating to one of the three sources of uncertainty.

For an overview of results on event-driven control, one can refer to [Garcia et al., 2010], [Miskowicz, 2016], and [Hetel et al., 2017]. We conclude with the following remark on the importance of event-driven control:

> Event-driven control has extensively been used in sampling problems to space out consecutive sampling instants, which contributes to alleviating the load on the underlying communication technologies. In a more general context (i.e. with or without feedback loops), event-driven communication forms an important part of the solutions to problems involving dynamical systems and communication technologies.

1.1.5.2 Limit on the Packet Size

All packet-based communication relies on sending packets, that is, messages consisting of a certain number of bits, in which data is encoded. A portion of these bits is typically reserved for protocol-specific information (information such as: who is sending the packet, who is it meant for, security confirmation bits,...) while the rest of the bits can freely be assigned. Let us thus consider the following example: a certain communication protocol allows for 4 physical bits to be sent at each communication. How many different messages can be sent based on these four bits? Logically, $2^4 = 16$, so 16 different messages. Note that in some sense, not sending a message constitutes a form of communication as well, and hence, one could consider that 17 different messages can be sent, by using and an extra "virtual" bit (which corresponds to no message being sent). For the remainder of this thesis however, we will not consider this possibility and hence n bits always imply that 2^n different messages can be encoded.

What if one has to use this communication protocol to transmit estimates of the state of the following dynamical system (logistic map[2], [May, 1976])

$$x[k+1] = \lambda x[k](1 - x[k]),$$

to a remote location, with the initial state known only to be in a certain set ($x[0] \in [0, 1]$), and $\lambda = 3.56995$? Is it possible to reconstruct an estimate $\hat{x}[k]$ of the state, with an arbitrarily small estimation error? A very basic answer would be that since it is possible to send 16 different messages and since the state $x[k]$ remains within the set $[0, 1]$ for all initial conditions [Strogatz, 1994], one might simply partition the set $[0, 1]$ into 16 equidistant intervals, and simply transmit the sequence of bits corresponding to the interval in which the state currently is

[2]This example will be studied in Chapter 3, where the data rate constrained consensus of a network of logistic maps is considered.

at each instant, to provide an estimate with precision 0.03125. This is however not a strategy that allows one to have arbitrarily small estimation errors since the estimation error cannot be reduced below 0.03125 unless one uses more bits to communicate or a different communication strategy. More on this problem, later in this thesis.

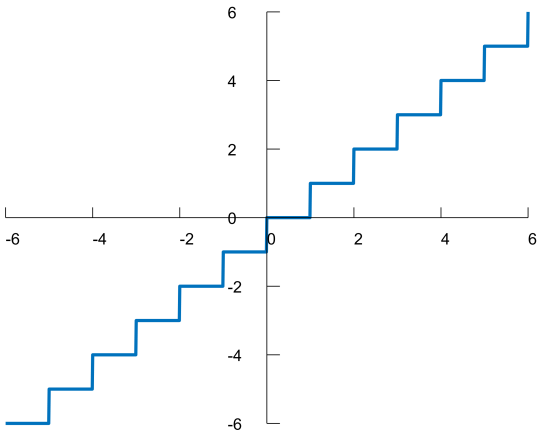

Figure 1.3: Example of a quantizer which maps real numbers to integers.

This example, and many other examples related to limits on the packet size, are better known in the control community as quantization problems. At its essence, quantization is the mathematical problem of mapping a variable that takes continuous values, to a discrete set. Figure 1.3 depicts an example of a simple quantizer, where numbers are mapped to the first preceding integer number (also known as the floor function).

Quantizers are devices or mappings constituted by a union of several quantization regions, which all correspond to a unique quantization value. The union of all quantization regions is called the quantization range. The quantizer maps all values laying inside a quantization region to the quantization value of that region. Depending on the chosen quantizer, the size of the quantization regions can be the same for all regions, or it can vary. Quantization policies can be static (memoryless) or dynamic (see definitions and references in [Ling and Lemmon, 2005]). The size of the regions determines the maximum quantization error when a point lies within the quantization range. If a point lies outside of the quantization range, it is generally mapped to the quantization value of the closest quantization region but the error is then typically very large. The quantization

error, and hence the precision of the quantizer is particularly important since it can have permanently destabilizing effects when it interacts with a control law that would have otherwise stabilized the system, as was shown in [Delchamps, 1990]. Other relevant works on the effects of quantization include [Montestruque and Antsaklis, 2007] which considers the effects of quantization in the case of linear systems with parametric uncertainty and [Nesic and Liberzon, 2009] where the authors use a Lyapunov function approach to tackle quantization problems for nonlinear systems.

The number of quantization regions is sometimes assumed to be infinite, in which case the quantization range is often infinite as well. Although this property is extremely valuable to avoid large errors when the quantized variable lies outside of the quantization range, for communication problems, such an assumption is not realistic. Indeed, to send a message which contains a quantized measurement, it is necessary to send a message which uniquely determines one of the quantization regions. To uniquely determine each quantization region, several bits should be sent, the number of which is proportional, up to a log term to the number of quantization regions. If the number of quantization regions is infinite, then the number of bits required to communicate is infinite as well, which is impossible. To cope with the fact that a finite number of quantization regions should be used, whilst maintaining a quantization region large enough that the quantized variable does not lie outside of it, quantizers with dynamic ranges are used. This solution was suggested in [Brockett and Liberzon, 2000], [Liberzon, 2003b], and [Liberzon and Nešić, 2007]. It consists of using a zoom-in/zoom-out procedure, where the size of the quantization regions depends on a parameter that is tuned online. The quantization range is thus adapted to always ensure that the quantized variable lies inside of the quantization range, whilst providing a low quantization error.

We conclude this section on quantization with the following remark:

> Quantization is an essential component of any problem involving communication technologies and dynamical systems. A good quantization algorithm guarantees a small quantization error whilst using as few quantization regions as possible, to reduce the load on the communication technology.

1.1.5.3 Combined Approach

In the two previous subsections, it was observed that limits on the transmission rate and the packet size correspond to sampling problems and quantization problems respectively. When both problems are tackled simultaneously, there

are two possible approaches. The first approach is to still consider both problems separately and to combine individual solutions to both problems into one solution. This is the approach followed, among others, by [Liberzon, 2003a, Li et al., 2012, Liu and Jiang, 2015, Li et al., 2016, Tallapragada and Cortes, 2016, Tanwani et al., 2016, Li et al., 2017, Liu and Jiang, 2019, Abdelrahim et al., 2019]. The common point of these works is that no characterisation of necessary/sufficient bit-rates is given. The second approach is to consider that the combination of limited transmission and limited packet size results in a new single problem: a data rate problem. This is the approach that will be followed in this thesis. There are two ways to deal with the data rate problem: as a capacity problem and as a minimal rate problem.

1. *Limited communication capacity:* The capacity of channel is an asymptotic quantity. Given that the channel has a maximum number of bits $b^+(\bar{s})$ that can be sent per time interval of length $\bar{s}$, the capacity is defined as (assuming that the limit exists)

$$c \coloneqq \lim_{\bar{s}\to\infty} \frac{b^+(\bar{s})}{\bar{s}}.$$

   The goal is, given a certain capacity, to find out what kind of properties can be guaranteed by messaging schemes that respect a channel capacity constraint (namely that the number of bits sent over any time interval of length $\bar{s}$ does not exceed the maximum $b^+(\bar{s})$).

2. *Minimal communication rate:* given that a certain property should be guaranteed (e.g., observability, certain performance objective...), find a messaging scheme which guarantees this property, whilst sending as few as possible bits per unit of time, on average. If at communication instants j, the messaging scheme sends messages of b_j bits, the rate R is defined as (assuming that the limit exists)

$$R \coloneqq \lim_{j\to\infty} \frac{1}{j} \sum_{i=0}^{j} b_i.$$

Note that in some works of the literature, the limits are replaced by lim sup or lim inf (depending on which is applicable), to avoid existence problems.

The differences between these two terminologies (capacity versus rate) are:

- In the first case, the restriction comes from the communication channel and the objective is to find the best property that holds, whilst in the second case, the restriction is that a certain property should hold and the objective is to find the messaging scheme that achieves this with the lowest rate. In the second case, there is no fixed constraint.

- Because of its definition, the channel capacity might require arbitrarily long time intervals between communications $\bar{s}$, which, from an application point of view, is not realistic. The rate, on the other hand, is simply an average per unit of time of the number of bits that are sent.

The earliest works on this topic include [Wong and Brockett, 1997] which tackles state estimation for a stochastic plant whilst using the minimal communication rate approach, [Elia and Mitter, 2001] which considers the stabilization of a linear system with quantized input with the minimal communication rate approach, and [Nair and Evans, 2003] which considers the stabilization of a discrete-time linear system with a capacity approach. On the topic of state estimation, we further note [Simsek et al., 2004, Martins et al., 2006, Sahai and Mitter, 2006, Matveev and Savkin, 2007] which all provide solutions for the state estimation with limited data rate but assume a feedback communication link.

The rest of this section is dedicated to presenting some recent developments on the data rate approach. It is organized around several categories of problems that are tackled in the literature: linear systems versus nonlinear systems, sequential communication schemes versus absolute communication schemes, and single systems versus multiple systems.

Linear Versus Nonlinear

The problems involving data rate constraints have already been studied extensively. One can find surveys of the results for single-system configurations in [Nair et al., 2007] and for multiple systems / networks of systems in [Baillieul and Antsaklis, 2007, Hespanha et al., 2007, Andrievsky et al., 2010].

Linear Time-Invariant systems can be written in the following form:

$$x[k+1] = Ax[k] + Bu[k], \qquad \text{or} \qquad \dot{x}(t) = Ax(t) + Bu(t).$$

Most bounds on rates or capacity are then given in terms of the eigenvalues of the matrix A, the singular values of the matrix A, or some variant thereof. This is because the matrix A drives the dynamics of the system, that is, it determines how fast or how slow the system evolves, which in turn influences the necessary data rate to describe the system. For nonlinear systems, such an analysis is not as simple.

For generic nonlinear systems of the following form

$$x[k+1] = f(x[k], u[k]), \qquad \text{or} \qquad \dot{x}(t) = f(x(t), u(t)),$$

a direct characterization is not possible. For discrete-time systems, one might think that, in analogy to what is done for linear systems, it is possible to consider the singular values of $\frac{\partial f}{\partial x}(x(t), u(t))$, the Jacobian of f (a technique close to the first Lyapunov method). Similarly, for continuous-time systems, the singular values of the flow could be used. This approach is not viable on its own however since f is nonlinear, the singular values taken at some point in the state-space are meaningless to describe the global dynamics.

Some early results for nonlinear systems with specific structures were obtained in [De Persis, 2003, Baillieul, 2004]. To obtain results for systems with more general forms, some, like Liberzon et al. [Liberzon and Hespanha, 2005], opted to generalize techniques that were originally designed for linear systems (in the case of Liberzon et al., [Liberzon, 2003a]) to nonlinear systems. Others, like Fradkov et al. [Fradkov et al., 2008a], employed a passivity-based approach to solve this problem but the approach that received the most attention is the entropy-based method.

Entropy is a physical property that originated in the 1800s to measure the amount of disorder, randomness, or uncertainty in a certain system. In [Shannon, 1948], the author established the link between communication problems and entropy, which opened up the way for Adler et al. to define topological entropy in [Adler et al., 1965]. We borrow the following informal description of topological entropy from [Nair et al., 2004]:

> *"Briefly, the idea behind this definition is to first fix an open cover for the space, through which each iteration of the map is observed, i.e., all that is known is the sets of the open cover in which the iterations fall. Each observed open set is then inverted to yield an open set in the initial state space. As the number of iterations increases, the family of all possible intersections of initial state open sets forms an increasingly fine open cover for the space. The topological entropy of the map is then obtained by supremizing the asymptotic rate of increase of the cardinality of this open cover over all observation open covers. This in some sense measures the fastest rate at which uncertainty about the initial state can be reduced, or equivalently the fastest rate at which initial state information can be generated."*

The paper of Nair et al. [Nair et al., 2004], was among the first to use a notion of entropy to describe the minimal data rates required in a problem involving a dynamical system and a communication channel. In their case, they introduced a concept called feedback topological entropy, which measures the required data rate to stabilize a nonlinear system with a data rate-constrained feedback loop. Since then, topological entropy was employed many times [Pogromsky and Matveev, 2016a,b, Matveev and Pogromsky, 2016] as well as several other

notions of entropy, to quantify sufficient and/or necessary bit-rates to achieves various properties:

- *Stabilization entropy [Colonius, 2012]:* quantifies the minimum bit rates for the exponential stabilization of systems:
- *Invariance entropy [Kawan, 2013, Colonius et al., 2013, Kawan, 2011]:* quantifies the rate necessary to achieve invariance of a compact subset of the state-space;
- *Topological entropy for uncertain system [Savkin, 2006]:* an extension of topological entropy for systems with uncertain input;
- *Estimation entropy [Liberzon and Mitra, 2018, Sibai and Mitra, 2017, Kawan, 2018] :* quantifies the sufficient bit-rate to approximate system trajectories up to an exponentially decaying error;
- *Restoration entropy [Kawan et al., 2020, Matveev and Pogromsky, 2019, Kawan et al., 2021]:* quantifies "the minimal rate at which sensory data should be transferred to an estimator so that the initial estimation accuracy can be reproduced somewhere in the future then be maintained and, moreover, exponentially improved" [Matveev and Pogromsky, 2019].

For a comprehensive introduction of entropy and its usage in dynamical systems, one can refer to [Downarowicz, 2011].

Sequential Versus Absolute

Another important difference relates to the messaging scheme that is used: whether it functions in an absolute way (the meaning of the message can be derived independently of the other messages) or a sequential way (the meaning of the message depends on the previous messages). The difference between both strategies is best illustrated through an example. Consider the game of chess which is played on a board of 8×8 black-and-white squares. The most important piece is the king, which moves along the board, one square at a time. At any time instant, there are at most 8 possibilities for the king to move to (as illustrated on Figure 1.4). Imagine a king moving randomly on an empty board of chess, one square at a time, starting at the bottom left corner (An example of such a random walk is shown in Figure 1.5). One person observes the movements of the king and has to transmit them to another person, who cannot see the board, with the objective that the second person always knows on which square the king currently is. Here are two possible strategies to communicate the current position of the king:

1. *Absolute strategy:* Using the traditional coordinate scheme for chess, which consists of labelling the columns from a to g and the rows from 1 to 8 (see Figure 1.5), the person can simply give the current position of the king at each instant. For the walk of Figure 1.5, this would give the sequence of messages: a2, b3, c3, d2, e3, e2, d1. Because there are 8 columns and 8 rows, it takes 3 bits to encode for the row and 3 bits to encode for the column which implies that for such a strategy, 6 bits have to be sent at each time instant.

2. *Relative Strategy:* Because the king only has at most 8 legal moves at each time instant, it is possible to simply describe the movement that the king employed and combine this information with the last position of the king to provide the coordinates of a new square on which the king currently resides. Figure 1.5, this would give the sequence of messages: top, top right, right, bottom right, top right, bottom, bottom left. Since there are 8 legal moves, 3 bits are necessary to encode all possible messages, and hence for this strategy, 3 bits per time instant are necessary.

What are the advantages and disadvantages of each strategy? In terms of rate, the relative strategy is of course better than the absolute strategy. Over 7 time instants, 21 fewer bits are required to communicate. This is since the rate depends only on the dynamics of the king, rather than on the size of the chessboard. If the chessboard was infinite, the absolute strategy would require an infinite number of bits whereas the relative strategy would still require only 3 bits.
There is however a major drawback of using the relative strategy: imagine someone makes noise at the same time the observer is telling one of the messages, and this message is not heard by the remote peer. In that case, it is impossible to reconstruct the position of the king because the messages are always relative to the last known position. This is a default that the absolute strategy does not suffer from because a single message is sufficient to reconstruct the current position of the king, independently of the previous messages. The difference between the absolute strategy and relative strategy is thus that there is a trade-off: rate versus robustness towards losses.

The above example illustrates the difficulty that comes from dealing with losses in the communication channel whilst being subject to data rate constraints. In this particular case, it is the quantization scheme that is particularly important (as was already discussed in subsection 1.1.5.2). The above two strategies apply to quantization: absolute and relative. The absolute strategy consists of quantizing in such a way that no other information that the message itself is necessary to reconstruct the state of the system accurately. This strategy has a cost, however: possibly many more quantization regions are necessary to cover the entire state-space. Exactly how many are necessary depends on the

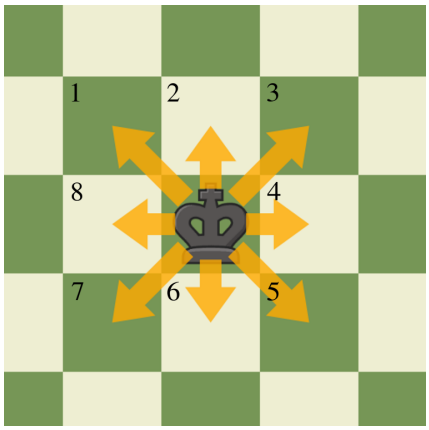

Figure 1.4: Possible moves of the king in the game of chess.

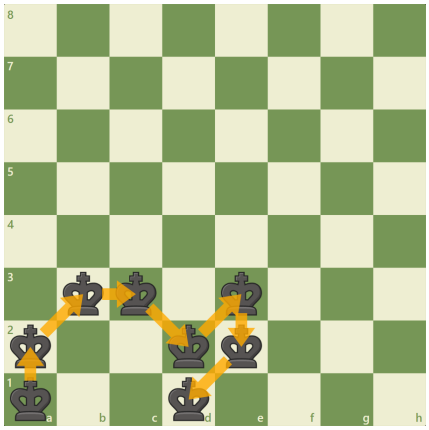

Figure 1.5: Example of a random walk of a king.

dimension of the state-space. For most systems, the state-space is simply $\mathbb{R}^n$ (or some subset thereof) but for some systems whose state-space has a fractal structure, other non-Euclidean dimensions are useful. Among them are the Hausdorff dimension ([Douady and Oesterle, 1980]), the Lyapunov exponents (which was proven to be related to the Hausdorff dimension in [Ledrappier and Young, 1985, Young, 1983], the limit capacity/box-counting dimension (see [Falconer, 1997, Takens, 1980] for a definition and [Rosenberg, 2020, Siegmund and Taraba, 2006] for numerical methods on how to compute it) and the Lyapunov dimension ([Kaplan and Yorke, 1978], which was proven to upper bound the box-counting dimension in [Hunt, 1996], and which can be computed via the second Lyapunov method ([Leonov, 2007, Kuznetsov, 2016]). For a general introduction to dimension theory for ODE's, one can refer to [Boichenko et al., 2005]. The relative quantization strategy does not suffer from this curse of dimensionality but its robustness towards losses remains a challenge. An example of sequential communication schemes can be found in [Matveev and Pogromsky, 2016].

It should also be noted that both strategies are not mutually exclusive. It is possible to alternate between both communication strategies to reduce the required communication rate whilst still maintaining some robustness towards losses. However, to the best of the author's knowledge, such a strategy has not been explored yet.

Single System Versus Multiple Systems

So far, most examples that were given included only one dynamical system and some communication device. In applications, it is very common to have not

just one but several systems interacting via wireless communication technologies. These are problems known as cooperation, decentralized schemes, and consensus/synchronization. These two forms of collaboration between dynamical systems should be understood as follows:

- *Cooperation:* Several dynamical systems interact to achieve a common control objective. *Example:* several mobile robots carrying a load together [Johnson and Bay, 1995].

- *Synchronization/Consensus:* Several dynamical systems interact such as their state-space trajectories coincide (up to a small, possibly vanishing, synchronization error). *Example:* formation control of flying vehicles [Eren et al., 2002].

One particular type of synchronization is master/slave(s) synchronization. In master/slave(s) synchronization, one system is considered the master while all others are considered slaves. The master system's state is then left unaffected while the slave systems have to track the master's state-space trajectories.

Communication problems generally require specific solutions to the multiple systems case. Some interesting results on consensus algorithms include [Yamaguchi et al., 2001], which provided a distributed consensus algorithms for a formation of robots, [Fax and Murray, 2004], which considered cooperative control of a formation of robots, [Olfati-Saber and Murray, 2004] which provides an algorithm for consensus in a network of integrators with fixed and switched topology and time-delays. For a more general reference on collaboration problems, one can refer to [Ren and Beard, 2008].

In terms of collaboration problems with data rate constraints, the earliest works include [Fradkov et al., 2008a] and [Fradkov et al., 2008a] which consider the synchronization of two chaotic systems and the master/slave synchronization of two systems respectively. Since then, more works on consensus in networks have been published, such as [Li et al., 2011] (average consensus in networks of linear systems with fixed topologies), [Li and Xie, 2011] (distributed average consensus with limited communication and varying network topology), [Liu et al., 2011] (average consensus in a network of systems with communication delays), [Dong, 2019] (consensus of a network of nonlinear systems with perturbations and a specific structure for the systems). To the best of the author's knowledge, no general results have been obtained for consensus in networks of nonlinear systems with generic structures.

This concludes the discussion on the solutions that have been provided so far for interactions between dynamical systems and communication technologies. In the next section, we discuss the contributions of this thesis to the field.

1.2 Research Problem and Contributions of this Thesis

As was highlighted in sections 1.1.2 and 1.1.3, many problems originate from the combination of one or several dynamical systems with communication technologies. A non-exhaustive list of such problems is:

I. Considering several sources of uncertainty at the same time. This is illustrated in Figure 1.6. Most problems fit in one of the three circles but to the best of the author's knowledge, few tackle two at the same time whilst none tackle all three. An open problem is thus to design solutions for problems combining two or three sources of uncertainty;

II. Developing communication protocols that deal with both limiting factors of wireless communication technologies (limited transmission rate and limited packet size) whilst also having robustness to the disturbances (packet losses and packet corruption);

III. Developing data-efficient algorithms for cooperation problems of dynamical systems with generic structures;

IV. Using an event-triggered approach to the combined rate problem, rather than for sampling and quantization problems separately;

V. Developing tools for control-oriented interactions between nonlinear systems with generic structures and communication technologies;

VI. Validating tools for interactions between dynamical systems and communication technologies through experiments.

This thesis aims to tackle some of these problems. The research problem could thus be summarized as follows:

> **Research Problem:** To develop general tools for control-oriented interactions between dynamical systems and communication technologies.

In this thesis, five different solutions are developed, each aiming to provide a (partial) solution to one or several of the aforementioned open problems.

1.2.1 First Contribution

The first important problem is the design of observers that function over communication channels with limited communication capacity. A single deterministic nonlinear system is connected to a remote location via a data rate constrained communication channel. Initially, only an estimate of the state is available at the remote location which implies that there is a **sensitivity to the initial conditions**. The objective is to design a communication protocol, in the form

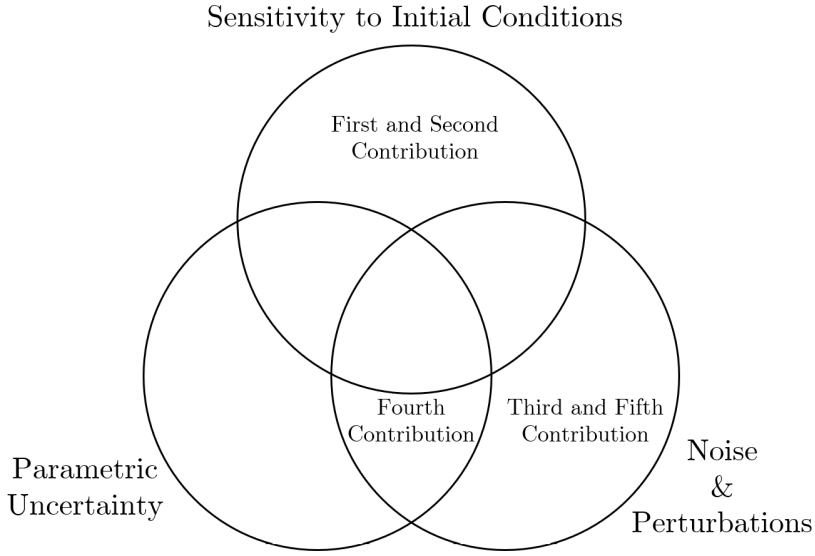

Figure 1.6: The three sources of uncertainty for dynamical systems and how the contributions of the thesis relate to them.

of several devices that generates estimates of the state of the system at the remote location. The protocol should satisfy two criteria: the error between the actual state and the estimate of the state should be either bounded, or vanishing, and the channel capacity should not be exceeded. In [Matveev and Pogromsky, 2016], the authors provide analytical bound on the minimum capacity required for several types of observability, but the communication protocol is not robust towards losses in the communication channel. Since channel dropouts are a common problem with wireless communication technologies (as was highlighted in 1.1.3), developing a scheme that is robust towards losses constitutes an important contribution. The first contribution of the thesis is (referring to items II. and V. of the above list):

First Contribution: A data rate constrained observation scheme for nonlinear systems (both continuous- and discrete-time), with sensitivity to initial conditions, that is robust towards losses in the communication channel is developed.

For this observer, analytical bounds on the minimum sufficient communication capacity are obtained to observe the system through a data rate constrained communication channel. The bound is proven to depend on the singular values of the Jacobian of the system's map (for the discrete-time case) and the singular

values of the Jacobian of the system's flow (for the continuous-time case), as well as the Lyapunov dimension of the state-space of the system. The novelty of this result is the robustness of the observer towards losses in the communication channel. It is developed in Chapter 2.

1.2.2 Second Contribution

As was highlighted in the part of section 1.1.5 dedicated to configurations with multiple systems, it is sometimes desirable for dynamical systems to interact to achieve some collective behaviour. If several dynamical systems interact through a network of data rate constrained communication channels, it is not always possible to achieve consensus. In particular, in some applications, the device that measures the state is not directly connected to the controller of each system, which implies that even to have an estimate of its own system's state, the controller needs to receive messages which are sent over the communication channel. When combined with a **sensitivity to initial conditions** of each system, it results in a problem that requires a specific solution to achieve and maintain consensus. So far, no general communication protocols have been derived for such configurations. Whether it is possible or not to achieve consensus, under what conditions on the adjacency matrix which describes the connections between the systems, what channel capacities are required, are all open problems to which we provide the following answers (referring to items III. and V.):

> **Second Contribution:** For a network of agents described by identical discrete-time dynamical systems, several consensus protocols (consisting of several interacting devices) are obtained (for different network topologies) which keep the agents in consensus whilst limiting the load on the communication.

For each of these consensus protocols, analytical bounds on the required channel capacity are provided in terms of quantities that depend on the equations describing the dynamical systems. The novelty of this result is the three consensus protocols for nonlinear systems. This result is developed in Chapter 3

1.2.3 Third Contribution

The third contribution is the design of a communication protocol for the remote observation through a data rate constrained communication channel of a Lipschitz-nonlinear continuous-time system with **state perturbations** and **measurement noise**. The problem is approached from the point of view of the rate, rather than capacity. Bounded state perturbations and measurement noise are assumed, which makes a deterministic LMI approach suitable for this problem. This forms the following contribution (referring to items IV. and V.):

Third Contribution: An event-triggered communication protocol for the remote observation of continuous-time Lipschitz-nonlinear systems with bounded state perturbations and measurement noise is developed.

For this remote observer, a bound on the maximum rate and maximum observation is provided, in function of tunable constants. Through the tunable constants, a trade-off is possible of precision for rate and vice-versa. The inclusion of the event-triggered scheme leads to a greatly reduced average communication rate, as is proven through simulations on various examples. The novelty of this result is an event-triggered scheme for generic continuous-time systems with Lipschitz-nonlinear structures. It is developed in Chapter 4

1.2.4 Fourth Contribution

The fourth contribution of the thesis is a solution for the remote observation of a steered system. A single discrete-time Lipschitz-nonlinear system is steered by an external, a priori unknown signal, which is measured with zero measurement error. This signal plays the role of a **parametric uncertainty**. The system is subject to **state perturbations** and only an output measurement with **measurement noise** is available. The objective is to send messages to a remote location such that an estimate of the state is available remotely. This property should be achieved whilst using as few bits per unit of time as possible. The following solution is developed (referring to items I., IV., and V.):

Fourth Contribution: An event-triggered communication protocol for the remote observation of steered discrete-time Lipschitz-nonlinear systems with bounded state perturbations and measurement noise is obtained.

Again, an event-triggered scheme is used to reduce the average number of communications. Bounds on the maximum observation error and maximum communication rate are provided in function of the system's constants. The novelty of this result consists of an event-triggered scheme for discrete-time systems with Lipschitz-nonlinear structures, perturbations, and parametric uncertainty. It is developed in Chapter 5.

1.2.5 Fifth Contribution

The fifth and final contribution of the thesis is a solution for the remote observation of unicycle-type robots, with experimental validation. A unicycle-type robot is driven by a steering signal. It is connected to a remote location via a data rate constrained communication channel. The objective is to reconstruct

the position of the robot, whilst using an as low as possible number of bits per unit of time. The following solution is developed (referring to items IV. and VI.):

Fifth Contribution: An event-triggered communication protocol for the remote observation of unicycle robots, with experimental validation, is developed and demonstrated.

Due to the experimental nature of the contribution, *perturbations* naturally occur, as well as *sensitivity to initial conditions*. An event-triggered scheme is used, to reduce the average number of communications. Bounds on the maximum observation error and communication rate are computed. The communication scheme is then tested through experiments on Turtlebot robots. The novelty of the result consists of implementing the protocol on a real-life problem. It is developed in Chapter 6.

1.3 Structure of the Thesis and List of Publications

Each chapter of this thesis corresponds to one publication in a peer-reviewed journal. The content of the papers has only been modified to reflect the comments of the members of the thesis committee and for layout issues. Each chapter is self-contained and can be read independently.
Chapter 2 corresponds to:

- Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. "Data-Rate Constained Observers of Nonlinear Systems". In: *Entropy* 21.282 (2019), pages 1-29, [Voortman et al., 2019].

which is the continuation of:

- Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. "Continuous Time Observers of Nonlinear Systems with Data-Rate Constraints". In: *Proceedings of the 5th IFAC Conference on Analysis and Control of Chaotic Systems*. Eindhoven, 2018, [Voortman et al., 2018a].

- Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. "A Data Rate Constrained Observer for Discrete Nonlinear Systems". In: *Proceedings of the 57th IEEE Conference on Decision and Control*. Miami Beach, 2018, [Voortman et al., 2018b].

Chapter 3 corresponds to:

- Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. "Data-Rate Constrained Consensus in Networks of Dynamical Systems", *Preprint submitted to Automatica*, [Voortman et al., 2020e].

which is the continuation of:

- Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. "Consensus of Nonlinear Systems with Data-Rate Constraints". In: *Proceedings of the 21st IFAC World Congress*. Berlin, 2020, [Voortman et al., 2020d].

Chapter 4 corresponds to:

- Q. Voortman, D. Efimov, A. Y. Pogromsky, J.-P. Richard, and H. Nijmeijer. "An Event-Triggered Observation Scheme for Systems with Perturbations and Data-Rate Constraints", *Preprint submitted to Automatica*, [Voortman et al., 2020a].

which is the continuation of:

- Q. Voortman, D. Efimov, A. Y. Pogromsky, J.-P. Richard, and H. Nijmeijer. "Event-triggered Data-efficient Observers of Perturbed Systems". In: *Proceedings of the 21st IFAC World Congress*. Berlin, 2020, [Voortman et al., 2020b].

Chapter 5 corresponds to:

- Q. Voortman, D. Efimov, A. Y. Pogromsky, J.-P. Richard, and H. Nijmeijer. "Tracking State with Limited Communications: an Event-Triggered Approach", *Preprint submitted to IEEE Transactions on Automatic Control*, [Voortman et al., 2021a].

which is the continuation of:

- Q. Voortman, D. Efimov, A. Y. Pogromsky, J.-P. Richard, and H. Nijmeijer. "Synchronization of Perturbed Linear Systems with Data-Rate Constraints". In: *Proceedings of the 59th IEEE Conference on Decision and Control*. Jeju Island, 2020, [Voortman et al., 2020c].

Chapter 6 corresponds to:

- Q. Voortman, D. Efimov, A. Y. Pogromsky, J.-P. Richard, and H. Nijmeijer. "Observing Mobile Robots with Data-Rate Constraints: a Case Study", *Accepted for publication at the 60th IEEE Conference on Decision and Control*. Austin, 2021, [Voortman et al., 2021b].

Chapter 2

Data-Rate Constrained Observers of Nonlinear Systems

In this chapter, the design of a data-rate constrained observer for a dynamical system is presented. This observer is designed to function both in discrete time and continuous time. The system is connected to a remote location via a communication channel which can transmit limited amounts of data per unit of time. The objective of the observer is to provide estimates of the state at the remote location through messages that are sent via the channel. The observer is designed such that it is robust toward losses in the communication channel. Upper bounds on the required communication rate to implement the observer are provided in terms of the upper box dimension of the state space and an upper bound on the largest singular value of the system's Jacobian. Results that provide an analytical bound on the required minimum communication rate are then presented. These bounds are obtained by using the Lyapunov dimension of the dynamical system rather than the upper box dimension in the rate. The observer is tested through simulations for the Lozi map and the Lorenz system. For the Lozi map, the Lyapunov dimension is computed. For both systems, the theoretical bounds on the communication rate are compared to the simulated rates.

2.1 Introduction

Ever since the widespread usage of wireless technologies, there has been a focus on data-rate problems for dynamical systems. These problems arise when networked technologies are employed in configurations where sensors, actuators, and controllers are placed at locations that are remote from one another. Extra complications arise from uncertainties in the system's parameters, initial conditions, sensor measurements, communication channels, and dynamics, such as in

the form of exogenous disturbances. These necessitate communication strategies that are efficient in terms of data rate and robust against all kinds of uncertainty. In this chapter, the focus will be placed on uncertainties in the initial conditions, and also on issues of losses in the communication channel.

Up until now, the main focus in the relevant control-oriented literature was on state estimation and stabilization problems. In the early 2000's, most of the research dealt with linear systems, for which many results have been obtained (see [Elia and Mitter, 2001, Nair et al., 2007, Baillieul and Antsaklis, 2007, Andrievsky et al., 2010] for extended surveys).

Early results on nonlinear systems (exemplified by [De Persis, 2003, Baillieul, 2004]) typically assumed special properties of the considered systems, and were also based on these properties. Proper adapting and extending techniques, originally developed for linear plants, opened the door for handling generic nonlinear systems and permitted the establishment of lower data-rate bounds sufficient for the observability and stabilizability of such systems; see, e.g., [Liberzon and Hespanha, 2005]. Another research trend was concerned with intrinsic characteristics of nonlinear systems that provide a somewhat exhaustive description of the bit-rate of information transmission under which a certain dynamic property (such as stability, invariance, observability) can be achieved. As a result, there appeared a whole series of extensions and modifications of classical topological entropy [Adler et al., 1965], such as feedback topological entropy [Nair et al., 2004], invariance entropy [Kawan, 2013, Colonius et al., 2013], topological entropy with regard to dynamic uncertainties [Savkin, 2006] and to uncertainties in the initial conditions [Kawan and Yüksel, 2018], estimation entropy [Liberzon and Mitra, 2018], and others (see, e.g., [Matveev and Savkin, 2009, Kawan, 2018, Sibai and Mitra, 2017, Pogromsky and Matveev, 2016a]). Finally, some papers such as [Fradkov et al., 2008a] have relied on passivity-based methods to provide bit-rate bounds.

One of the objectives of the chapter is to use non-Euclidean concepts of the set dimension as an alternative to the aforementioned notions of entropy. Among the non-Euclidean concepts of set dimensions, the best known is, maybe, the Hausdorf dimension [Douady and Oesterle, 1980]. Another related characteristic is the upper box dimension [Falconer, 1997], which is sometimes referred to as the limit capacity [Takens, 1980]. These two dimensions—the entropy and Lyapunov exponent—were proven to be closely related to one another in [Kawan, 2011, Young, 1983, Ledrappier and Young, 1985]. Both dimensions are based on covering a set with balls of infinitesimally small size (a technique which is very similar to the idea of partitioning the state-space and using symbolic dynamics to describe the dynamical system — see, e.g., [Stojanovski et al., 1997]); both can assume non-integer values, and both may be smaller than the dimension of the hosting Euclidean space. These concepts have been much inspired by studies of fractals and research on chaotic attractors of dynamical systems. The latter is a primary incentive for our interest in these dimensions, which may be

a non-integer for chaotic attractors and provide a somewhat deeper insight into the issue of their dimensionality than the ordinary Euclidean dimension.

Unfortunately, there are still no general analytical techniques for computing the two aforementioned dimensions for chaotic attractors. Numerical methods remain the main tool used by scientists and engineers to estimate these dimensions [Siegmund and Taraba, 2006]. In this chapter, an alternative to this numerical approach is developed. The alternative is based on the so-called Lyapunov dimension, in which the upper bounds are the above two dimensions [Hunt, 1996]. Moreover, its advantage resides in the fact that the Lyapunov dimension can be computed analytically by using the second Lyapunov method [Leonov, 2007, Kuznetsov, 2016], which leads to analytical upper bounds. By following this alternative, we obtain a fully analytical lower bound on the communication data-rate under which reliable estimation of the system's state becomes feasible. When doing so, we consider a generic nonlinear system and focus attention on its behavior within a given invariant set, which may be a chaotic attractor, for example.

Apart from this bound, the design of a particular observer that ensures such an observability is also presented. The observer is composed of a sampler, a quantizer, a data-rate constrained channel, and a decoder. All components interact in order to build estimates of the state at a remote location in real time. The observer can ensure arbitrarily high precision of estimation with a communication rate that remains below the channel capacity. Moreover, the proposed observer is robust against delays and losses in the communication channel, which is a valuable property for applications where delays and losses are a common occurrence. This robustness is achieved without any feedback in the communication channel, which is atypical for most data-rate constrained observers in the current literature [Sahai and Mitter, 2006, Martins et al., 2006, Simsek et al., 2004, Matveev and Savkin, 2007] and constitutes the novel contribution of the chapter.

This chapter is both a generalization and an extension of [Voortman et al., 2018a,b]. We provide a unified solution for both continuous and discrete-time systems. In addition to providing proofs for all the results that were presented in the two aforementioned conference papers, the problem statement is extended to also include delays in the communication channel.

In Section 2.2, we define the types of systems to be observed, as well as the observer notations, and provide a definition for observability with data-rate constraints. Section 2.3 introduces the proposed observer. In Section 2.4, preliminary criteria for observability of the plant are offered, which are converted into a fully analytical form in Section 2.5. Section 2.6 illustrates the general theory via handling two examples: the Lozi map and the Lorenz system. For both systems, the necessary data-rates are computed and simulations that confirm the theoretical results are provided.

2.2 Problem Statement

In this section, we introduce the problem statement. The general setting is that of a dynamical system and two peers connected together via a communication channel. Both peers have full knowledge about the dynamics of the system. Meanwhile, only one of them has direct access to the system's current state and fully measures it. The task is to provide estimates of the state to the other peer by sending messages through the communication channel. This channel is discrete (i.e., the variety of transmittable messages is finite) and has delays, losses, and limited data-rate. The effects due to data-rate constraints and delays are explicitly modeled in this section, whereas the issue of message losses is discussed separately in Remark 2.8. Two types of delays will be considered in turn. A *processing delay* is also incurred, since the channel can transmit only a given and finite number of bits c per unit time, and so a B-bits message can wholly arrive at the receiving end of the channel not earlier than B/c time units after the transmission of this message is commenced. A *transmission delay* is caused by holding up the progress in the ideal routine of bits transfer, which may occur as a result of, e.g., resolving competition with third parties for shared resources of the communication medium or network.

In order to solve the stated problem, we will develop a particular type of observer. In this section, we will only introduce notations concerned with the observer. Operation of its components will be described in the next section.

2.2.1 Observed Dynamical System

We consider a dynamical system $\{\varphi^t\}_{t\in T}$ on an open set $S \subset \mathbb{R}^n$, paying special attention to a certain subset $S_0 \subset S$. Here,

- T is the set of time periods, which is either $\mathbb{Z}^+$ or $\mathbb{R}^+$;
- $\varphi^t : S \to S$ is the evolution function that gives the system state $x(t) = \varphi^t(x_0)$ at time $t \in T$, provided that the initial state is x_0;
- S_0 is the focus of our interest in the system.

Specifically, we are interested only in trajectories that start in S_0 and remain there afterwards:

$$x(t) \in S_0 \qquad \forall t \in T. \tag{2.1}$$

Assumption 2.1. *The dynamical system at hand is time-invariant:* $\varphi^t \circ \varphi^s = \varphi^{t+s}\ \forall t, s \in T$.

Assumption 2.2. *The set* S_0 *is a bounded forward invariant* $\varphi^t(S_0) \subset S_0\ \forall t \in T$, *its closure* $\overline{S}_0$ *lies in* S.

Typical examples of sets S_0 which satisfy the aforementioned assumption are chaotic attractors. Indeed, chaotic attractors are bounded and forward invariant by nature. If additionally the closure of the attractor lies in S, all three conditions of the assumption are met and the chaotic attractor can be used as the set S_0.

Our main interest is in systems for which complex and possibly chaotic long-horizon dynamics arise from rather regular short-horizon behavior. The last feature is partly substantiated by the following.

Assumption 2.3. *For any $t \in T$, the evolution function $\varphi^t : S \to S$ is continuously differentiable.*

Depending on the "time-set" option, two types of dynamical systems will be considered.

(1) Discrete time systems: $T = \mathbb{Z}^+$, and the system evolves as follows:

$$x(t+1) = \psi(x(t)) \quad t \in \mathbb{Z}^+, \qquad ,$$

where $\psi : S \to S$ is a given mapping. In this case,

$$\varphi^t(\cdot) \coloneqq \underbrace{\psi(\dots\psi(\cdot))}_{t \text{ times}} \quad \text{and} \quad \psi = \varphi^1.$$

Assumption 2.1 holds, and Assumption 2.3 is met if, and only if ψ is continuously differentiable.

(2) Continuous time systems: $T = \mathbb{R}^+$ and the evolution of the system is described by an ordinary differential equation (ODE):

$$\dot{x}(t) = f(x(t)) \quad t \in \mathbb{R}^+, \tag{2.2}$$

where $f : S \to \mathbb{R}^n$ is a continuously differentiable vector field. So for any $x_0 \in S$, the solution $x(t, x_0)$ of the Cauchy problem $x(0) = x_0$ for the ODE (2.2) exists, and is unique; it can be extended to the right on the maximal interval $[0, T(x_0))$. However, Equation (2.2) not ineluctably defines a dynamical system on S, since, not necessarily, $T(x_0) = \infty$ for all $x_0 \in S$. Insofar as the right-hand side of Equation (2.2) is not defined outside S, this extendability $T(x_0) = \infty$ means, in particular, that the solution $x(t, x_0), x_0 \in S$ never attempts to leave the set S; i.e., the set S is forward invariant. So when dealing with ODE, we always assume that all its solutions that start in S at $t = 0$ can be extended on $[0, \infty)$ while remaining in S.

The following proposition can be proved by retracing the arguments from Section 2.2 in [Khalil, 2002].

Proposition 2.4. *Whenever the vector field $f : S \to \mathbb{R}^n$ is smooth (i.e., continuously differentiable) and the ODE* (2.2) *has the just-stated extendability and*

invariance properties, this ODE gives rise to a dynamical system $\{\varphi^t\}_{t\in\mathbb{R}^+}$ *on* S ($\varphi^t(x_0) := x(t, x_0)$)*, which satisfies Assumptions* 2.1 *and* 2.3*. Moreover,* $\varphi^t(x)$ *and its first derivatives, with respect to* t *and* x*, are continuous functions of* t *and* x*.*

2.2.2 Architecture of the Observer, Notations, and General Traits of the Communication Channel

We assumed that the current state $x(t)$ was observed in full at a certain *measurement site* but is needed at time t at a remote location, where data can be communicated only via a discrete channel. The channel is discrete in the sense that first, it is constrained to carry messages that are drawn from a finite set, and second, the messages can be communicated only one at a time and, while the channel is busy transmitting a previous message, it is closed for the next transmission.

The purpose of the observer is to arrange and manage transmissions across the channel and to finally build, at time t and at the remote location, an estimate $\hat{x}(t)$ of the current state $x(t)$ with a pre-specified exactness. The formal definition of the last notion is as follows.

Definition 2.5. *A number* $\epsilon > 0$ *is called an "exactness of observation" if there exists* $\bar{t}_0 < \infty$ *such that*

$$\|x(t) - \hat{x}(t)\| \le \epsilon, \qquad \forall t \in T : t \ge \bar{t}_0.$$

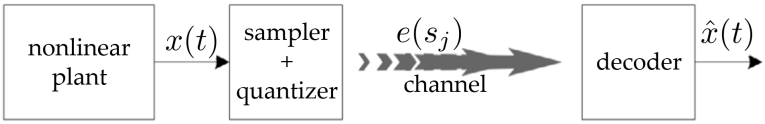

Figure 2.1: Structure of the observer.

As is illustrated in Figure 2.1, an *observer* $\mathscr{O}$ is defined as a composition consisting of a *sampler* $\mathscr{S}$, *quantizer* $\mathscr{Q}$, and *decoder* $\mathscr{D}$; the sampler and quantizer together form a *coder* $\mathscr{C}$:

$$\mathscr{O} \text{ is composed of } \underbrace{\mathscr{S} \text{ and } \mathscr{Q}}_{\mathscr{C}} \text{ and } \mathscr{D}.$$

- The sampler and quantizer are built at the measurement site and have access to the dynamics $\{\varphi^t\}$ of the system, the set S_0, the current state $x(t)$, and the desired exactness of observation ϵ.
- The decoder is built at the remote site $\mathcal{L}$ and has access to the system dynamics $\{\varphi^t\}$, the set S_0, the desired exactness of observation ϵ, and the messages transmitted across the channel.

The roles and structures of the observer components are as follows.

The Sampler generates the (*sampling*) instants $s_j \in T$ (where at every one of these instants $t = s_j$, transmission of another message $e(s_j)$ is initiated):

$$s_{j+1} = \mathscr{S}(\{\varphi^t\}_{t\in T}, S_0, x(s_j), s_j, \epsilon) > s_j, \quad s_0 = 0, \qquad j \in \mathbb{Z}^+. \tag{2.3}$$

Also, the sampler builds a finite *alphabet* $\mathcal{A}_j$ from which the message $e(s_j)$ should be drawn at time s_j for subsequently communicating across the channel:

$$\mathcal{A}_j = \mathscr{A}(\{\varphi^t\}_{t\in T}, S_0, x(s_j), s_j, \epsilon), \qquad j \in \mathbb{Z}^+. \tag{2.4}$$

The alphabet is thus permitted to depend on s_j.

The Quantizer forms the message $e(s_j) \in \mathcal{A}_j$ to be dispatched

$$e(s_j) = \mathscr{Q}(\{\varphi^t\}_{t\in T}, S_0, x(s_j), s_j, \epsilon), \qquad \forall j \in \mathbb{Z}^+. \tag{2.5}$$

The Decoder generates state estimates based on the previously received messages:

$$\hat{x}(t) = \mathscr{D}(\{\varphi^\tau\}_{\tau\in T}, S_0, \{(e(s_j), \bar{s}_j)\}_{j\in J(t)}, \epsilon), \tag{2.6}$$

where $\bar{s}_j$ is the time when the message $e(s_j)$ arrives at the remote site $\mathcal{L}$ and $J(t) := \{j : \bar{s}_j \le t\}$. If no message has arrived yet, $J(t) = \emptyset$ and the meaningless $\{(e(s_j), \bar{s}_j)\}_{j\in\emptyset}$ is replaced by an arbitrarily pre-specified symbol, e.g., $0 \in \mathbb{Z}^+$. The observer has to fit the constraints and capabilities of the channel, which are as follows.

c.1) The channel correctly transfers any message $e(s_j) \in \mathcal{A}_j$ to the receiving end provided that the message processing time τ_j^{pr} and the size of the message are in balance:

$$\log_2 \operatorname{card}(\mathcal{A}_j) \le b(\tau_j^{\mathrm{pr}}). \tag{2.7}$$

Here, $b(\tau)$ is a channel-dependent function that gives the number of bits processable by the channel during any time period of length τ.

c.2) As the processing time increases to infinity, the average number of bits transmittable per unit of time stabilizes and converges to a certain value $c \in \mathbb{R}^+$, called the (bit-rate) channel *capacity*:

$$\exists c := \lim_{\tau\to\infty} \frac{b(\tau)}{\tau}. \tag{2.8}$$

c.3) The channel is closed for the next message until all bits of the current message $e(s_j)$ have been processed, but is open afterwards.

c.4) On its way to the destination point $\mathcal{L}$, any message $e(s_j)$ incurs a transmission delay τ_j^{tr}:

$$\bar{s}_j = s_j + \tau_j^{\mathrm{pr}} + \tau_j^{\mathrm{tr}}, \tag{2.9}$$

where $\bar{s}_j$ is the time when the whole of the message $e(s_j)$ arrives at $\mathcal{L}$, and the processing time τ_j^{pr} plays the role of a processing delay here.

c.5) The transmission delays are upper-bounded: $\tau_j^{\mathrm{tr}} \leq \tau_+^{\mathrm{tr}} < \infty$.

To correctly transmit messages, the sampler should balance the chosen alphabet and the message processing time τ_j^{pr} in accordance with Equation (2.7), and respect the requirement: $s_{j+1} \geq s_j + \tau_j^{\mathrm{pr}}$.

2.2.3 Observability via Channels with Limited Bit-Rate Capacity

The objective of the observer is to guarantee observability, as defined in the following definition.

Definition 2.6. *A system $\{\varphi^t\}_{t\in T}$ is said to be observable on the set S_0 via a communication channel if, for any $\epsilon > 0$, there exists an observer Equations (2.3)–(2.6) that operates via this channel and ensures the requested exactness of observation ϵ for any trajectory satisfying Equation* (2.1).

Observability is classically defined as a property of the system itself. However, in the current context, finite data rate makes observability critically dependable on the employed communication channel. So, by following [Pogromsky and Matveev, 2016a, Matveev and Pogromsky, 2017, Voortman et al., 2018a,b], observability is introduced as a property of the pair "system + channel". In Definition 2.6, the reference to existence of an observer in fact conveys the idea of most effectively utilizing the properties of the system and the potentialities of the channel, where if their clever use may result in a reliable and exact state estimate at the receiving end of the channel, the pair is sealed with the stamp, "observable".

2.3 Design of the Proposed Observer

Since we are interested only in trajectories satisfying Equation (2.1), our discussion of the observer design is confined to the case where $x(s_j) \in S_0 \ \forall j$ in Equations (2.3)–(2.5).

We will introduce an observer that is determined by the following four entities:

e.1) $s_+ \in T$ — the period $s_{j+1} = s_j + s_+$ between consecutive dispatches of messages via the channel;

e.2) P — a symmetric and positive definite $n \times n$-matrix;

e.3) $\delta(\epsilon, s_+) > 0$ — a function of $\epsilon > 0$ and s_+ for which

$$\|\hat{x} - x\|_P \leq \delta(\epsilon, s_+) \text{ and } \hat{x}, x \in S_0 \Rightarrow \left\|\varphi^t(\hat{x}) - \varphi^t(x)\right\|_P \leq \epsilon \ \forall t \in T, t \leq s_+. \tag{2.10}$$

e.4) $\{B_\delta^P(q_k)\}_{k=1}^K$ — a finite covering of the compact set $\overline{S}_0$ with $K = K(\epsilon, s_+)$ balls (with respect to the norm $\|\cdot\|_P$) centered in $q_i \in S_0$ and with a radius of $\delta = \delta(\epsilon, s_+)$ each.

Here, the centers q_i may also depend on ϵ, s_+.

Lemma 2.7. *Let Assumptions 2.2 and 2.3 hold, and for any $\tau \in T$, the derivatives of φ^t are bounded over $t \in T, t \leq \tau$ and x from some τ-dependent neighborhood of S_0. Then, a function $\delta(\epsilon, s_+) > 0$ with the property (2.10) exists.*

The proof of this lemma simply follows from the continuous differentiability of φ^t and the boundedness of of the derivates.

The proposed observer operates as follows.

Procedure 1. (Observer)

o.1) The sampler $\mathcal{S}$ (Equations (2.3) and (2.4)) carries out the following actions:

$$\begin{aligned} s_{j+1} = &\mathcal{S}(\{\varphi^t\}_{t\in T}, x(s_j), s_j, \epsilon) := s_j + s_+, \qquad s_0 := 0. \\ &\mathcal{A}(\{\varphi^t\}_{t\in T}, x(s_j), s_j, \epsilon) := \{1, \ldots, K\}, \end{aligned}$$

i.e., the alphabet substantiates the numbering of the balls from e.4).

o.2) The quantizer $\mathcal{Q}$ finds an element $B_\delta^P(q_k)$ of the covering from e.4) that contains $x(s_{j+1}) = \varphi^{s_+}(x(s_j)) \in S_0$ and sends its index k over the channel:

$$\mathcal{Q}(\{\varphi^t\}_{t\in T}, x(s_j), s_j, \epsilon) = k,$$

o.3) The decoder $\mathcal{D}$ performs the following operations at time $t \in T$:

- *Extracts the index k from the last message received at a time $\theta \leq s_i$, where $i := \lfloor t/(s_+) \rfloor$ (If no message has been received yet, k is assigned an arbitrarily pre-specified value, e.g., 1.).*
- *By using the centers from e.4), forms the current state estimate*

$$\hat{x}(t) := \varphi^{(t-s_i)}(q_k). \tag{2.11}$$

Several comments on this observer are as follows:

- In o.2), we do not address the case $x(s_{j+1}) \notin S_0$ due to the reason stated at the onset of the section.
- The proposed design assumes that both the coder and decoder have access to s_+ from e.1) and the covering from e.4).
- The observer uses a fixed alphabet $\{1, \dots, K\}$, which is shared by the coder and the decoder, where K is the number of balls of radius ϵ in the covering of $\bar{S}_0$ (as defined in e.4))
- The quantizer sends data about the estimate q_k of not the current $x(s_j)$, but the forward-time state $x(s_{j+1})$, which is computed from the measured $x(s_j)$ by using the known transition map $\varphi^{s_+}(\cdot)$.
- The idea behind this relies on the expectation that these data will be received prior to s_{j+1} and put in use at due time, $t = s_{j+1}$. Then, the exactness of estimation will be δ at this time.
- These data are also used to estimate the state on the subsequent time interval $\{t \in T : s_{j+1} \leq t < s_{j+2}\}$ via applying the matching transition map to the just-discussed estimate at time $t = s_{j+1}$. By Equation (2.10), this guarantees the exactness of estimation $\|x(t) - \hat{x}(t)\|_P \leq \epsilon$ on this interval.

In order for the proposed observer to be able to *operate correctly* via a given communication channel, the message $e(s_j)$ initiated at time s_j should be fully processed and received prior to s_{j+1}. (This, in particular, implies that the messages arrive in order: $\bar{s}_{j+1} > \bar{s}_j$.) Due to c.1), c.4), and c.5), correct operation occurs whenever there exists a solution $\tau^{\,\mathrm{pr}} \in T$ to the following two inequalities:

$$\log_2 K(\epsilon, s_+) \leq b(\tau^{\mathrm{pr}}), \qquad s_+ \geq \tau^{\mathrm{pr}} + \tau_+^{\mathrm{tr}}. \tag{2.12}$$

We recall that τ_+^{tr} is an upper bound on the transmission delay. In the typical case where $K(\epsilon, \tau)$ is an increasing function of τ and modulo the possibility to choose s_+, Equation (2.12) reduces to only one inequality:

$$\log_2 K(\epsilon, \tau^{\mathrm{pr}} + \tau_+^{\mathrm{tr}}) \leq b(\tau^{\mathrm{pr}}). \tag{2.13}$$

Anyhow, the inequalities depend on both the system (via $K(\cdot, \cdot)$) and channel (via $b(\cdot), \tau_+^{\mathrm{tr}}$). This means that correct operation in fact requests a certain level of conformity between the system and the channel.

The conditions for correct operation will be fleshed out in the next section. We conclude the section with a comment on observability and a remark on how the observer proceeds when a loss occurs.

Observation 1. *The following statements are true:*

i) Let the proposed observer correctly operate for a given $\epsilon > 0$ and s_+. Then, for any trajectory satisfying Equation (2.1)*, the desired exactness of observation ϵ is ensured with respect to the norm $\|\cdot\|_P$;*

ii) Let a communication channel be given. Also, let any $\epsilon > 0$ small enough be coupled with some s_+ so that the proposed observer with these ϵ and s_+ operates correctly via the channel at hand. Then, the system $\{\varphi^t\}_{t\in T}$ is observable on the set S_0 via this communication channel.

All observability conditions that will be established in this chapter are nothing but implications of ii) in this observation. This means that these conditions ensure the correct operation of the proposed observer modulo's proper and feasible choice of its parameters. In other words, whenever these conditions are satisfied, a reliable state estimate can be obtained by means of this observer.

Remark 2.8. *Suppose that messages may be lost when transmitting over the communication channel. If a loss does occur, the message q_k which was last received is used in Equation* (2.11) *not only during the intended time interval (from s_i to s_{i+1}), but also during the subsequent time intervals until the next successful transmission. Certainly, there is no guarantee that the estimation accuracy will be within the desired ϵ on these extra intervals. However, as soon as a new message arrives, this accuracy is restored due to the very design of the observer. This robustness against losses is achieved without any feedback in the communication channel (i.e., the coder is not notified when losses occur on the channel), unlike many competing schemes [Sahai and Mitter, 2006, Martins et al., 2006, Simsek et al., 2004, Matveev and Savkin, 2007].*

Note that although the robustness towards losses is one of the main advantages of the designed observer, the discussion on the situation with losses is relatively short in this chapter. This is intentional as the authors believe that providing any further comments would require some form of a statistic model for the losses. For brevity such a discussion is left for further research. This remark extends on the situation where the message is not lost, but corrupted so that an incorrect q_k is occasionally used in Equation (2.11).

2.4 Criteria for Observability of the System

A problem with the conditions (2.12) and (2.13) is that they use the function $K(\cdot,\cdot)$ from e.4), for which there is a lack of constructive techniques to compute, or at least to assess it from its "parents": the dynamics $\{\varphi^t\}$, and the set S_0. In this section, we make a first step to overcome this deficit; whereas the function $K(\cdot,\cdot)$ is a by-product of the coalesce of the dynamics and set, we re-master the conditions into a form where separate characteristics of the dynamics and the set are employed.

2.4.1 The Size of Finite Covering

Inspired by e.4), we start with the question: How many balls of a common radius δ are needed to cover a given bounded set? Though not articulated thus far, our interest in fact focuses on the high exactness of estimation $\delta \approx 0$. This, in turn, motivates asymptotical analysis as $\delta \to 0$. A response to these concerns is partly given by the concept of an upper box-counting dimension $\bar{d}_B$, which is defined as follows.

Definition 2.9 ([Falconer, 1997])**.** *The upper box-counting dimension $\bar{d}_B(F)$ of a bounded set $F \subset \mathbb{R}^n$ is given by*

$$\bar{d}_B(F) := \limsup_{\delta \to 0} \frac{\log N_\delta(F)}{-\log \delta}. \tag{2.14}$$

Here, $N_\delta(F)$ can be defined in any of the following ways, with all of them resulting in a common value (2.14)*:*

1. *The smallest number of closed balls of radius δ that cover F;*
2. *The smallest number of closed balls of radius δ and centers in F that cover F;*
3. *The smallest number of cubes of side δ that cover F;*
4. *The number of δ-mesh cubes that intersect F;*
5. *The smallest number of sets of diameter at most δ that cover F;*
6. *The largest number of disjoint balls of radius δ with centers in F.*

Also, the quantity (2.14) *does not depend on the choice of the norm in (i),(ii), (v), and (vi).*

It follows that for arbitrarily small $\varkappa > 0$, the number of δ-balls with centers in F that are needed to cover F does not exceed $\delta^{-(\bar{d}_B(F)+\varkappa)}$ for all sufficiently small $\delta > 0$.

As is well-known [Falconer, 1997], $\bar{d}_B(F) = \bar{d}_B(\overline{F}) \leq n$ for any bounded set $F \subset \mathbb{R}^n$ and $\bar{d}_B(F) = n$ if the interior of F is not empty, $F_1 \subset F_2 \Rightarrow \bar{d}_B(F_1) \leq \bar{d}_B(F_2)$, and $\bar{d}_B(F_1 \cup \ldots \cup F_k) = \max\{\bar{d}_B(F_1), \ldots, \bar{d}_B(F_k)\}$, $k \in \mathbb{Z}^+$. The box-counting dimension may assume non-integer values; for example, $\bar{d}_B(F) = 1/\log_2 3$ for the middle-thirds of the Cantor set $F \subset \mathbb{R}$.

Our particular interest is in dynamical systems and their invariant sets S_0 with $\bar{d}_B(S_0) < n$; this case does hold for some chaotic systems and complex attractors S_0.

2.4.2 Balance between the Initial and Forthcoming Estimation Exactness, Respectively

Now, we are going to study relations between the initial exactness δ of the state x estimate $\hat{x}$ and the implied forthcoming exactness ϵ during the time horizon of duration s_+. This study is aimed at building the component e.3) of which the proposed observer is composed, among others. We recall that this component is a function $\delta(\epsilon, s_+)$ for which Equation (2.10) holds.

The *growth rate* of the system $\{\varphi^t\}$ on the set S_0 is defined to be:

$$g(S_0) := \lim_{\delta\to 0} \overline{\lim_{t\to\infty}}\, t^{-1} \log_2 \sup_{\theta\in T:\theta\le t} \sup_{x\in B_\delta(y), y\in S_0} \|A_\theta(x)\|\,, \tag{2.15}$$

where $A_\theta(x) := \frac{\partial\varphi^\theta}{\partial x}(x)$ is the Jacobian matrix of the map $\varphi^\theta(\cdot)$ at point x and $\log_2\infty := \infty$. It is well-defined for all sufficiently small δ, since $x \in S$ in Equation (2.15), thanks to the following.

Lemma 2.10. *There exists $\delta^0 > 0$ such that $B_{\delta^0}(y) \subset S$ for any $y \in \overline{S}_0$.*

The proof of this lemma is trivial and thus omitted from this document.

In Equation (2.15), the limit $\lim_{\delta\to 0}$ exists since the subsequent quantity decays as δ decreases. Since all norms $\|\cdot\|$ in the space of $n \times n$-matrices are equivalent, it is easy to see that $g(S_0)$ does not depend on the choice of the norm.

Among other components, the proposed observer uses a function $\delta(\epsilon, s_+)$ with a special property described in e.3). Now, we show how such a function can be built from $g(S_0)$.

Lemma 2.11. *Let $g(S_0) < \infty$. For any $\hat{g} > g(S_0)$ and any positive definite $n \times n$-matrix P, there exists a function $\delta(\epsilon, s_+)$ with the property e.3) that is given by*

$$\delta(\epsilon, s_+) = \epsilon 2^{-\hat{g}s_+} \tag{2.16}$$

for all sufficiently small $\epsilon > 0$ and sufficiently large s_+.

The proof of this lemma is provided in Appendix 2.A.

2.4.3 Correct Operation of the Observer and a Criterion for Observability

By bringing the pieces together, we arrive at the following.

Proposition 2.12. *Suppose that Assumptions 2.1–2.3 hold and the system has a finite growth rate $g(S_0)$ on the set S_0. Consider a communication channel with capacity c. If*

$$c > g(S_0)\bar{d}_B(S_0), \tag{2.17}$$

the system $\{\varphi^t\}_{t\in T}$ is observable on the set S_0 via this communication channel in the sense of Definition 2.6.

The proof of this proposition is provided in Appendix 2.A. The previous inequality strongly resembles other inequalities in the context of entropy in dynamical systems that link dimensions, Lyapunov exponents, and entropy (see [Kawan, 2011, Young, 1983, Ledrappier and Young, 1985]). Note that the above bound is suboptimal, in the sense that for noiseless communication channels, several results requiring lower capacities have already been design. One can refer to Section VI of [Kawan and Yüksel, 2018] for an overview of said results. The goal of this chapter is to provide an alternative observer which is robust towards losses in the communication channel.

Remark 2.13. *The bounded transmission delay τ_j^{tr} from Equation* (2.9) *and its upper bound from c.5) do not affect the condition* (2.17) *for observability.*

2.5 Constructive Estimates and Analytical Bounds

In this section, we make the next and final step for obtaining tractable conditions for observability. The road to this is via the development of techniques for assessing growth rate and the box-counting dimension. A technique will be employed in both these cases that is similar in spirit to the second Lyapunov method.

2.5.1 Lyapunov-Like Function

The characteristic trait of the classic Lyapunov function $v(\cdot)$ is its decay along the trajectories of the system. In the current context, we are not interested in such a decay. Instead, our interest is focused on the rate at which an infinitesimally small ball is expanded under the transition mapping φ^t. The smallness implies that this mapping is well-approximated by the first two terms of its Taylor series, and so the rate in question is nothing but the expansion rate due to the Jacobian matrix $A_t(x)$ defined in Equation (2.15). The deformation of a ball under a linear mapping A is described by the singular values of A; in particular, the maximal of them is the norm of A and may be used in Equation (2.15). If P is a symmetric positive definite matrix and $\mathbb{R}^n$ is endowed with the P-related norm $\|\cdot\|_P$, these values are the square roots of the solutions of the algebraic equation $\det[A^\intercal PA - P] = 0$ repeated in accordance with their algebraic multiplicities and ordered from large to small. With these in mind, we introduce a function $v(\cdot) : S \to \mathbb{R}$ with special properties whose description uses the t-step increment of this function:

$$\Delta^t v(x) \coloneqq v(\varphi^t(x)) - v(x). \tag{2.18}$$

Assumption 2.14. *There exist $d \in [0, n]$, a bounded function $v : S \to \mathbb{R}$,*

constant $\Lambda \geq 0$*, and symmetric positive definite matrix* $P \in \mathbb{R}^{n\times n}$*, such that*

$$\Delta^t v(x) + \sum_{i=1}^{\lfloor d \rfloor} \log_2 \lambda_i(t,x) + (d - \lfloor d \rfloor) \log_2 \lambda_{\lfloor d \rfloor + 1}(t,x) \leq \Lambda t, \tag{2.19}$$

$\forall x \in S,\ \forall t \in T : 1 \geq t > 0$*, where* $\lambda_1(t,x) \geq \cdots \geq \lambda_n(t,x)$ *are the roots of the algebraic equation*

$$\det\left[A_t(x)^\intercal P A_t(x) - \lambda P\right] = 0 \tag{2.20}$$

repeated in accordance with their algebraic multiplicities and ordered from large to small, and $\log_2 0 \coloneqq -\infty$*.*

In the discrete-time case, only $t = 1$ is concerned in Equation (2.19). In the continuous-time case, Equation (2.19) is imposed only within the finite time horizon of duration 1.

With $P = I_n$, Equation (2.20) reduces to $\det\left[A_t(x)^\intercal A_t(x) - \lambda I_n\right] = 0$ and $\lambda_i(x,t)$ are the squares of the standard singular values of the Jacobian matrix $A_t(x)$. For a generic P, the roots $\lambda_i(x,t)$ can also be reduced to standard singular values. Indeed, let U be the symmetric and positive definite "square root" of the symmetric and positive definite matrix $P = U^2$. The solutions of Equation (2.20) are evidently identical to those of $\det\left[U^{-1}A_t(x)^\intercal U U A_t(x) U^{-1} - \lambda I_n\right] = 0$, and so the $\lambda_i(x,t)$'s are the squares of the ordinary singular values of the matrix $UA_t(x)U^{-1}$. This matrix is similar to $A_t(x)$, and so these two matrices represent a common linear transformation in various bases. Thus, the role of P is, in fact, that of a linear coordinate transformation in pursuit of ease of building Λ and $v(\cdot)$.

Assumption 2.14 will be utilized for assessment of both quantities that we are interested in. Specifically, it will be used with $d = 1$ and arbitrary Λ to upper-estimate the growth rate (2.15) of the system; this estimate is given by Λ. With $\Lambda = 0$ and some $d \in [0, n]$, it will be used to establish an upper bound on the upper box dimension of the invariant set S_0; this bound is given by d.

In the case of a continuous time system, the next proposition provides an alternative to computing the transition maps $\varphi^t, t \in (0,1]$ and checking infinitely many inequalities (2.19), each for its own $t \in (0,1]$, when verifying Assumption 2.14. To state this proposition, we introduce the Jacobian matrix of the right-hand side in Equation (2.2):

$$J(x) = \frac{\partial f}{\partial x}(x).$$

Proposition 2.15. *Let there exist* $d \in [0,n]$*, a continuously differentiable bounded function* $w : S \to \mathbb{R}$*, constant* $\Gamma \geq 0$*, and a symmetric positive definite matrix* $P \in \mathbb{R}^{n\times n}$*, such that*

$$\dot{w}(x) + \sum_{i=1}^{\lfloor d \rfloor} \gamma_i(x) + (d - \lfloor d \rfloor)\gamma_{\lfloor d \rfloor + 1}(x) \leq \Gamma, \qquad \forall x \in S, \tag{2.21}$$

where $\dot{w}(x) = \frac{\partial w}{\partial x} f(x)$ *and* $\gamma_i(x)$ *are the solutions of the algebraic equation*

$$\det\left(J(x)^\intercal P + PJ(x) - \gamma P\right) = 0 \tag{2.22}$$

ordered from largest to smallest $(\gamma_1(x) \geq \cdots \geq \gamma_n(x))$ *and repeated in accordance with their algebraic multiplicity. Then, Assumption 2.14 holds with the particular* P *and* d *of Equation* (2.22)*,* $v(x) = \frac{w(x)}{\ln 2}$ *and* $\Lambda = \frac{\Gamma}{\ln 2}$*.*

This result is proved in Appendix 2.B.

2.5.2 Analytical Upper Bound on the System's Growth Rate and Related Conditions for Observability

Proposition 2.16. *Let Assumptions 2.1–2.14 hold with* $d = 1$ *and* $\Lambda \geq 0$ *in the last of them. Then, the growth rate* (2.15) *of the system on* S_0 *obeys the following bound:*

$$g(S_0) \leq \frac{\Lambda}{2}.$$

The proof of this proposition is provided in Appendix 2.B.

By combining Propositions 2.12 and 2.16, we arrive at the following.

Theorem 2.17. *Suppose that Assumptions 2.1–2.14 hold with* $d = 1$ *and* $\Lambda \geq 0$ *in the last of them, and consider a communication channel with capacity* c*. If*

$$c > \frac{\Lambda \bar{d}_B(S_0)}{2}, \tag{2.23}$$

the system $\{\varphi^t\}_{t\in T}$ *is observable on the set* S_0 *via this communication channel in the sense of Definition* 2.6*.*

The observation schemes proposed in [Pogromsky and Matveev, 2011, Matveev and Pogromsky, 2016] can sometimes work under the channel rates smaller than that given in Theorem 2.17. This improved rate comes at a price: these schemes are not robust against losses in the communication channel.

In [Liberzon and Hespanha, 2005], the observer requires some feedback in the channel and a channel rate of the form $n \log_2 L$, where L is the Lipschitz constant of the mapping φ^1. The estimate (2.23) is less conservative, both because $L \geq \Lambda/2$ (if L is related to a norm of the form $\|\cdot\|_P$) and $n \geq \bar{d}_B(S)$, with $\geq \mapsto >$ in some cases. Moreover, the scheme from [Liberzon and Hespanha, 2005] does not enjoy robustness against losses in the communication channel.

Finally, Corollary 6.2.1 of [Leonov, 2007] provides an estimate for the topological entropy by using a result from [Ito, 1970], which is identical to our estimate of the rate c with identical assumptions.

2.5.3 Analytical Bounds on the Upper Box Dimension and Final Conditions for Observability

A drawback of Theorem 2.17 is that it uses the upper box dimension, whereas there are no general techniques to compute this dimension analytically. To compensate for this drawback, we will use results from [Hunt, 1996, Kuznetsov, 2016] to replace the upper box dimension by its upper estimate in the form of another well-known kind of dimension, i.e., the so-called Lyapunov dimension. The benefit from this is that the latter can be estimated analytically.

We start by introducing the necessary definitions, including those of the Lyapunov dimension of a map in a point, of a map over a set, and of a dynamical system. Next, we will recall the required results from [Hunt, 1996, Kuznetsov, 2016], and finally, we will provide the general results of this chapter, which offers analytical conditions for observability under a finite communication bit-rate.

Definition 2.18. *For any $t \in T$, the* singular value function *of $A_t(x)$ of order $d \in [0, n]$ at point $x \in \mathbb{R}^n$ is defined as*

$$\omega_d\left(A_t(x)\right) \coloneqq \begin{cases} 1, & d = 0, \\ \sigma_1(A_t(x)) \dots \sigma_d(A_t(x)), & d \in \{1, .., n\}, \\ \sigma_1(A_t(x)) \dots \sigma_{\lfloor d \rfloor + 1}(A_t(x))^{d - \lfloor d \rfloor}, & d \in (0, n) \setminus \{1, .., n\text{-}1\}. \end{cases}$$

Here $\sigma_1(A) \geq \ldots \geq \sigma_n(A)$ are the singular values of the $n \times n$-matrix A.

Definition 2.19 ([Kuznetsov, 2016])**.** *For any $t \in T$, the Lyapunov dimension of the map $\varphi^t(\cdot)$ at the point $x \in S$ is given by*

$$d_L(\varphi^t, x) \coloneqq \sup\{d \in [0, n] : \omega_d\left(A_t(x)\right) \geq 1\}.$$

Definition 2.20 ([Kuznetsov, 2016])**.** *For any $t \in T$, the Lyapunov dimension of the map $\varphi^t(\cdot)$ with respect to the invariant set S_0 is given by*

$$d_L(\varphi^t, S_0) \coloneqq \sup_{x \in S_0} d_L(\varphi^t, x) = \sup_{x \in S_0} \sup\{d \in [0, n] : \omega_d\left(A_t(x)\right) \geq 1\}.$$

Definition 2.21 ([Kuznetsov, 2016])**.** *The Lyapunov dimension of the dynamical system $\{\varphi^t\}_{t \geq 0}$ with respect to the invariant set S_0, is defined as*

$$d_L(\{\varphi^t\}_{t \in T}, S_0) \coloneqq \inf_{t \in T} \sup_{x \in S_0} d_L(\varphi^t, x) = \inf_{t \in T} \sup_{x \in S_0} \sup\{d \in [0, n] : \omega_d\left(A_t(x)\right) \geq 1\}.$$

For the sake of completeness, we provide the results that we borrowed from [Hunt, 1996, Kuznetsov, 2016].

Theorem 2.22 ([Hunt, 1996])**.** *Let Assumptions 2.1–2.3 hold. Then,*

$$\bar{d}_B(S_0) \leq d_L(\varphi^1, S_0).$$

Corollary 2.23 ([Hunt, 1996])**.** *Let the hypotheses of Theorem 2.22 be true. Then for all $t \in T : t \geq 1$,*

$$\bar{d}_B(S_0) \leq d_L(\varphi^t, S_0).$$

The following proposition is essentially a reformulation of Theorem 2 from [Kuznetsov, 2016].

Proposition 2.24. *Let Assumptions 2.1–2.3 be true. Suppose also that Assumption 2.14 holds with some $d \in [0, n]$ and $\Lambda = 0$. Then, for sufficiently large $l > 0$, the following inequality is valid:*

$$d_L(\{\varphi^t\}_{t\in T}, S_0) \leq d_L(\varphi^l, S_0) \leq d. \tag{2.24}$$

The proof of this proposition is provided in Appendix 2.B.

In some cases, the inequalities in Equation (2.24) take place as equalities. Specifically, the following proposition is valid, which is a reformulation of Proposition 3 and Corollary 3 from [Kuznetsov, 2016].

Proposition 2.25 ([Kuznetsov, 2016])**.** *Suppose that at one of the equilibrium points of the dynamical system $\{\varphi^t\}_{t\in T} : x_{\text{eq}} \equiv \varphi^t(x_{\text{eq}})$, the matrix $A_1(x_{\text{eq}})$ has the simple real eigenvalues $\lambda_1(x_{\text{eq}}), \ldots, \lambda_n(x_{\text{eq}})$. Let us consider a non-singular matrix U, such that*

$$UA(x_{\text{eq}})U^{-1} = \operatorname{diag}(\lambda_1(x_{\text{eq}}), \ldots, \lambda_n(x_{\text{eq}})) \tag{2.25}$$

where $|\lambda_1(x_{\text{eq}})| \geq \cdots \geq |\lambda_n(x_{\text{eq}})|$, which matrix does exist thanks to the first assumption of the proposition. Let $\varphi_U : w \to U\varphi^1(U^{-1}w)$ be the transition mapping after the linear coordinate change. Suppose that Assumption 2.14 holds with some d and $\Lambda = 0$, and additionally, we have

$$d_L(\varphi_U, Ux_{\text{eq}}) = d.$$

Then, for any compact invariant set $S_0 \ni x_{\text{eq}}$ of $\{\varphi^t\}_{t\in T}$, the following equation holds

$$d_L(\{\varphi^t\}_{t\in T}, S_0) = d.$$

Now we are in a position to state the main result of the chapter, which is clear from Theorem 2.17 and Proposition 2.24.

Theorem 2.26. *Let Assumptions 2.1–2.3 be true. Suppose also that Assumption 2.14 holds twice: first, with $d = 1$ and some $\Lambda = \bar{\Lambda} \geq 0$ and second, with $\Lambda = 0$ and some $d = \bar{d} \in [0, n]$. Consider a communication channel with capacity c. If*

$$c > \frac{\bar{\Lambda}\bar{d}}{2}, \tag{2.26}$$

the system $\{\varphi^t\}_{t\in T}$ is observable on the set S_0 via this communication channel in the sense of Definition 2.6.

2.6 Examples

In this section, we apply the previous theory to two celebrated prototypical chaotic systems: the smoothened Lozi map and the Lorenz system. For the smoothened Lozi map, we will compute the Lyapunov dimension and provide a bound on the channel rate above where the associated dynamical system is observable via the channel at hand. We will then test this bound via computer simulations of the proposed observer to show that the established theoretical rates are close to the actual practical rates. For the Lorenz system, we borrow upper estimates of the Lyapunov dimension and the largest singular value of the Jacobian from [Pogromsky and Matveev, 2011, Leonov et al., 2016], respectively, to provide a bound on the channel rate by using Theorem 2.26. Like in the previous example, we will also test this bound via computer simulations.

2.6.1 The Smoothened Lozi Map

The Lozi map [Lozi, 1978, Elhadj, 2013] is a modification of the Henon map. The Lozi map is not continuously differentiable, and so does not meet Assumption 2.3. We examine its continuously differentiable analog introduced in [Aziz-Alaoui et al., 2001] by smoothing the Lozi map at the fracture point. The respective smoothened map acts according to the following formula

$$\varphi_\alpha : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} 1 - a f_\alpha(x_1) + b x_2 \\ x_1 \end{pmatrix}, \tag{2.27}$$

where a, b, and $\alpha \ll 1$ are positive parameters and

$$f_\alpha(x) = \begin{cases} |x|, & \text{if } |x| \geq \alpha; \\ \frac{x^2}{2\alpha} + \frac{\alpha}{2}, & \text{if } |x| < \alpha. \end{cases} \tag{2.28}$$

If $1 + a - b > 0$ and $\alpha < (a + 1 - b)^{-1}$, the smoothened Lozi map has an equilibrium

$$x_+ = \left(\frac{1}{1 + a - b}, \frac{1}{1 + a - b} \right).$$

If $1 - a - b < 0$ and $\alpha < (a - b - 1)^{-1}$ in addition to the previous inequalities, there exists a second equilibrium

$$x_- = \left(\frac{1}{1 - a - b}, \frac{1}{1 - a - b} \right).$$

In this section, we adopt the following.

Assumption 2.27. *The following inequalities hold:*

$$\begin{aligned} &a, b, \alpha > 0, \\ &1 - a < b < 1, \\ &\alpha < (a + 1 - b)^{-1}. \end{aligned}$$

This assumption implies that two equilibria exist, and that they are unstable. Moreover, $b < 1$ ensures $d_L(\{\varphi_\alpha^t\}_{t\geq 0}, K) < 2$ for the associated discrete-time dynamical system. We start by giving more insight into the Lyapunov dimension of the smoothened Lozi map.

Theorem 2.28. *Let Assumption 2.27 hold. Then, for any compact invariant set S_0 of the map* (2.27)*, the following inequality is valid*

$$d_L(\{\varphi_\alpha^t\}_{t\in T}, S_0) \leq \overline{d} := 2 - \frac{\log_2 b}{\log_2\left(\sqrt{a^2 + 4b} - a\right) - 1} \tag{2.29}$$

and Assumption 2.14 holds with $\Lambda = 0$ and $d = \overline{d}$. Moreover, if $x_+ \in S_0$, inequality (2.29) *holds as equality:*

$$d_L(\{\varphi_\alpha^t\}_{t\in T}, S_0) = 2 - \frac{\log_2 b}{\log_2\left(\sqrt{a^2 + 4b} - a\right) - 1}.$$

The proof of this theorem is provided in Appendix 2.C.

In order to use the observer from Section 2.3 for the smoothened Lozi map (2.27), we need to choose a compact and invariant set S_0. For the original (i.e., non-smooth) Lozi map, such a set exists whenever the following conditions are met: [Misiurewicz, 1980]

$$0 < b < 1, \tag{2.30}$$
$$a > 0, \tag{2.31}$$
$$2a + b < 4, \tag{2.32}$$
$$b < \frac{a^2 - 1}{2a + 1}, \tag{2.33}$$
$$a\sqrt{2} > b + 2. \tag{2.34}$$

Moreover, when the previous inequalities hold, the set S_0 is the closure of the unstable manifold of the unstable equilibrium x_+. It is still unknown whether they guarantee the same for the smoothened Lozi map (2.27). To the best of the authors' knowledge, no conditions that guarantee the existence of such a set for the map (2.27) are available in the literature. In the following, we will assume that Equations (2.30)–(2.34) are sufficient to ensure the existence of a compact and invariant set S_0. Our simulations with the parameters $a = 1.7$

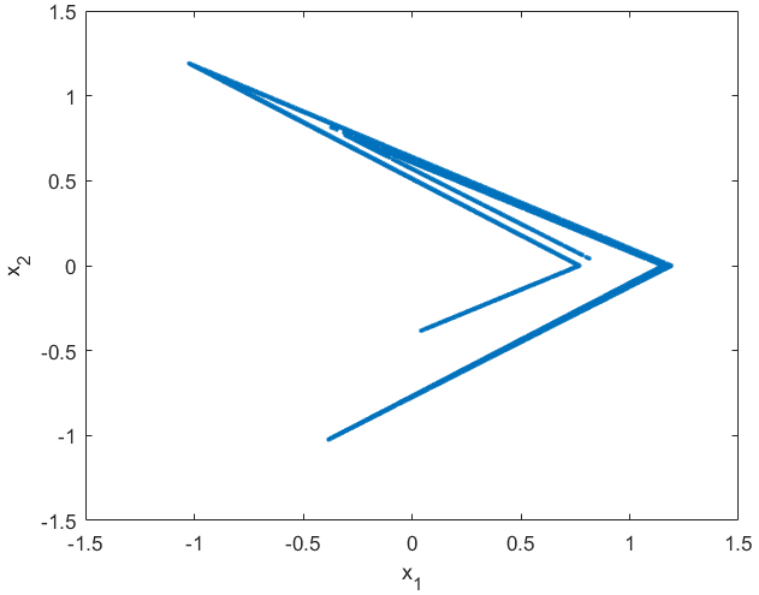

Figure 2.2: Typical trajectory with 10,000 steps of the smoothened Lozi map with $a = 1.7$, $b = 0.3$, and $b = 10^{-5}$.

and $b = 0.3$, $\alpha = 10^{-5}$, which verify Equations (2.30)–(2.34), provide evidence, illustrated in Figure 2.2, in favor of this hypothesis. By combining Theorem 2.26 with Theorem 2.28 and an estimate of the largest singular value of the concerned Jacobian given in [Matveev and Pogromsky, 2016], we arrive at the following.

Corollary 2.29. *Let Assumption 2.27 hold, and let S_0 be a compact invariant set of the smoothened Lozi map* (2.27)*. Then, the associated dynamical system is observable on the set S_0 via any communication channel whose capacity*

$$c > \frac{\left[2\log_2\left(\frac{\sqrt{a^2+4b}-a}{2}\right) - \log_2 b\right]\left[\log_2\left(\frac{\sqrt{a^2+4b}+a}{2}\right)\right]}{\log_2\left(\frac{\sqrt{a^2+4b}-a}{2}\right)}.$$

Proof. Theorem 13 from [Pogromsky and Matveev, 2016a] yields that Assumption 2.14 holds with $d = 1$ and

$$\Lambda = 2\log_2\left(\sqrt{a^2+4b}+a\right) - 2.$$

Theorems 2.26 and 2.28 complete the proof. □

Corollary 2.29 implies that for these parameters, the associated dynamical system is observable for any channel rate above 1.2013 bits/s. To verify whether this lower bound on the channel rate can be improved, we employed the observer from Section 2.3 whose parameters were experimentally tuned to ensure a prespecified exactness of observation ϵ during the first 1000 steps. The following values were considered: $\epsilon = 0.5, 0.2, 0.1, 0.05$. An accompanying objective of experimentally tuning was to minimize the size of the alphabet K employed for data encoding, or, in other words, the channel capacity c^* requested by the observer. The best values of the capacity can be found in Table 2.1. It shows that for high exactness, the system can be observed with a channel rate slightly below the theoretical estimate. For the lowest exactness, the experimental rate barely exceeds the theoretical bound. However, in any case, this bound seems to be pretty close to the experimental result.

Table 2.1: Results of the simulations on the smoothened Lozi map.

	$\epsilon = 0.2$	$\epsilon = 0.1$	$\epsilon = 0.075$	$\epsilon = 0.05$
K (1)	1×10^6	2×10^6	2×10^7	3.5×10^7
c^* (bits/s)	1.0924	1.1499	1.169431	1.212770

2.6.2 The Lorenz System

In this section, we apply our previous theoretical results to the Lorenz system. The Lorenz system [Lorenz, 1963] is a well-known example of a continuous-time system where, for certain values of its parameters, it displays chaotic behavior. The system equations are:

$$\begin{cases} \dot{x}_1 &= -\sigma x_1 + \sigma x_2, \\ \dot{x}_2 &= \rho x_1 - x_1 x_3 - x_2, \\ \dot{x}_3 &= x_1 x_2 - \beta x_3, \end{cases} \tag{2.35}$$

where σ, ρ, and β are positive parameters. If $\rho < 1$, the system has a single globally asymptotically stable equilibrium: the origin. For $\rho > 1$, this equilibrium becomes a hyperbolically unstable saddle-point. In addition, two equilibria appear. In this chapter, we assume $\rho > 1$. We will apply our findings to the system with its chaotic attractor as the set S_0. As is well-known [Sparrow, 1982], the conditions on the parameters ($\sigma > 0,\ \beta > 0,\ \rho > 1$) suffice to ensure the presence of a compact invariant set. Moreover, to compute the Lyapunov dimension, we adopt the following assumption which is taken from [Leonov et al., 2016].

Assumption 2.30. *Let the following hold:*

$$\begin{aligned}\rho &> 1,\\ \rho &\geq \frac{\beta^3 - 2\beta^2 + 6\beta^2\sigma - 3\beta\sigma^2 - 6\beta\sigma + \beta}{3\sigma^2} + 1,\\ \sigma\rho &> (\beta+1)(\beta+\sigma),\end{aligned}$$

and either

$$\sigma^2(\rho-1)(\beta-4) \leq 4\sigma(\sigma\beta+\beta-\beta^2) - \beta(\beta+\sigma-1)^2,$$

or the following equation has two distinct solutions, ν

$$\begin{aligned}&(2\sigma-\beta+\nu)^2(\beta(\beta+\sigma-1)^2 - 4\sigma(\sigma\beta+\beta-\beta^2) + \sigma^2(\rho-1)(\beta-4))\\ &\quad + 4\beta\nu(\sigma+1)(\beta(\beta+\sigma-1)^2 - 4\sigma(\sigma\beta+\beta-\beta^2) - 3\sigma^2(\rho-1)) = 0 \quad (2.36)\end{aligned}$$

and

$$\begin{cases}\sigma^2(\rho-1)(\beta-4) > 4\sigma(\sigma\beta+\beta-\beta^2) - \beta(\beta+\sigma-1)^2\\ \nu_1 > 0\end{cases},$$

where ν_1 is the largest root of Equation (2.36).

Any solution of the Lorenz system that starts at $t = 0$ can be extended on $[0, \infty)$ [Leonov, 2007], and thus has the extendability property discussed just after Equation (2.2). Hence, the differential Equations (2.35) give rise to a dynamical system on $S := \mathbb{R}^3$ in the sense of Section 2.2.1.

Proposition 2.31. *Let Assumption 2.30 hold, and let S_0 be a compact invariant set of the Lorenz system. Then, this system is observable on the set S_0 via any communication channel whose capacity is:*

$$c > \frac{\left[\sqrt{(\sigma\text{-}1)^2 + 4\rho\sigma}\text{-}\sigma\text{-}1\right]\left[3\sqrt{(\sigma\text{-}1)^2 + 4\rho\sigma}\text{-}2\beta + \sigma + 1\right]}{2\ln 2\left[\sqrt{(\sigma\text{-}1)^2 + 4\rho\sigma} + \sigma + 1\right]}.$$

Proof. Since the right-hand side of the equations in Equation (2.35) are polynomial, Assumptions 2.1 and 2.3 hold by Proposition 2.4; Assumption 2.2 is true due to the choice of S_0. It is easy to see by inspection of the proof of Theorem 4.3 from [Pogromsky and Matveev, 2011] that the assumptions of Proposition 2.15 hold with $d = 1$,

$$\Gamma = \sqrt{(\sigma\text{-}1)^2 + 4\rho\sigma} - \sigma - 1,$$

and the matrix P defined by (16) in [Pogromsky and Matveev, 2011]. So Proposition 2.15 guarantees that Assumption 2.14 holds with $d = 1$ and

$\Lambda = \frac{1}{2\ln 2}\left[\sqrt{(\sigma\text{-}1)^2 + 4\rho\sigma} - \sigma - 1\right]$. As is shown in Section 4 from [Leonov et al., 2016], Assumption 2.14 also holds with $\Lambda = 0$ and

$$d = \overline{d} = 3 - \frac{2(\sigma + \beta + 1)}{\sqrt{(\sigma - 1)^2 + 4\sigma\rho} + \sigma + 1}.$$

Theorem 2.26 completes the proof. □

We have performed simulation studies similar to those carried out for the previous example. Starting from various initial conditions in S_0, we have simulated the observer for various chosen ϵ, each with its own chosen δ and associated covering. In our simulations, we used $\sigma = 10$, $\rho = 28$, $\beta = \frac{8}{3}$, which verify Assumption 2.30. For these parameters, the theoretical rate bound is $c > 40.975$ bit/s. The results of the simulations can be seen in Table 2.2. Once again, it can be seen that the experimentally found rate is below or very close to the theoretical rate. This confirms that our theoretical results correctly predict the rate.

Table 2.2: Results of the simulations on the Lorenz system.

	$\epsilon = 0.2$	$\epsilon = 0.1$	$\epsilon = 0.075$	$\epsilon = 0.05$
K (1)	364758	714701	1448880	3.5×10^7
c^* (bits/s)	19.814	30.739	34.614	43.714

2.7 Conclusion

In this chapter, we have presented an observer for both discrete and continuous-time nonlinear systems. We have provided bounds on the necessary data-rates to implement the observer. We have proven that this observer can be implemented on any channel with a finite delay parameter and a channel rate $c > \Lambda \bar{d}_B(S_0)/2$, where $\Lambda/2$ is an upper bound on the largest singular value of the Jacobian and $\bar{d}_B(S_0)$, the upper box dimension of the compact invariant set of the system. By combining results from several other papers, we have provided an analytical bound to the channel rate that depends on the Lyapunov dimension, rather than the upper box dimension. These analytical bounds have been computed for the smoothened Lozi map and the Lorenz system. For the smoothened Lozi map, we computed the Lyapunov dimension. Simulations of the observer on both of these systems have proven that the theoretical rate is closely related to the actual rate required to implement the observer.

The following notations are used in this manuscript:

- $\mathbb{Z}^+$: the set of nonnegative integers;
- $\mathbb{R}^+$: the set of nonnegative real numbers;

- $|x|$: the absolute value of x;
- $\mathrm{card}(F)$: the cardinality of a set F;
- $\lfloor s \rfloor = \max_{t\in\mathbb{Z}} t,\ \text{s.t. } t \leq s$;
- I_n: the identity matrix of dimension $n \times n$;
- $M^\intercal$: the conjugate transpose of the matrix M;
- $M^{-\intercal}$: the inverse of the transpose of the matrix M;
- $\|x\|_2 = \sqrt{x^\intercal x}$;
- $\|x\|_P = \sqrt{x^\intercal P x}$, with P symmetric and positive definite;
- $\mathrm{B}_\delta(x)$: the ball in $\|\cdot\|_2$ of radius δ centered in x;
- $\mathrm{B}^P_\delta(x)$: the ball in $\|\cdot\|_P$ of radius δ centered in x.

Appendices

2.A Proofs of Section 2.4

In this appendix, we present the proofs of of the results stated in Section 2.4.

2.A.1 Proof of Lemma 2.11

We introduce the P-associated spectral norm of a $n \times n$-matrix:

$$\|A\|_P = \max_{x:\|x\|_P=1} \|Ax\|_P.$$

Based on Equation (2.15), Lemma 2.10, and the remark following that lemma, we first pick $\delta^0 > 0$ such that

$$B_{\delta^0}(x) \subset S \ \forall x \in \overline{S}_0$$

and

$$\overline{\lim_{t\to\infty}}\, t^{-1} \log_2 \sup_{\theta\in T:\theta\leq t} \sup_{x\in B_{\delta^0}(y), y\in S_0} \left\| \frac{\partial \varphi^\theta}{\partial x}(x) \right\|_P < \hat{g}.$$

Based on this inequality, we then pick $\varsigma \in T$ so that for all $t \in T, t \geq \varsigma$, we have

$$\left\| \frac{\partial \varphi^\theta}{\partial x}(x) \right\|_P < 2^{\hat{g}t} \quad \text{if } \theta \in T, \theta \leq t,\ x \in B_{\delta^0}(y), y \in S_0. \tag{2.37}$$

Now, we consider $\epsilon \leq \delta^0$, $s_+ \geq \varsigma$, and $\delta = \delta(\epsilon, s_+)$ defined in Equation (2.16). Let the context of Equation (2.10) holds, i.e., $\hat{x}, x \in S_0$ and $\|\hat{x} - x\|_P \leq$

$\delta(\epsilon, s_+) = \epsilon 2^{-\hat{g}s_+} \in (0, \delta^0]$. Then, the segment $[x, \hat{x}] = \{(1-\theta)x+\theta\hat{x} : 0 \le \theta \le 1\}$ lies in the balls $B_{\delta^0}(x)$, and so in S by the choice of δ^0. By using the mean value inequality Theorem 9.19 from [Rudin, 1976], we have for any $\theta \in T, \theta \le s_+$,

$$. \left\|\varphi^\theta(\hat{x}) - \varphi^\theta(x)\right\|_P \le \sup_{z\in[x,\hat{x}]} \left\| \frac{\partial \varphi^\theta}{\partial x}(z)[\hat{x}-x] \right\|_P \le \|x-\hat{x}\|_P \sup_{z\in[x,\hat{x}]} \left\| \frac{\partial \varphi^\theta}{\partial x}(z) \right\|_P .$$

Now, we note that the conditions from Equation (2.37) are fulfilled for $t := s_+, x := z$, and $y := \hat{x}$. Hence

$$\left\|\varphi^\theta(\hat{x}) - \varphi^\theta(x)\right\|_P \le \epsilon 2^{-\hat{g}s_+} \times 2^{\hat{g}s_+} = \epsilon.$$

Thus, we see that Equation (2.10) is true. It remains to extend the function $\delta(\epsilon, s_+)$ from $\epsilon \in (0, \delta^0], s_+ \in T, s_+ \ge \varsigma$ on all $\epsilon > 0$ and $s_+ \in T$ by putting $\delta(\epsilon, s_+) := \delta\big(\min\{\epsilon, \delta^0\}, \max\{s_+, \varsigma\}\big)$. □

2.A.2 Proof of Proposition 2.12

We first pick $\hat{g} > g(S_0)$ and $\varkappa > 0$ so close to $g(S_0)$ and 0, respectively, that

$$c > \hat{g}[\bar{d}_B(S_0) + \varkappa]. \tag{2.38}$$

We also pick a positive definite $n \times n$-matrix P and borrow the function $\delta(\epsilon, s_+)$ from Lemma 2.11. We also consider $\epsilon > 0$ and $s_+ \in T$ so small and large, respectively, that Equation (2.16) holds. As was remarked just after Definition 2.9, the set S_0 can be covered by no more than $\delta^{-(\bar{d}_B(S_0)+\varkappa)}$ δ-balls centered in S_0 for all small enough $\delta > 0$. By reducing $\epsilon > 0$, if necessary, we make $\delta = \delta(\epsilon, s_+)$ small enough in this sense irrespective of $s_+ \in T$. Then, no more than

$$\delta(\epsilon, s_+)^{-(\bar{d}_B(S_0)+\varkappa)} = \epsilon^{-(\bar{d}_B(S_o)+\varkappa)} 2^{\hat{g}[\bar{d}_B(S_0)+\varkappa]s_+}$$

δ-balls centered in S_0 are needed to cover S_0. Hence, the integer floor of the right-hand side of this equation is the function $K(\epsilon, s_+)$ from e.4). So, the condition (2.13) for the correct operation of the observer from Section 2.3 takes the form

$$\begin{gathered} -(\bar{d}_B(S_0) + \varkappa)\log_2 \epsilon + \hat{g}[\bar{d}_B(S_0) + \varkappa](\tau^{\mathrm{pr}} + \tau_+^{\mathrm{tr}}) \le b(\tau^{\mathrm{pr}}) \\ \Leftrightarrow \frac{\hat{g}[\bar{d}_B(S_0) + \varkappa]\tau_+^{\mathrm{tr}} - (\bar{d}_B(S_0) + \varkappa)\log_2 \epsilon}{\tau^{\mathrm{pr}}} + \hat{g}[\bar{d}_B(S_0) + \varkappa] \le \frac{b(\tau^{\mathrm{pr}})}{\tau^{\mathrm{pr}}}. \end{gathered}$$

By invoking Equations (2.8) and (2.38), we see that the last inequality can be satisfied by picking τ^{pr} which is large enough. Then, by picking $s_+ \ge \tau^{\mathrm{pr}} + \tau_+^{\mathrm{tr}}$ in accordance with Equation (2.12), we ensure the correct operation of the observer. The statement ii) from Observation 1 completes the proof. □

2.B Proofs of Section 2.5

2.B.1 Proof of Proposition 2.15

For the dynamical system given by Equation (2.2), the matrices $A_t(x)$ defined in Equation (2.15) obey the equations

$$\frac{\partial A_t}{\partial t}(x) = J[\varphi^t(x)]A_t(x), \qquad A_0(x) = I_n. \tag{2.39}$$

Hence,

$$A_\Delta(x) = I_n + \Delta J(x) + \Delta M_x(\Delta), \quad \forall \Delta > 0, \tag{2.40}$$

where $M_x(\Delta) \to 0$ as $\Delta \to 0$. Moreover, this convergence is uniform over x from any compact subset K of S. Indeed, for $x \in K$, Equation (2.39) yields that

$$\begin{aligned} M_x(\Delta) &= \Delta^{-1}[A_\Delta(x) - A_0(x)] - J(x) = \Delta^{-1}\int_0^\Delta \frac{\partial A_t}{\partial t}(x)\, dt - J(x) \\ &= \Delta^{-1}\int_0^\Delta \Big[J[\varphi^t(x)]A_t(x) - J(x)\Big]\, dt = \Delta^{-1}\int_0^\Delta \Big[\Psi(t,x)] - \Psi(0,x)\Big]\, dt, \end{aligned}$$

where $\Psi(t,x) := J[\varphi^t(x)]A_t(x)$. Due to the last claim of Proposition 2.4, the function $\Psi(t,x)$ is continuous. Hence it is uniformly continuous on the compact set $[0,1] \times K$ and so

$$\mu_K(\Delta) := \max_{t\in[0,\Delta], x\in K} \|\Psi(t,x) - \Psi(0,x)\| \to 0 \quad \text{as} \quad \Delta \to 0.$$

It remains to note that for all $x \in K$,

$$\begin{aligned} \|M_x(\Delta)\| &\le \Delta^{-1}\left\|\int_0^\Delta \Big[\Psi(t,x) - \Psi(0,x)\Big]\, dt\right\| \le \Delta^{-1}\int_0^\Delta \Big\|\Psi(t,x) - \Psi(0,x)\Big\|\, dt \\ &\le \mu_K(\Delta). \end{aligned}$$

By invoking Equation (2.40), we see that

$$A_\Delta(x)^\intercal P A_\Delta(x) = P + \Delta\left[J(x)^\intercal P + PJ(x)\right] + \Delta Q_x(\Delta), \tag{2.41}$$

where

$$\begin{aligned} Q_x(\Delta) := {} & \Delta J(x)^\intercal P J(x) + \Delta M_x(\Delta)^\intercal P M_x(\Delta) + P M_x(\Delta) + M_x(\Delta)^\intercal P \\ & + \Delta J(x)^\intercal P M_x(\Delta) + \Delta M_x(\Delta)^\intercal P J(x) \to 0 \quad \text{as } \Delta \to 0 \end{aligned}$$

uniformly over $x \in K$ due to the foregoing since the continuous function $J(x)$ is bounded on the compact set K. Let $U = U^\intercal > 0$ be the positive definite square

root $P = U^2$ of P. Equation (2.20) with $t := \Delta$ can be rewritten by virtue of Equation (2.41) as follows

$$\begin{gathered}
0 = \det[A_\Delta(x)^\intercal P A_\Delta(x) - \lambda P] \\
\Updownarrow \\
0 = \det\Big\{\Delta\left[J(x)^\intercal P + P J(x)\right] + \Delta Q_x(\Delta) - (\bar\lambda - 1)P\Big\} \\
\Updownarrow \\
\det\Big\{\left[J(x)^\intercal P + P J(x)\right] + Q_x(\Delta) - \tfrac{\bar\lambda - 1}{\Delta} P\Big\} = 0 \\
\Updownarrow \\
\det\Big\{\left[U^{-1}J(x)^\intercal U + U J(x) U^{-1}\right] + U^{-1} Q_x(\Delta) U^{-1} - \tfrac{\bar\lambda - 1}{\Delta} I_n\Big\} = 0.
\end{gathered}$$

Thus we see that $[\lambda_i(\Delta, x] - 1)/\Delta$ are the ordinary eigenvalues of the symmetric matrix $U^{-1}J(x)^\intercal U + UJ(x)U^{-1} + U^{-1}Q_x(\Delta)U^{-1}$, which goes to $U^{-1}J(x)^\intercal U + UJ(x)U^{-1}$ as $\Delta \to 0$ uniformly over $x \in K$. Meanwhile the eigenvalues continuously depend of the symmetric matrix by Corollary 6.3.8 [Horn and Johnson, 2013]. It follows that

$$[\lambda_i(\Delta, x) - 1]/\Delta = \eta_i(x) + \omega_i(x, \Delta),$$

where $\omega_i(x, \Delta) \to 0$ as $\Delta \to 0$ uniformly over $x \in K$ and $\eta_i(x)$ are the solutions for the eigenvalue problem

$$\det\Big[U^{-1}J(x)^\intercal U + UJ(x)U^{-1} - \eta I_n\Big] = 0 \Leftrightarrow \det\Big[J(x)^\intercal P + PJ(x) - \eta P\Big] = 0.$$

The last equation is identical to Equation (2.22), and so $\eta_i(x) = \gamma_i(x)$. By using the previous equations together, we see that

$$\lambda_i(\Delta, x) = 1 + \Delta\gamma_i(x) + \Delta\omega_i(x, \Delta), \quad \Omega_i(K, \Delta) := \sup_{x \in K} |\omega_i(x, \Delta)| \to 0 \text{ as } \Delta \to 0. \tag{2.42}$$

Now we invoke d from Proposition 2.15. For any $n \times n$ matrix B, we denote by $\bar\lambda_i(B)$ the roots of the algebraic equation

$$\det[B^\intercal P B - \lambda P] = 0 \tag{2.43}$$

enumerated from large to small, and put

$$\bar\omega_d(B) := \bar\lambda_1(B) \cdots \bar\lambda_{\lfloor d \rfloor}(B)[\bar\lambda_{\lfloor d \rfloor + 1}(B)]^{d - \lfloor d \rfloor}$$

By combining the generalized Horn inequality [Leonov, 2007] with Lemma 8.1 in [Pogromsky and Matveev, 2011] (which relates to the roots of Equation (2.43) with the concept of the singular value of a matrix), we infer that $\bar\omega_d(BC) \le \bar\omega_d(B)\bar\omega_d(C)$ for any $n \times n$ matrices B, C. It follows that for any sequence $B_0, \ldots, B_m$ of such matrices

$$\bar\omega_d(B_m B_{m-1} \cdots B_0) \le \prod_{j=0}^{m} \bar\omega_d(B_j).$$

Now we pick an arbitrary $t > 0$ and denote $\Delta_r := t/r, t_r(j) := j\Delta_r$ for any natural r and $j \in \mathbb{Z}^+$. Since the system is time-invariant, we have

$$A_t(x) = A_{\Delta_r}[\varphi^{t_r(r-1)}(x)]A_{\Delta_r}[\varphi^{t_r(r-2)}(x)]\cdots A_{\Delta_r}[\varphi^{t_r(1)}(x)]A_{\Delta_r}[\varphi^{t_r(0)}(x)].$$

As a result, we see that in Equation (2.19),

$$\begin{aligned}
\mathfrak{A} &:= \sum_{i=1}^{\lfloor d\rfloor} \log_2 \lambda_i(t,x) + (d - \lfloor d\rfloor)\log_2 \lambda_{\lfloor d\rfloor+1}(t,x) = \log_2 \bar{\omega}_d[A_t(x)] \\
&\le \sum_{j=0}^{r-1} \log_2 \bar{\omega}_d\big\{A_{\Delta_r}[\varphi^{t_r(j)}(x)]\big\} \\
&= \sum_{j=0}^{r-1} \left\{ \sum_{i=1}^{\lfloor d\rfloor} \log_2 \lambda_i[\Delta_r, \varphi^{t_r(j)}(x)] + (d - \lfloor d\rfloor)\log_2 \lambda_{\lfloor d\rfloor+1}[\Delta_r, \varphi^{t_r(j)}(x)] \right\} \\
&= \sum_{i=1}^{\lfloor d\rfloor}\sum_{j=0}^{r-1} \log_2 \lambda_i[\Delta_r, \varphi^{t_r(j)}(x)] + (d - \lfloor d\rfloor)\sum_{j=0}^{r-1} \log_2 \lambda_{\lfloor d\rfloor+1}[\Delta_r, \varphi^{t_r(j)}(x)] \\
&= \sum_{i=1}^{\lfloor d\rfloor}\sum_{j=0}^{r-1} \log_2 \big\{1 + \Delta_r\gamma_i[\varphi^{t_r(j)}(x)] + \Delta_r\omega_i[\varphi^{t_r(j)}(x), \Delta_r]\big\} \\
&+ (d - \lfloor d\rfloor)\sum_{j=0}^{r-1} \log_2 \big\{1 + \Delta_r\gamma_{\lfloor d\rfloor+1}[\varphi^{t_r(j)}(x)] + \Delta_r\omega_{\lfloor d\rfloor+1}[\varphi^{t_r(j)}(x), \Delta_r]\big\}.
\end{aligned}$$

We proceed by using the elementary inequality $\log_2(1+x) \le x/\ln 2$ and the quantity $\Omega_i(K,\Delta)$ defined in Equation (2.42) for the compact set $K := \{\varphi^\theta(x) : 0 \le \theta \le t\}$, which contains all points of the form $\varphi^{t_r(j)}(x), j = 0, \ldots, r$,

$$\begin{aligned}
&\sum_{j=0}^{r-1} \log_2 \big\{1 + \Delta_r\gamma_i[\varphi^{t_r(j)}(x)] + \Delta_r\omega_i[\varphi^{t_r(j)}(x), \Delta_r]\big\} \\
&\le \frac{\Delta_r}{\ln 2}\sum_{j=0}^{r-1} \gamma_i[\varphi^{t_r(j)}(x)] + \frac{\Delta_r}{\ln 2}\sum_{j=0}^{r-1} \omega_i[\varphi^{t_r(j)}(x), \Delta_r] \\
&\le \frac{\Delta_r}{\ln 2}\sum_{j=0}^{r-1} \gamma_i[\varphi^{t_r(j)}(x)] + \frac{r\Delta_r}{\ln 2}\Omega_i(K, \Delta_r),
\end{aligned}$$

where $r\Delta_r = t$ by the definition of $\Delta_r = t/r$. Thus

$$\mathfrak{A} \le \frac{\Delta_r}{\ln 2} \sum_{j=0}^{r-1} \underbrace{\sum_{i=1}^{\lfloor d \rfloor} \gamma_i[\varphi^{t_r(j)}(x)] + (d - \lfloor d \rfloor)\gamma_{\lfloor d \rfloor + 1}[\varphi^{t_r(j)}(x)]}_{\mathfrak{B}} + \Omega(\Delta_r), \quad \text{where}$$

$$\Omega(\Delta) := \frac{t}{\ln 2} \left[\sum_{i=1}^{\lfloor d \rfloor} \Omega_i(K, \Delta) + (d - \lfloor d \rfloor)\Omega_{\lfloor d \rfloor + 1}(K, \Delta) \right] \to 0 \quad \text{as} \quad \Delta \to 0.$$

Now we estimate $\mathfrak{B}$ by employing Equation (2.21)

$$\begin{aligned} \mathfrak{A} &\le \frac{\Delta_r}{\ln 2} \sum_{j=0}^{r-1} \left\{ \Gamma - \dot{w}([\varphi^{t_r(j)}(x)] \right\} + \Omega(\Delta_r) \\ &= t\frac{\Gamma}{\ln 2} - \frac{1}{\ln 2} \sum_{j=0}^{r-1} \Delta_r \dot{w}([\varphi^{t_r(j)}(x)] + \Omega(\Delta_r). \end{aligned}$$

Here, the sum is the Riemann sum of the continuous function $\dot{w}([\varphi^{\theta}(x)]$ of $\theta \in [0, t]$, and $\mathfrak{A}$ does not depend on r. So be letting $r \to \infty$ and by invoking that $\dot{w}(x) = \frac{\partial w}{\partial x}(x) f(x)$, we get

$$\begin{aligned} &\sum_{i=1}^{\lfloor d \rfloor} \log_2 \lambda_i(t, x) + (d - \lfloor d \rfloor) \log_2 \lambda_{\lfloor d \rfloor + 1}(t, x) \\ &\quad \le t\frac{\Gamma}{\ln 2} - \frac{1}{\ln 2} \int_0^t \frac{\partial w}{\partial x}[\varphi^{\theta}(x)] f[\varphi^{\theta}(x)] \, d\theta \\ &\quad = t\frac{\Gamma}{\ln 2} - \frac{1}{\ln 2} \left\{ w[\varphi^{\theta}(x)] - w[\varphi^{0}(x)] \right\} \end{aligned}$$

Thus we have arrived at Equation (2.19) modulo the definitions of $\Lambda = \Gamma / \ln 2$ and $v(x) = w(x)/\ln 2$ from Proposition 2.15. It remains to note that the function $v(x)$ is bounded since $w(x)$ has this property by the assumptions of Proposition 2.15. □

2.B.2 Proof of Proposition 2.16

We first note that Equation (2.19) with $d = 1$ means that whenever $x \in S, t \in T, 1 \ge t > 0$, we have

$$\Delta^t v(x) + 2 \log_2 \|A_t(x)\|_P \le \Lambda t. \tag{2.44}$$

We are going to show that for any natural r, Equation (2.44) holds whenever $t \in T, 0 < t \le r$, arguing by induction on r. As a result, we will show that Equation (2.44) is valid for all $t \in T, t > 0$.

For $r = 1$, this claim is initially given. Suppose that it is true for some r, and consider $t \in T \cap (0, r+1]$. If $t \in T \cap (0, r]$, Equation (2.44) is true by the induction hypothesis. Let $t > r$. Then $t = r + \theta$, where $\theta \in T, 0 < \theta \leq 1$. By putting $y := \varphi^r(x)$ and invoking Assumption 2.1 and Equation (2.15), we see that

$$\begin{aligned}\varphi^t = \varphi^\theta \circ \varphi^r \Rightarrow A_t(x) = A_\theta(y)A_r(x) \Rightarrow \|A_t(x)\|_P \leq \|A_\theta(y)\|_P \|A_r(x)\|_P \\ \Rightarrow \log_2 \|A_t(x)\|_P \leq \log_2 \|A_\theta(y)\|_P + \log_2 \|A_r(x)\|_P; \\ \Delta^t v(x) = v(\varphi^t(x)) - v(x) = v(\varphi^{r+\theta}(x)) - v(\varphi^r(x)) + v(\varphi^r(x)) - v(x) \\ = \Delta^\theta v(y) + \Delta^r v(x),\end{aligned}$$

where the start of the second line is due to Equation (2.18). By using these relations, we arrive at Equation (2.44) via adding Equation (2.44) with $t := \theta$ to the inequality

$$\Delta^r v(x) + 2\log_2 \|A_r(x)\|_P \leq \Lambda r,$$

which holds by the induction hypothesis. Thus, Equation (2.44) is true whenever $t \in T, t > 0$.

Now we introduce a finite upper bound $\overline{v} \geq |v(x)|\ \forall x \in S$ on the bounded function $|v(\cdot)|$. Let $x \in S$. Then $\varphi^t(x) \in S$ and so Equation (2.18) yields that $|\Delta^t v(x)| \leq 2\overline{v}$. By Equation (2.44),

$$\log_2 \|A_t(x)\|_P \leq \Lambda t/2 + \overline{v}.$$

As was remarked, $g(S_0)$ does not depend on the matrix norm $\|\cdot\|$ in Equation (2.15). Meanwhile, Lemma 2.10 ensures that $x \in B_\delta(y), y \in S_0 \Rightarrow x \in S$ for all small enough $\delta > 0$. Hence

$$\begin{aligned}g(S_0) &= \lim_{\delta \to 0} \overline{\lim_{t \to \infty}}\, t^{-1} \sup_{\theta \in T: \theta \leq t} \sup_{x \in B_\delta(y), y \in S_0} \log_2 \|A_\theta(x)\|_P \\ &\leq \lim_{\delta \to 0} \overline{\lim_{t \to \infty}}\, t^{-1} \sup_{\theta \in T: \theta \leq t} \sup_{x \in B_\delta(y), y \in S_0} [\Lambda\theta/2 + \overline{v}] = \Lambda/2.\end{aligned}$$ □

2.B.3 Proof of Proposition 2.24

Proof. From Assumption 2.14, we have that $\lambda_i(t, x)$ are the roots of the equation

$$\det(A_t(x)^\intercal P A_t(x) - \lambda P) = 0. \tag{2.45}$$

Since P is positive definite and symmetric, we can decompose it as

$$P = U^\intercal U,$$

where U is nonsingular. Equation (2.45) can thus be rewritten as

$$\det(A_t(x)^\intercal U^\intercal U A_t(x) - \lambda U^\intercal U) = 0.$$

If we premultiply with $U^{-\intercal}$ and postmultiply with U^{-1}, the solutions λ_i are unchanged and the equation becomes

$$\det(U^{-\intercal}A_t(x)^\intercal U^\intercal U A_t(x)U^{-1} - \lambda I_n) = 0.$$

We thus know that $\lambda_i(t,x)$ are the eigenvalues of the matrix

$$U^{-\intercal}A_t(x)^\intercal U^\intercal U A_t(x)U^{-1}$$

or the square of the singular values of the matrix $UA_t(x)U^{-1}$. From Equation (2.14) with the same d and $\Lambda = 0$, we have that

$$\Delta^t v(x) + \sum_{i=1}^{\lfloor d \rfloor} \log_2 \lambda_i(t,x) + (d - \lfloor d \rfloor) \log_2 \lambda_{d+1}(t,x) < 0,$$

$\forall x \in S,\ t \in T : 0 < t \leq 1$, which thus also implies that

$$\frac{1}{\log_2 e}\left(\frac{1}{2}\Delta^t v(x) + \sum_{i=1}^{\lfloor d \rfloor} \frac{1}{2}\log_2 \lambda_i(t,x) + \frac{(d - \lfloor d \rfloor)}{2}\log_2 \lambda_{d+1}(t,x)\right) < 0,$$

$\forall x \in S,\ t \in T : 0 < t \leq 1$, where e is Euler's number. This can be rewritten as

$$\frac{1}{2\log_2 e}\Delta^t u(x) + \sum_{i=1}^{\lfloor d \rfloor} \ln\sqrt{\lambda_i(t,x)} + (d - \lfloor d \rfloor)\ln\sqrt{\lambda_{d+1}(t,x)} < 0,$$

$\forall x \in S,\ t \in T : 0 < t \leq 1$.

We now apply Theorem 2 from [Kuznetsov, 2016] with $t = 1$, $S = U$, $\lambda_i(x,S) = \sqrt{\lambda_i(1,x)}$ and $V(x) = \frac{1}{2\log_2 e}v(x)$ to obtain

$$d_L(\{\varphi^t\}_{t\geq 0}, S_0) \leq d_L(\varphi^T, S_0) \leq d,$$

for T sufficiently large. □

2.C Proofs of Section 2.6

Proof of Theorem 2.28

Proof. Proposition 2.24 yields that to prove the first sentence from the conclusion of Theorem 2.28, it suffices to justify Assumptions 2.1–2.3 and Assumption 2.14 with $\Lambda = 0$ and $d := \bar{d}$ defined in Equation (2.29) for the discrete-time dynamical system associated with the smoothened Lozi map. Assumptions 2.1–2.3 do hold since the map (2.27) is smooth and S_0 is a compact invariant set by the assumptions of Theorem 2.28. It remains to check Assumption 2.14, where $t = 1$ in Equation (2.19) since we are in discrete time now.

We are going to justify Assumption 2.14 with $P = \mathrm{diag}\{1, b\}$. Since we have that

$$A(x) = \begin{pmatrix} -af'_\alpha(x) & b \\ 1 & 0 \end{pmatrix},$$

Equation (2.20) becomes

$$\det\left(A(x)^\intercal P A(x) - \lambda P\right) = 0 \tag{2.46}$$

$$\Updownarrow \tag{2.47}$$

$$\det\left[\begin{pmatrix} a^2\left(f'_\alpha(x)\right)^2 + b & -af'_\alpha(x)b \\ -af'_\alpha(x)b & b^2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & b\lambda \end{pmatrix}\right] = 0. \tag{2.48}$$

To simplify the notations, we introduce $\bar{f} \coloneqq f'_\alpha(x)$. Equation (2.48) admits two solutions

$$\begin{aligned} \lambda_1(x) &= \frac{1}{4}\left(\sqrt{a^2\bar{f}^2 + 4b} + a|\bar{f}|\right)^2, \\ \lambda_2(x) &= \frac{1}{4}\left(\sqrt{a^2\bar{f}^2 + 4b} - a|\bar{f}|\right)^2. \end{aligned} \tag{2.49}$$

Since by Equation (2.28),

$$1 \geq |\bar{f}| \geq 0, \tag{2.50}$$

given Assumption 2.27, we know that $\max_{x\in S_0} \lambda_1(x)$ is always larger than one (in particular, for all $|x| \geq \alpha$, $\bar{f} = 1$). This implies that $d_L(\{\varphi^t_\alpha\}_{t\in T}, S_0) > 1$ and that $1 < d < 2$. Note that the latter inequality is strict as a result of Assumption 2.27. After some computations and due to Equation 2.49, the left-hand side of Equation (2.19) becomes

$$\begin{aligned} &\log_2(\lambda_1(x)) + (d - \lfloor d \rfloor)\log(\lambda_2(x)) \\ &= 2\left[1 - (d - \lfloor d \rfloor) + \log_2 b + (d - \lfloor d \rfloor - 1)\log_2\left(\sqrt{a^2\bar{f}^2 + 4b} - a|\bar{f}|\right)\right]. \end{aligned}$$

Which, using Equation (2.50) can be upper bounded in the following way

$$\begin{aligned} &\log_2(\lambda_1(x)) + (d\text{-}\lfloor d \rfloor)\log(\lambda_2(x)) \\ &\leq 2\left[1\text{-}(d\text{-}\lfloor d \rfloor) + \log_2 b + (d\text{-}\lfloor d \rfloor\text{-}1)\log_2\left(\sqrt{a^2 + 4b}\text{-}a\right)\right]. \end{aligned}$$

To satisfy Assumption 2.14 with some d and $\Lambda = 0$, we are looking for $d - \lfloor d \rfloor$ such that the left-hand side of the previous equation is negative. We thus obtain the following sufficient condition

$$2\left(1 - (d - \lfloor d \rfloor) + \log_2 b + (d - \lfloor d \rfloor - 1)\log_2\left(\sqrt{a^2 + 4b} - a\right)\right) < 0$$

which, using Assumption 2.27, can be rewritten as

$$(d - \lfloor d \rfloor) > 1 - \frac{\log_2 b}{\log_2\left(\sqrt{a^2+4b} - a\right) - 1}.$$

Note that from the conditions from Assumption 2.27, the following is always true $0 < (d - \lfloor d \rfloor) < 1$. Since Assumption 2.14 with $d = \bar{d}$ and $\Lambda = 0$ is verified, the proof is completed by applying Proposition 2.24 which yields

$$d_L(\{\varphi_\alpha^t\}_{t\geq 0}, S_0) \leq \bar{d} = 2 - \frac{\log_2 b}{\log_2\left(\sqrt{a^2+4b} - a\right) - 1}.$$

To prove the second part of this theorem, we will use Proposition 2.25. We consider the following diagonalizing coordinate change matrix

$$U = \begin{bmatrix} \frac{1}{\sqrt{a^2+4b}} & \frac{a+\sqrt{a^2+4b}}{2\sqrt{a^2+4b}} \\ \frac{-1}{\sqrt{a^2+4b}} & \frac{-a+\sqrt{a^2+4b}}{2\sqrt{a^2+4b}} \end{bmatrix}$$

which is nonsingular and well-defined for all parameters satisfying Assumption 2.27.

To compute $d_L(\varphi_U, Ux_{\text{eq}})$, we first compute

$$\lambda_1(x_+) = \sqrt{\frac{a^2+2b+a\sqrt{a^2+4b}}{2}}, \tag{2.51}$$

$$\lambda_2(x_+) = \sqrt{\frac{a^2+2b-a\sqrt{a^2+4b}}{2}}. \tag{2.52}$$

From Assumption 2.14 with $d = \bar{d}$ and $\Lambda = 0$, we have $\lambda_1(x_+) > 1$ and $\lambda_2(x_+) < 1$ which means we have $\bar{d} \in [1, 2)$. We thus compute

$$\begin{aligned} d_L(\varphi_U, Ux_+) = \sup\{d \in [0,n] : \omega_d(UA(Ux_+)U^{-1}) > 1\} \\ \Leftrightarrow \lambda_1(x_+)\left[\lambda_2(x_+)\right]^{(\bar{d}-\lfloor \bar{d} \rfloor)} = 1. \end{aligned}$$

Using Equations (2.51) and (2.52), and after some computations, we obtain that

$$\bar{d} - \lfloor \bar{d} \rfloor = \frac{-\log_2\left(\sqrt{\frac{a^2+2b+a\sqrt{a^2+4b}}{2}}\right)}{\log_2\left(\sqrt{\frac{a^2+2b-a\sqrt{a^2+4b}}{2}}\right)}.$$

This latter equation can be rewritten as

$$\bar{d} - \lfloor \bar{d} \rfloor = 1 - \frac{\log_2 b}{\log_2\left(\sqrt{a^2+4b} - a\right) - 1}.$$

Since we assumed $x_+ \in S_0$, all conditions of Proposition 2.24 are verified. We thus obtain

$$d_L(\{\varphi_\alpha^t\}_{t\geq 0}, S_0) = 2 - \frac{\log_2 b}{\log_2\left(\sqrt{a^2+4b}-a\right)-1}.$$

□

Chapter 3

Consensus in Networks of Dynamical Systems with Limited Communication Capacity

This chapter provides a solution for the preservation of consensus in a network of several agents, described by discrete-time nonlinear dynamical systems. Consensus preservation is the problem of maintaining a certain distance between the states of several systems, given that the systems' initial states are close to each other. The agents are equipped with both a smart sensor, capable of measuring the state of the system and performing some computations, and a controller. The sensors and controllers are placed at locations that are remote from one another. A network of communication channels with limited transmission capacity connects the agents by allowing the sensors to send messages to the controllers of their system as well as to the controllers of other systems. The controllers use the messages that they receive to steer the agents such that the consensus is preserved. Sensors and controllers that achieve this feature are called consensus-preserving protocols. In this chapter, three distinct consensus-preserving protocols are presented, each with an increasing degree of interaction between the systems. For each of these protocols, a theorem providing conditions on the minimum sufficient communication capacity to implement them is presented. The protocols are tested by simulations for a network of logistic maps and a network of Hénon maps. For both of these networks, analytical bounds on the sufficient transmission capacities in the communication channels to implement the protocols are provided. These bounds are compared to rates observed in simulations of the consensus-preserving protocols.

3.1 Introduction

In modern society, wireless technologies such as Wi-Fi and Bluetooth are omnipresent and it is thus no surprise that these technologies have many applications in the industrial world. For the field of dynamics and control, these technologies have given birth to a completely new subfield that studies interactions between dynamical systems and wireless communication technologies. These communication technologies generally take the form of communication channels that can only transmit limited numbers of bits per unit of time and can be subject to losses or corruption of the messages they carry. In many modern industrial applications, one or several dynamical systems, together with their sensors and actuators, are connected only by such communication channels. Examples of such applications are networks of distributed sensors that communicate via Wi-Fi, drones which fly in formation, the cooperative driving of automated vehicles.

In this chapter, we will investigate lossless communication channels with limited transmission capacities. For dynamical systems, interactions with such constrained channels become problematic when they are combined with one or several sources of uncertainties. Indeed, according to Shannon (see Shannon [1948]), uncertainty can be seen as information and in the case of dynamical systems, this uncertainty generally takes one or several of the four following forms: uncertainty in and sensitivity to initial conditions, parametric uncertainties, unmodelled system dynamics, or noise/disturbances. The challenge is to design communication protocols to transmit the information generated by the uncertainties via the data rate constrained channels to remote locations such that all the information needed to solve the posed control/estimation problem is transmitted whilst respecting the data rate constraints.

The earliest works on the interaction between dynamical systems and communication channels with limited transmission capacities were concerned with the design of observers and controllers that function over data rate constrained channels. The earliest works focus on linear dynamical systems (see e.g. Wong and Brockett [1997] or Elia and Mitter [2001] and references therein). Most of the problems involving linear systems have now been solved (an overview of those results can be found in Nair et al. [2007], Baillieul and Antsaklis [2007], Andrievsky et al. [2010], Yüksel and Başar [2013], and Fang et al. [2017]).

Although some results appeared not much later for nonlinear systems, most of these early results presupposed very specific structures in the nonlinear systems (see e.g. in De Persis [2003] and Baillieul [2004]). More general results were obtained in Nair et al. [2004] and Liberzon and Hespanha [2005]. The first paper used a concept called feedback entropy to provide meaningful bounds on the necessary channel rates while the second paper used techniques that were first designed for linear dynamical systems and extended them to the nonlinear case. Since then, several different notions of entropy have been introduced in

an effort to provide general results about necessary and/or sufficient data rates for observers and controllers of nonlinear systems with data rate constrained communication channels (see Kawan [2013], Colonius et al. [2013], Matveev and Savkin [2009], Kawan [2018], Sibai and Mitra [2017], Liberzon and Mitra [2016], and Pogromsky and Matveev [2016a]).

Since one of the important topics in the field of dynamics and control is the problem of synchronization and/or consensus of dynamical systems, it should come as no surprise that this problem was also studied in the subfield of dynamical systems and data rate constrained communication channels. Two of the first works that studied the problem of consensus of systems that communicate with limited data rates are Fradkov et al. [2008a] and Fradkov et al. [2008b] where the problem of master/slave synchronization of two nonlinear systems was considered. Since then, several more papers have focused on this problem. Some results were obtained for linear systems, such as in Li et al. [2011], where the problem of average consensus in networks of linear systems with fixed topologies and limited data rates was tackled. The paper Li and Xie [2011] provided an extension of the aforementioned results with a time-varying network topology. In You and Xie [2011] the specific effects of network topology and data rate constraints were studied and in Li et al. [2016] an event-triggered approach was developed for the average consensus of linear systems with data rate constraints. In Meng and Li [2014] and Meng et al. [2017], the authors develop an observer-based solution for the coordination problem with quantization. Some works on consensus for nonlinear systems were also done such as Dong [2019], which studied consensus for networks of nonlinear systems with data rate constraints.

In this work, we consider the problem of consensus preservation in a network of identical nonlinear dynamical systems. Several agents, with identical dynamics, have initial states that are close to each other. All agents are equipped with a smart sensor (a smart sensor is a sensor that is also capable of performing some computations) and a controller. The sensor and controller are placed at locations remote from one another. The sensors are connected to the controllers, including that of their own agent, via a network of communication channels with limited transmission capacities. A particular controller has access to data about the state of its own associated agent only if a communication channel links this controller with the sensor of this agent. The data rate constrained channels can only transmit symbols from finite-sized lists of symbols that are generally referred to as alphabets. The goal of this chapter is to design the sensors, alphabets, and decoder-controllers such that the initial distance between the states of the agents is preserved over time, up to a time-invariant multiplication factor. A combination of sensors, alphabets, and controllers which achieves this feature is called a consensus-preserving protocol. In this chapter, we present three consensus-preserving protocols. This work extends Voortman et al. [2020d], both because two new consensus-preserving protocols are added and because the considered class of systems is broader. For all protocols, the design of the

alphabets is based on an idea that is very similar to the idea of separating the state-space into partitions and using symbolic dynamics (see Morse and Hedlund [1938]) to describe the dynamical system. The main result of this work resides in three theorems, one for each consensus-preserving protocol, which provide analytical upper bounds on transmission capacities in the communication channels sufficient to implement the consensus-preserving protocol. The bounds on the capacities are proven to depend on the singular values of the Jacobian of the system. The novelty of this work is to consider the situation where the sensors and controllers of each agent are connected only via the limited capacity communication channels, which, to the best of the authors' knowledge, has not been considered in the literature before.

The structure of the chapter is as follows. Firstly, the problem statement along with all necessary definitions are exposed in Section 3.2. Next, in Section 3.3, the method which will be used to communicate whilst limiting the data rate is described. In Section 3.4, three different algorithms, that solve the problem exposed in the problem statement, are developed. The chapter continues with Section 3.5, which contains the main results of the chapter: theorems that provide upper bounds on the minimum sufficient capacities required to implement the consensus-preserving protocols of the previous section. Finally, in Section 3.6, examples of how these theorems can be used are given. The problems of consensus preservation for a network of logistic maps and a network of Hénon maps are considered. For both of these problems, the sufficient capacities are computed from the theorems of the previous section and compared with the rates of simulations.

3.2 Problem Statement

We consider a network of k agents described by the following discrete-time dynamical systems:

$$x_i(t+1) = f(x_i(t), u_i(t)), \qquad x_i(0) = x_{i0}, \tag{3.1}$$

for $i \in \{1, \dots, k\}$, where $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ is a nonlinear mapping, $u_i \in \mathbb{R}^{n_u}$ are the control inputs, $x_{i0} \in X$ are initial states, and X is a given set of them. We define the distance $\mathrm{dist}(x, X)$ from a point x to a set X as $\mathrm{dist}(x, X) \coloneqq \inf_{y \in X} \|x - y\|_P$, where $\|x\|_P \coloneqq \sqrt{x^\intercal P x}$, for a positive definite matrix P that will be defined further in this chapter. We impose the following regularity assumptions on the mapping f and the state-space trajectories of the system.

Assumption 3.1. *The function f is continuously differentiable on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_u}$. The set X of initial states is bounded and there exists $\xi > 0$ such that for all $x \in \mathbb{R}^{n_x}$ verifying $\mathrm{dist}(x, X) \leq \xi$, $f(x, 0) \in X$.*

Further in this chapter, a covering of X with balls in norm $\|\cdot\|_P$ will be considered. Assumption 3.1 implies that such a covering exists and that, provided that the balls in the covering are of small enough radius and a zero control input is applied, all trajectories starting within the covering end up in X. The agents are connected through a network of communication channels. The connections in the network are described by a communication adjacency matrix $\mathfrak{A} \in \mathbb{R}^{k\times k}$. Each entry $\mathfrak{a}_{ij}$ of this matrix is 1 if there is a communication channel transmitting messages $m_{ij}(t)$ from agent i to agent j and 0 otherwise. We denote by J the set of all pairs (i, j) such that $\mathfrak{a}_{ij} = 1$. Figure 3.1 depicts a particular network of four agents.

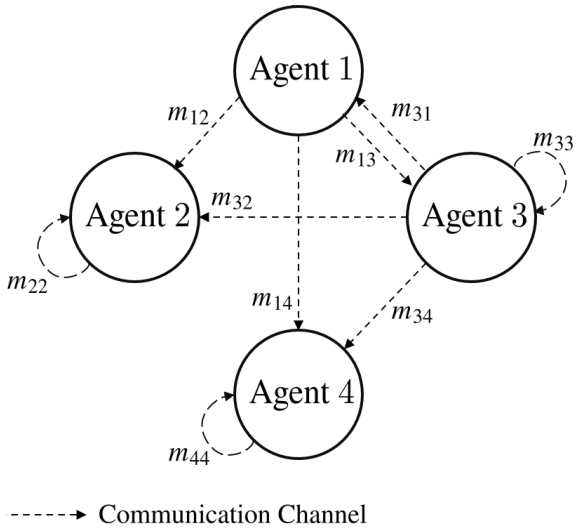

Figure 3.1: The structure of a particular network with 4 agents.

Each agent is equipped with a smart sensor and a smart controller which are responsible for sending (sensor) and receiving (controller) messages over the communication channels. The structure of a single agent is depicted on Figure 3.2.

Informally, the challenge could be described as follows: given that the states of the systems are close to each other initially, design the sensors and controllers of the agents such that the states of their systems remain close to each other whilst using bit-rates below the transmission capacities of the communication channels. Each smart sensor i is composed of several encoders $\mathcal{E}_{ij}$: one for each of the agents it is connected to. Encoder $\mathcal{E}_{ij}$ sends messages $m_{ij}(t)$ to controller j. All messages are sent simultaneously by the encoders at communication time instants $t \in S := \{0, \bar{s}, 2\bar{s}, \ldots\}$, where $\bar{s} > 0$ is a tunable constant. The encoders

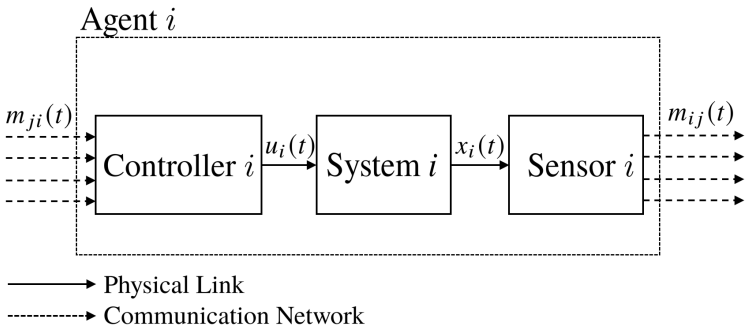

Figure 3.2: The structure of a single agent.

$\mathcal{E}_{ij}$ utilize the past messages $M_{ij}(t)$ that they sent, $M_{ij}(t) := \{m_{ij}(s) : s \in S, s < t\}$, $\forall i, j \in J$. It is assumed that all agents are aware of a shared initial estimate of the state $\hat{x}_0 \in X$ which verifies

$$\|\hat{x}_0 - x_i(0)\|_P \leq \delta, \tag{3.2}$$

$\forall i \in \{1, \ldots, k\}$. It is also assumed that the sensors measure the state of their own system without any measurement error. The encoder equations are

$$m_{ij}(t) = \mathcal{E}_{ij}(f, \bar{s}, \delta, x_i(t), \hat{x}_0, M_{ij}(t)) \tag{3.3}$$

for $i, j \in J,\ t \in S$.

For any time interval $\bar{s}$ between two consecutive communications, each of the channels has its own maximum number of bits $b_{ij}^{+}(\bar{s})$ that can be transmitted and depends on the length of the time interval $\bar{s}$. The bound b_{ij}^{+} implies that the messages $m_{ij}(t)$ that are sent have to be drawn from finite-sized alphabets. Each communication channel between sensor i and controller j has its own alphabet function $\mathcal{A}_{ij}$ which determines the length of the alphabet at time t and hence the variety of messages it can transmit. The alphabet is a list of symbols which are indexed from 1 to $l_{ij}(t) < \infty$ where the last index $l_{ij}(t)$ and thus the size of the alphabet at time t is determined by the alphabet function $\mathcal{A}_{ij}$. The alphabet of each channel is known by both agents which communicate via the channel. The alphabet function equations are

$$l_{ij}(t) = \mathcal{A}_{ij}(f, \bar{s}, \delta, \hat{x}_0, M_{ij}(t)), \tag{3.4}$$

for $i, j \in J,\ t \in S$. Since the messages have to be elements of the alphabet of their channel, the restriction on the messages generated by (3.3) is that

$$m_{ij}(t) \in \{1, ..., l_{ij}(t)\}, \tag{3.5}$$

for $i, j \in J,\ t \in S$. At every communication instant, the logarithm of the size of the alphabet should always be inferior or equal to the maximum number of

bits that can be transmitted during any time interval between communications. The following should thus hold

$$\log_2(l_{ij}(t)) \leq b_{ij}^+(\bar{s}), \tag{3.6}$$

for $i, j \in J,\ t \in S$.

On the controller side, we assume that the messages are received one time instant before the next communication, that is, $\bar{s} - 1$ time instants later, at the end of the transmission interval (the time between two consecutive communications $\bar{s} > 0$ prevents the channel from transmitting infinite amounts of data instantaneously). Each controller uses the previously received messages $\bar{M}_j(t) \coloneqq \{m_{ij}(s) : s \in S, s \leq t - \bar{s} + 1, i \in \{1, \ldots, k\} : \mathfrak{a}_{ij} = 1\}$ to form the control input, and is described by an equation of the form

$$u_j(t) = \mathcal{U}_j(f, \bar{s}, \delta, \hat{x}_0, \bar{M}_j(t)) \tag{3.7}$$

$j \in \{1, \ldots, k\},\ \forall t \geq 0$.

To formally define the objective of this chapter, we need a quantity that measures the sum of the channel's transmission capacities from all channels that each agent is connected to. We first define each channel's transmission capacity $\mathfrak{c}_{ij}$ as

$$\mathfrak{c}_{ij} \coloneqq \liminf_{s \to \infty} \frac{b_{ij}^+(s)}{s}. \tag{3.8}$$

Note that this definition means that for any $\alpha > 0$ and $\mathfrak{c}_{ij} > \alpha$ there exists $\bar{s} < \infty$ such that $\alpha\bar{s}$ bits can be sent during any time interval of duration $\bar{s}$. The **outgoing communication capacity** c_i of agent i is then defined as

$$c_i \coloneqq \min_{j:\mathfrak{a}_{ij}=1} \mathfrak{c}_{ij} \sum_{q=1}^{k} \mathfrak{a}_{iq}. \tag{3.9}$$

The outgoing communication capacity of an agent is thus the product of the minimum transmission capacity among all the channels that the agent's smart sensor is connected to times the number k_c of controllers that the agent is connected to. Hence for any $\alpha > 0$, $\mathfrak{c}_i > \alpha k_c$ implies that there exists a time interval $\bar{s} < \infty$ such that $\alpha\bar{s}$ bits can be sent over each of those communication channels during that time interval.

We pose the following problem: several systems have initial states all within a distance of 2δ of each other (see (3.2)). Design the sensors, alphabet functions, and controllers such that their interactions preserve the consensus, i.e. that the states of the systems remain within a certain distance $G\delta$ of each other for $t > 0$, where G is a positive constant. Moreover, this consensus-preserving property should hold for all small enough δ's with a common G. Note that this property does not necessarily require all systems to be controlled, but at least some of

them should be, particularly the systems that are sensitive to initial conditions, such as chaotic systems.

The constant G will be referred to as a **consensus factor** and is defined as follows.

Definition 3.2. *Let (3.2) hold for k systems with a particular $\delta > 0$. Then, if the quantity $G < \infty$ satisfies $\|x_i(t) - x_j(t)\|_P \leq G\delta$, $\forall t \geq 0$, $\forall i, j \in \{1, \ldots, k\}$, it is called a* ***consensus factor****.*

All smart sensors, alphabet functions, and smart controllers together form a consensus-preserving protocol. The protocol should not only maintain the consensus but also function whilst respecting the channel capacity constraints (as stemming from (3.6), (3.8), and (3.9)). If a particular protocol achieves both these features, it is said to preserve consensus which is more formally defined as follows.

Definition 3.3. *A consensus protocol (3.3), (3.4), (3.7) preserves consensus of k agents with outgoing communication capacities c_i as defined in (3.9) if both of the following conditions hold*

(i) There exists $G < \infty$, $\delta^* > 0$, such that for all $\delta : 0 < \delta \leq \delta^*$, G is a consensus factor as defined in Definition 3.2 with those particular δ;

(ii) The messages generated by the encoders respect the restriction on the choice of messages (3.5) and the alphabet functions respect the channel capacity constraints (3.6).

This chapter presents several consensus-preserving protocols (3.3), (3.4), (3.7).

3.3 Rationale Behind the Alphabet for Communication

A first important feature of a consensus-preserving protocol is how the sensors and controllers communicate. To preserve the consensus between the agents, the controllers need information about the state of the system they control, as well as information about a common trajectory that is going to be followed. Since the communication channels can only transmit limited numbers of bits during any time interval, sending the full state is not possible. There is thus a necessity of quantizing the state so that it can be encoded into messages of finite sizes. These messages can then be used to generate estimates of the states which will be used to control the systems. This discretization involves designing the alphabets which are lists of symbols that are used for communication. To make the protocols more graspable, we develop the design of the alphabets in a separate section where we describe one possible approach to this design. Note that the alphabets only describe a possible method to encode the information about

the state or the common trajectory into messages. Which sensor should send what information to what controller is another part of the consensus-preserving protocol and will be discussed in the next section. In this section, we describe the design of the alphabets and not the alphabet functions that determine the cardinality of the alphabets (as defined in (3.4)).

In this section, the role of the quantities δ and $\bar{s}$ will be highlighted. Up to now, how these should be chosen has not been discussed. The method to build the alphabets which we will now describe simply requires that δ is chosen strictly positive and small enough (at least smaller than ξ from Assumption 3.1) and that $\bar{s}$ is an integer larger than zero.

The idea behind the alphabet (which is partially based on techniques from Matveev and Pogromsky [2016], Pogromsky and Matveev [2011], Voortman et al. [2018a], Voortman et al. [2018b], and Voortman et al. [2019]) is to cover the set of initial conditions X with balls of size δ, where δ is the initial distance (the left part of Fig. 3.3 depicts such a covering). We will use the notation $B_\delta(x)$ for a ball of radius δ, centred in x, in the norm $\|\cdot\|_P$. The rationalities of this section are valid for any positive definite P. In the following definition, we will use the notation $|V|$ to denote the cardinality of a set V.

Definition 3.4. *Given X and $\delta > 0$, we will denote by $V \subset X$ a finite set of points $v_q \in V$ such that $X \subseteq \bigcup_{q=1}^{|V|} B_\delta(v_q)$.*

The set V is a part of our protocol, it is thus known to all agents (and their connected devices). This set is used to form a covering of X with balls: each element in V corresponds to the center of one ball in the covering. This in turn provides a simple method to discretize X: each point of X is mapped to the center of one of the balls in which it lies, i.e. points are discretized to elements of V. Assumption 3.1 states that X is bounded which implies that the cardinality of V can be chosen finite. The set V cannot be used to communicate estimates because it might require a very large number of balls to cover X, and in particular, as δ tends to zero, the number of elements in V increases unboundedly. We use another feature of the covering to overcome this difficulty. In the absence of input, and given $\bar{s} \geq 1$ and $\delta > 0$, one can compute the image $f^{\bar{s}}(B_\delta(v_q), 0)$ of the ball q through the mapping f applied $\bar{s}$ times with zero input. This situation is illustrated in Fig. 3.3.

The images of these balls

$$I_q \coloneqq f^{\bar{s}}(B_\delta(v_q), 0), \tag{3.10}$$

are subsets of X by Assumption 3.1 and can thus be covered by selecting balls with centres in V. To each set I_q we associate a set V_q of such centres. A family of these sets $\{V_q\}$ is defined as follows:

Definition 3.5. *Given X, $\delta > 0$, and V, for each $v_q \in V$, we denote a finite set $V_q \subseteq V$ of points $v_p^q \in V_q$, $p \in \{1, \ldots, |V_q|\}$ such that $I_q \subseteq \bigcup_{p=1}^{|V_q|} B_\delta(v_p^q)$, where I_q is defined in (3.10). The family $\{V_q\}$ is the collection of these sets.*

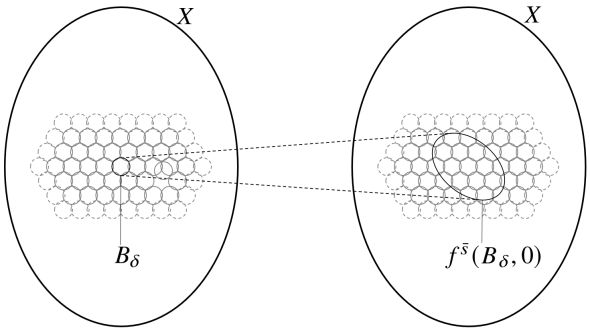

Figure 3.3: Every ball B_δ of the covering of the state-space X is mapped into an ellipsoid after $\bar{s}$ time-steps.

As a part of our protocol, a set V and a family $\{V_q\}$ are computed and provided to all sensors and controllers. For each of the balls with a center v_q in V, the indexes of the balls in V_q form an alphabet. Each symbol in this alphabet corresponds to the center of one of the balls used to cover I_q (See Fig. 3.4).

Remark 3.6. *The sets V_q depend on I_q, which are the images of the balls through the mapping with a zero input applied. The fact that the input is assumed to be zero is intentional and will be taken into account in the design of the consensus-preserving protocols in the next section.*

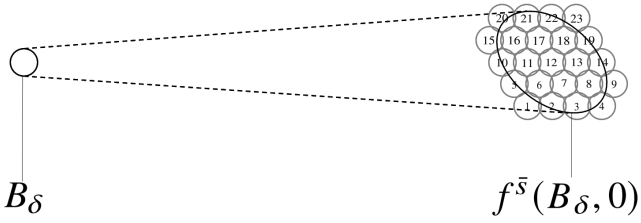

Figure 3.4: Alphabet for the ball B_δ of Fig. 3.3, $l_{ij} = 23$

The sensors then use the sets V and $\{V_q\}$ to send messages in the following way. If a controller possesses an estimate v_q of the current state of its system, then it knows which alphabet is used for communication (namely $\{1, \ldots, |V_q|\}$), since all alphabets are known at the start of the protocol. Because the sensor has access to the full state of its system, it can easily compute the future state of the system with zero input (by applying the mapping $f(\cdot, 0)$ to the current state of the system). To transmit an estimate of the future state to the controller, the encoder which communicates with that controller then simply needs to transmit the index p of the center of the ball v_p^q in which the state will be in $\bar{s}$ time-steps (Fig. 3.4 presents this idea schematically) to provide a future estimate of

the state of its system to that controller. The alphabets thus allow any sensor to send estimates of the state of its system to any of the controllers, provided that they are connected via a communication channel with sufficient capacity to transmit the message.

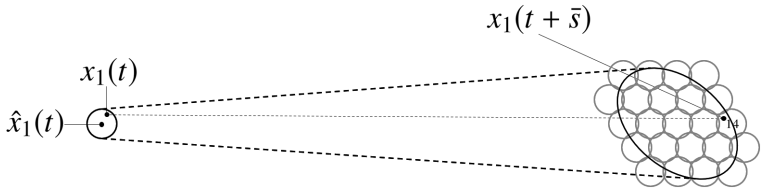

Figure 3.5: Particular situation with the state and estimate. The state evolves in such a way that it ends up in the ball with index 14 after $\bar{s}$ time steps. The message to communicate a new estimate would then be "14".

The presented design of the alphabets of this section will be used to build the alphabet functions of the next section where we will discuss consensus-preserving protocols in full.

3.4 Consensus-preserving Protocols

In this section, the consensus-preserving protocols, in the form of equations of encoders, alphabet functions, and controllers are presented. For the clarity and simplicity of the upcoming mathematical developments, we will only consider mappings with the following form

$$f(x_i(t), u_i(t)) = \varphi(x_i(t)) + u_i(t). \tag{3.11}$$

Note that it is possible to adapt the upcoming consensus-preserving protocols to situations with non-additive inputs. To achieve this, it is necessary to assume reachability/controllability for the systems (see Nijmeijer and van der Schaft [2013]) and to adapt the protocols to take into account the time it would take to drive the systems from one point to another. This heavily impacts the clarity of the presentation, however, and given the current complexity of the mathematical developments, is left for further research.

We will denote $\varphi^s(\cdot)$ the mapping φ applies s times to itself (i.e. $\varphi^s(\cdot) = \varphi(\ldots\varphi(\cdot))$). Three distinct protocols will be presented, each requiring a higher outgoing communication capacity. To avoid redundancy, the first subsection will contain the aspects that are identical in the functioning of all three protocols. We will then dedicate a separate subsection to every particular protocol. For each of them, we will briefly describe the underlying idea and then present the equations.

3.4.1 Common Elements of the Protocols

Some parts of the consensus protocols are similar, in this subsection we describe these common parts. We denote by q_t the index of the ball with center v_{q_t} in which the states $x_i(t)$ are contained at the communication time $t \in S$. We use the notation $\bar{t}$ for the last communication instant. For the sake of brevity, we will not repeat the arguments of the protocol functions and instead use the abbreviated notation: $\mathcal{A}_{ij}(\cdot), \mathcal{E}_{ij}(\cdot)$, and $\mathcal{U}_i(\cdot)$.

We design the controller equations (3.7) such that the input of all systems is zero at all time instants except the ones preceding the instants of communication (i.e. $u_i(t) = 0,\ \forall t : (t+1) \in S$). Note that due to the transmission delay, these are also the instants at which the previous messages reach the controllers. To control only at the time instants preceding communications, all protocols operate in periods of $\bar{s} \geq 2$. Note that this design allows us to consider the mapping with zero input when constructing the subcoverings, as was foreshadowed by Remark 3.6.

By properly controlling the systems, the protocol guarantees that at the communication instants, all states $x_i(t)$ are within the ball of radius δ centred in v_{q_t}. The index q_t of the ball in which the states of all of the systems are at the communication instants is assumed to be always known by all sensors and controllers. The control input is applied at the time step preceding the communication instants. Via the messages they receive, all controllers compute a common ball in which all systems should be at the next communication instant. If a controller receives no messages, it applies zero control input. The control input that is applied takes the form of a shift vector in the state-space which is added to the dynamics of the system (as in (3.11)). It corresponds to the difference between the center of the ball in which all systems should be at the next time step, and the ball in which the controller's system would be, had no input be applied.

The alphabet functions (3.4) are

$$\mathcal{A}_{ij}(\cdot) = |V_{q_t}|, \tag{3.12}$$

$t \in S, i, j \in J$. The alphabet function thus corresponds to the cardinality of the covering of the set I_{q_t} (as defined in (3.10)).

The encoder functions (3.3) are

$$\mathcal{E}_{ij}(\cdot) = \arg \min_{p \in \{1,\ldots,|V_{q_t}|\}} \left\| \varphi^{\bar{s}}(x_i(t)) - v_p^{q_t} \right\|_P, \tag{3.13}$$

$t \in S, i, j \in J$. Each encoder thus sends the index of the ball in which its system will be in $\bar{s}$ timesteps (should no input be applied) to the controller it is connected to.

3.4.2 The Pacemaker Protocol

The pacemaker solution relies on an external pacemaker to determine the ball in which the state of all systems should be contained. This decision happens independently of the current states of the systems. The trajectory generated by the pacemaker is denoted $r(t)$ and is defined as follows

$$r(t) \coloneqq \arg\min_{v \in V} \left\| \varphi^t(\hat{x}_0) - v \right\|_P .$$

It is a sequence of points, which are selected from the points in the set V, such that they are the closest points to the solution to (3.1) with $\hat{x}_0$ as an initial condition. This trajectory is known by all controllers since they have access to both $\hat{x}_0$ and V. In order to track this trajectory, the controllers require an estimate of the state of their own system, which implies that the communication adjacency matrix is $\mathfrak{A} = I_k$ (where I_k is the $k \times k$ identity matrix). The controller equations (3.7) of the pacemaker protocol are for,

$$\mathcal{U}_i(\cdot) = \begin{cases} 0, & \text{if } (t+1) \notin S \\ r(t+1) - v_{m_{ii}(\bar{t})}^{q_{\bar{t}}}, & \text{if } (t+1) \in S. \end{cases}, \qquad t \geq 0.$$

Remark 3.7. *It might seem peculiar at first sight that tracking a predetermined target is regarded as consensus. Given the previously enunciated definition for consensus preservation, this protocol is however inevitable. Not only does this protocol preserve consensus, but it is also the protocol that will be presented that requires the smallest outgoing communication capacity, as will be proven in the next section.*

3.4.3 The Master/Slaves Protocol

The master/slaves protocol relies on one of the systems being the master system and the remaining $k-1$ systems being the slave systems. The master system is left uncontrolled and all the slaves track the master system's trajectory. The slave systems are uncoupled. For simplicity's sake, we will assume that system 1 is the master while the next $k-1$ systems are the slaves. The communication adjacency matrix is

$$\mathfrak{A} = \mathfrak{A}_{\text{ms}} \coloneqq \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{3.14}$$

Note that the element $\mathfrak{a}_{11}$ of this matrix is zero, which is because the first system will not be controlled and hence doesn't need an estimate of its state.

The controller equations (3.7) of the master/slaves protocol are for $t \geq 0$,

$$\begin{aligned} \mathcal{U}_1(\cdot) &= 0, \\ \mathcal{U}_i(\cdot) &= \begin{cases} 0, & \text{if } (t+1) \notin S, \\ v^{q_{\bar{t}}}_{m_{1i}(\bar{t})} - v^{q_{\bar{t}}}_{m_{ii}(\bar{t})}, & \text{if } (t+1) \in S, \end{cases} , \qquad \forall i \neq 1, \end{aligned}$$

3.4.4 The Mutual Protocol

Our third consensus-preserving protocol is closer to what is traditionally understood as consensus: part or all of the systems decide on a common trajectory. Out of the k systems, k_m systems ($k \geq k_m \geq 2$) are decision-makers and thus exchange information to decide on a common trajectory to be tracked while $k_s = k - k_m$ systems are followers which track the trajectory without participating in its generating. Without any loss of generality, we will assume that the systems numbered from 1 till k_m are decision-makers and the remaining ones are followers. Among the group of decision-makers, all systems exchange estimates of their states with each other. They also send an estimate of their states to the controller of their agent as well as the controller of the followers. The followers only need to send an estimate of their state to their controller to be actuated. Note that since the systems receive an estimate of their state, this protocol with $k_m = 1$ is not the same as the master/slaves protocol. The communication adjacency matrix is

$$\mathfrak{A} = \mathfrak{A}_{\text{mc}} \coloneqq \begin{bmatrix} \mathbb{1}_{k_m \times k_m} & \mathbb{1}_{k_m \times k_s} \\ 0_{k_s \times k_m} & I_{k_s} \end{bmatrix}. \tag{3.15}$$

The mutual protocol controller equations (3.7) are,

$$\mathcal{U}_i(\cdot) = \begin{cases} 0, & \text{if } (t+1) \notin S, \\ \arg\min\limits_{v \in V_{q_{\bar{t}}}} \left\| \sum_{j=1}^{k_m} \frac{v_{m_{ji}(\bar{t})}}{k_m} - v \right\|_P - v^{l_{\bar{t}}}_{m_{ii}(\bar{t})}, & \text{if } (t+1) \in S. \end{cases}$$

The difference between the mutual protocol and the master/slave protocol is that the master does not receive an estimate of its state because it is left uncontrolled (notice that the first entry of $\mathfrak{A}_{ms}$ is zero). The master/slave protocol is thus not the same as the mutual protocol with $k_m = 1$. This explains why both protocols are presented separately. In the next section, we will prove that the sufficient outgoing communication capacity is smaller for the master/slave protocol than for the mutual protocol with $k_m = 1$. Note that the different consensus-preserving protocols that have been presented here are by no means the only protocols that preserve consensus in the sense of Definition 3.3.

3.5 Resulting Rates

The three protocols from the previous section rely on the tunable constants $\delta > 0$ and $\bar{s} \geq 2$. As was mentioned in Section 3.3, the choice of δ and $\bar{s}$ has a capital

impact on the size of the alphabets and hence, on the required outgoing communication capacity. In this section, we provide the main results of the chapter, namely three theorems, one for each of the protocols, which provide sufficient conditions on the communication adjacency matrix and outgoing communication capacities c_i to implement the protocols.

As was briefly explained in Section 3.3, one of the determining factors is how a ball of radius δ expands/contracts under the influence of the mapping φ. It is well-known that the image of a ball under a linear mapping is an ellipsoid whose semi-axes are the right-singular vectors of the matrix associated with this linear map and the length of the axes are the singular values multiplied by the original radius of the ball. In the nonlinear case, the singular values of the Jacobian of the mapping φ have the same effect as long as the original ball is small enough such that the higher-order terms are neglectable. For nonlinear systems, these singular values are generally state-dependent which means that it is difficult to provide upper bounds on the expansion/contraction rate depending on those singular values.

One possibility to get rid of the state-dependency in the singular values of the Jacobian is to use an assumption from Matveev and Pogromsky [2016]. We first introduce the following notations $A^s(x) \coloneqq \frac{\partial \varphi^s}{\partial x}(x)$, $A(x) \coloneqq A^1(x)$, and $X_\xi \coloneqq \{x \in \mathbb{R}^{n_x} | \text{dist}(x, X) < \xi\}$, where ξ is taken from Assumption 3.1. We then impose the following assumption.

Assumption 3.8. *Matveev and Pogromsky [2016] There exist continuous and bounded on X_ξ functions $v_d : \mathbb{R}^n \to \mathbb{R}$, constants $\Lambda_d \geq 0$, $d \in \{1, \ldots, n_x\}$, and a positive definite $n_x \times n_x$ matrix $P = P^\intercal$ such that*

$$\Delta v_d(x) + \sum_{i=1}^{d} \log_2 \lambda_i(x) \leq \Lambda_d, \qquad \forall x \in X_\xi \tag{3.16}$$

for all $d \in \{1, \ldots, n_x\}$, where $\log_2(0) \coloneqq -\infty$ and $\lambda_1(x) \geq \cdots \geq \lambda_n(x) \geq 0$ are the roots of

$$\det\left(A^\intercal(x) P A(x) - \lambda P\right) = 0 \tag{3.17}$$

repeated according to their algebraic multiplicities and $\Delta v_d(x) = v_d(\varphi(x)) - v_d(x)$.

Note that this assumption constitutes an extension of the results obtained in Voortman et al. [2020d] since a wider class of systems can be considered now, thanks to the addition of the auxiliary functions v_d. In the upcoming results, we will use the quantity $\bar{v}$, which is defined as

$$\bar{v} \coloneqq \sup_{x \in X_\xi,\ d \in \{1,\ldots,n\}} |v_d(x)| \tag{3.18}$$

and is finite since the functions v_d are bounded on X_ξ (by Assumption 3.8). We also define

$$\bar{\Lambda} \coloneqq \max_{d\in\{1,\dots,n\}} \Lambda_d. \tag{3.19}$$

The matrix P from Assumption 3.8 is used to define the norm $\|\cdot\|_P$ and hence also the balls used in the construction of the alphabet that was discussed in Section 3.3.

The $\lambda_i(x)$ of the previous assumption are in fact the square singular-values of $A(x)$ expressed in a different coordinate basis. Indeed, decomposing the matrix P as $P = U^\intercal U$ where U is non-singular (since P is positive definite and symmetric such a decomposition always exists) allows us to rewrite (3.17) as $\det\left(A^\intercal(x)U^\intercal UA(x) - \lambda U^\intercal U\right) = 0$ which, since U is non-singular, has the same solutions as $\det\left(U^{-\intercal}A^\intercal(x)U^\intercal UA(x)U^{-1} - \lambda I_n\right) = 0$. The solutions of this equation are thus the squares of the singular values of $U^{-\intercal}A^\intercal(x)U^\intercal UA(x)U^{-1}$, which is equivalent to saying that they are the squares of the singular values of $A(x)$ in a different coordinate basis. One important consequence is that for the norm $\|\cdot\|_P$, the following inequality holds $\|A(x)x\|_P \leq \sqrt{\lambda_1(x)}\,\|x\|_P$.

With Assumption 3.8 in mind, we now present the main contributions of this chapter: the theorems that provide sufficient conditions on the outgoing communication capacities to implement the consensus-preserving protocols. The first theorem provides the bounds for the pacemaker protocol.

Theorem 3.9. *Let Assumptions 3.1 and 3.8 and (3.2) hold for k systems (3.1). Then there exists $\bar{s} > 0$ such that the pacemaker protocol with the communication adjacency matrix $\mathfrak{A} = I_k$ preserves the consensus of the systems over channels with outgoing communication capacities $c_i > \frac{\bar{\Lambda}}{2}$, and consensus factor $G = 2^{\frac{\Lambda_1 \bar{s}+\bar{v}+2}{2}}$.*

The proof of this theorem is provided in Appendix 3.A. The second theorem provides the bounds for the master/slaves protocol.

Theorem 3.10. *Let Assumptions 3.1 and 3.8, and (3.2) hold for k systems (3.1). Then there exists $\bar{s} > 0$ such that the master/slave protocol with the communication adjacency matrix $\mathfrak{A} = \mathfrak{A}_{\mathrm{ms}}$ as defined in (3.14) preserves the consensus of those systems over any channels with outgoing communication capacities $c_1 > (k-1)\frac{\bar{\Lambda}}{2}$, $c_i > \frac{\bar{\Lambda}}{2}$, $\forall i \in \{2,\dots,k\}$, and consensus factor $G = 2^{\frac{\Lambda_1 \bar{s}+\bar{v}+2}{2}}$.*

The proof of this theorem is provided in Appendix 3.A. The third and last theorem provides the bounds for the mutual protocol.

Theorem 3.11. *Let Assumptions 3.1 and 3.8, and (3.2) hold for k systems (3.1). Then there exists $\bar{s} > 0$ such that the mutual protocol with the communication adjacency matrix $\mathfrak{A} = \mathfrak{A}_{\mathrm{mc}}$ as defined in (3.15) preserves the consensus of k systems over any channels with outgoing communication capacities*

$c_i > k\frac{\bar{\Lambda}}{2}, \forall i \in \{1, \ldots, k_m\}$, $c_i > \frac{\bar{\Lambda}}{2}$, $\forall i \in \{k_m + 1, \ldots, k\}$ *and consensus factor* $G = 2^{\frac{\Lambda_1 \bar{s} + \bar{v} + 2}{2}}$.

The proof of this theorem is provided in Appendix 3.A.

The three protocols are ranked in terms of increasing interactions. In this context, interactions mean how many systems' current states are taken into account when deciding on the next point in the common trajectory. The pacemaker protocol requiring the least interactions (none of the systems' states are used to determine the common trajectory) whilst the mutual protocol requires the most (with $k_m = k$, all the system' states are taken into account). These interactions come at a cost, if we denote by c^* the total outgoing communication capacity ($c^* = \sum_i c_i$), then we have that

$$c^*_{\text{pm}} = \frac{k\bar{\Lambda}}{2},\ c^*_{\text{ms}} = \frac{(2k-2)\bar{\Lambda}}{2},\ c^*_{\text{mc}} = \frac{(k_m(k-1)+k)\bar{\Lambda}}{2}, \tag{3.20}$$

where the subscripts pm, ms and mc denote the pacemaker, master/slaves and mutual protocol respectively. Note that for $k > 2$, we have $c^*_{\text{pm}} < c^*_{\text{ms}} < c^*_{\text{mc}}$. Higher levels of interactions thus require higher total outgoing communication capacities.

3.6 Examples

In this section, we illustrate the use of the previous theorems by applying them to two examples. We consider the problem of consensus preservation first for a network of logistic maps, then for a network of Hénon maps. For both of these networks, we will apply Theorems 3.9, 3.10, and 3.11 to find the sufficient outgoing communication capacities to implement the consensus-preserving protocols. These capacities are then compared to the rates used by the protocols in simulations to validate the theoretical bounds.

3.6.1 The Logistic Map

The logisitic map is a single-dimensional map, introduced in May [1976], with the following state equation

$$\varphi_\gamma : \{x \to \gamma x\,(1-x)\}\,, \tag{3.21}$$

with state-space $X = [0, 1]$, where γ is a positive parameter. For $\gamma \in [0, 4]$, the set X is positively invariant. Depending on γ, the system has one or two equilibria $x^{\text{eq}}_{1,2} \in X$:

$$x^{\text{eq}}_1 = 0, \qquad\qquad x^{\text{eq}}_2 = \frac{\gamma - 1}{\gamma},$$

where the latter equilibrium only exists for $\gamma \in (1, 4]$. The system displays a wide array of different behaviors (see e.g. Strogatz [1994] for a detailed description of the behaviors). One particular property of the system is that for most values of $\gamma \geq 3.56995$, the system displays chaotic behavior. We thus consider k systems with the following equations

$$x_i(t+1) = \varphi_\gamma(x_i(t)) + u_i(t).$$

This system does not satisfy the final part of Assumption 3.1: there is no $\xi > 0$ such that ξ-neighbourhood of X is attractive. However, this assumption is only made to ensure that for any point belonging to the covering of X with balls, its image through the mapping of the system is inside of X (a property that is necessary for our protocols). Since $[0, 1]$ can be covered with balls in $\mathbb{R}$ such that no point of the balls fall outside of $[0, 1]$, the consensus-preserving protocols and their associated theorems can still be applied. Applying the three previous theorems to the logistic maps gives the following proposition

Proposition 3.12. *Theorems 3.9, 3.10, and 3.11 hold for k logistic maps with $\bar{\Lambda} = 2\log_2\gamma$ and $G = 2\gamma^{\bar{s}}$.*

Proof: In order to prove the proposition, we need to verify Assumption 3.8. We start by computing the Jacobian of the system $\frac{\partial \varphi_\gamma}{\partial x} = \gamma - 2\gamma x$. We thus have that $\lambda_1(x) = (\gamma - 2\gamma x)^2$ is a solution of (3.17) for the logistic map. Over the state-space $X = [0, 1]$, $\lambda_1(x)$ has a global maximum at $x = 0$: $\lambda_1(0) = \gamma^2$. We thus have that Assumption 3.8 holds with $P = 1$, $\Lambda_1 = \bar{\Lambda} = 2\log_2(\gamma)$, and $v_1 = 0$. □

For the simulations, we consider two logistic maps each with the same parameter $\gamma = 3.97$ and $X = [0, 1]$. For the mutual protocol, we consider the case $k_m = 2$ only. Proposition 3.12 holds with $\bar{\Lambda} = 3.9782$. This implies that the theoretical bounds on the capacities are $c_{\text{pm}} > 3.9782$, $c_{\text{ms}} > 3.9782$ and $c_{\text{mc}} > 7.9564$. Note that the pacemaker protocol and master/slaves protocol have the same bound on the total outgoing communication capacity, which is logical since in both cases, each sensor transmits estimates of its state to only one controller. The mutual protocol requires twice as large of a total outgoing communication capacity, as follows from (3.20). This is due to the fact that each of the sensors needs to send estimates of their state to both controllers, which requires twice as many communications as for the other protocols.

A Monte-Carlo simulation method was used for all three protocols. For each protocol, the two systems were given random initial states in $[0, 1]$, within a distance δ of each other, 1000 different times. For each of those 1000 different initial conditions, the systems and the consensus-preserving protocols were then simulated for 1000 timesteps for various choices of $\bar{s}$. In all cases, the consensus was preserved. The choices of δ and the associated results are displayed in Table 3.1. The quantity G is dimensionless while the total communication rates R^* are given in bits per time instant.

	$\bar{s}=2$ $\delta=10^{-2}$	$\bar{s}=5$ $\delta=10^{-4}$	$\bar{s}=10$ $\delta=10^{-8}$	$\bar{s}=20$ $\delta=10^{-13}$
G	3.97	1972	194194×10^{6}	1.89×10^{12}
R^*_{pm}	4.98	3.17	2.58	2.28
R^*_{ms}	4.97	3.18	2.60	2.27
R^*_{mc}	9.99	6.38	5.18	4.58

Table 3.1: Results for the logistic maps.

We make the following comments about these results. For $\bar{s}=2$, the rates R^* are above the theoretical thresholds c^* (which are 3.9782, 3.9782 and 7.9564). For larger $\bar{s}$ the resulting communication rates are below the theoretical capacity threshold. This effect is accentuated for the largest choices of $\bar{s}$. This is partially due to the fact that the worst initial states (i.e. the initial conditions requiring the highest rate) are not selected every time. The reduction in communication rate comes at a price: the consensus factor is larger for larger $\bar{s}$. There is thus a trade-off between the rate and consensus factor. Note that the small differences between the rates R^*_{pm} and R^*_{ms} are due to the random nature of Monte-Carlo type simulation methods.

3.6.2 The Hénon map

The Hénon map is a two-dimensional discrete-time dynamical system that was first introduced in Henon [1976]. It is described by the following map

$$\varphi_H : \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} a + bx_2 - x_1^2 \\ x_1 \end{pmatrix} \right\}, \tag{3.22}$$

where a and b are positive parameters. The most studied parameters are the combination $a=1.4$ and $b=0.3$ for which simulations show both bounded and unbounded trajectories. Moreover, for these parameter values, the system has a strange attractor which is a typical manifestation of chaotic properties. The system has two equilibrium points of (x^+, x^+) and (x^-, x^-)

$$x^+ = \frac{b\text{-}1 + \sqrt{(b\text{-}1)^2 + 4a}}{2}, \qquad x^- = \frac{b\text{-}1 - \sqrt{(b\text{-}1)^2 + 4a}}{2}$$

For the parameter values $a=1.4$ and $b=0.3$ both equilibria are saddle points. As was proven in Henon [1976], one of the particularities of the system is that is has a positively invariant set: the quadrilateral $ABCD$ with vertices $A=(-1.33, 0.42)$, $B=(1.32, 0.133)$, $C=(1.245, -0.14)$, $D=(-1.06, -0.5)$. The equilibrium (x^+, x^+) lies inside of this quadrilateral while the other equilibrium point lies outside of it. We thus consider k Hénon maps of the following form

$$x_i(t+1) = \varphi_H(x_i(t)) + u_i(t)$$

with the aforementioned quadrilateral $ABCD$ as the set X and $\xi = 10^{-3}$ for Assumption 3.1. Applying the three theorems to these Hénon maps gives the following proposition

Proposition 3.13. *Theorems 3.9, 3.10, and 3.11 hold for k Hénon maps with* $\bar{\Lambda} = 2\log_2\left(\sqrt{(x^-)^2 + b} - x^-\right)$ *and* $G = 2\left(\sqrt{(x^-)^2 + b} - x^-\right)^{\bar{s}}$.

The proof of this proposition simply follows from the proof of Theorem 15 from Matveev and Pogromsky [2016].

For the simulations, we considered 4 Hénon maps with the parameter values $a = 1.4$ and $b = 0.3$. The objective of the simulations are to compare the total outgoing communication capacities c^* for the pacemaker protocol, master/slave protocol, and the mutual protocol with $k_m \in \{2, \ldots, 4\}$. Proposition 3.13 holds with $\bar{\Lambda} = 3.41$. This implies that the theoretical bounds on the total outgoing communication capacities are $c_{\text{pm}} > 6.82$, $c_{\text{ms}} > 10.23$, $c_{\text{mc}} > 17.05$ ($k_m = 2$), $c_{\text{mc}} > 22.16$ ($k_m = 3$), and $c_{\text{mc}} > 27.28$ ($k_m = 4$). Monte Carlo methods were again used for simulations: 1000 times, the four systems were given random initial conditions within the quadrilateral $ABCD$ and within a distance of 2δ of each other. For each of these initial states, the state-space trajectories of the systems as well as the consensus-preserving protocols were simulated for 1000 timesteps. For all simulations the consensus was preserved. The results in terms of total communication rate are displayed in Table 3.2.

We make the following observations about the simulations. For $\bar{s}$, the rates are again above the theoretical thresholds. For $\bar{s} \geq 3$, the rates are below the theoretical capacities (for the same reasons as with the previous example). More cooperation between the systems in determining the target trajectory indeed requires a higher communication rate, as can be seen from the increasing rates from one protocol to the next. The consensus factor again increases large for the larger choices of $\bar{s}$. There is thus again a trade-off between the resulting rate and consensus factor. The large consensus factors imply that the systems can drift relatively far away from each other, in comparison with how close to each other they start initially. In practice, however, and given the choices of δ, even in the case of $\bar{s} = 20$, the systems are never further than 0.00367 units of distance away from each other.

3.7 Conclusion

This chapter provided a solution to the problem of consensus preservation in a network of nonlinear dynamical systems which communicate over channels with limited transmission capacities. The systems in the network were equipped with smart sensors and controllers which were placed at locations remote from one another and thus forced them to use the channels to communicate with each other. The problem of consensus preservation was introduced. This problem implies

	$\bar{s}=2$ $\delta=10^{-3}$	$\bar{s}=5$ $\delta=10^{-7}$	$\bar{s}=10$ $\delta=10^{-10}$	$\bar{s}=20$ $\delta=10^{-13}$
G	21.25	736	2.72×10^{5}	3.68×10^{10}
R^{*}_{pm}	10.47	6.23	4.82	4.12
R^{*}_{ms}	15.71	9.35	7.23	6.17
$R^{*}_{\mathrm{mc}}\,(k_m=2)$	26.18	15.59	12.06	10.29
$R^{*}_{\mathrm{mc}}\,(k_m=3)$	34.03	20.26	15.67	13.38
$R^{*}_{\mathrm{mc}}\,(k_m=4)$	41.89	24.94	19.29	16.46

Table 3.2: Results for the Hénon maps.

steering systems that start in a small vicinity of one another, such that they remain close to each other. Sensors and controllers which preserved the consensus were referred to as consensus-preserving protocols. The chapter then answered the following questions: "What protocols preserve consensus?", "What are the sufficient outgoing communication capacities necessary to implement such protocols?", and "How do these quantities depend on the system's equations?". Several answers, in the form of consensus-preserving protocols, were provided. Together with these protocols, three theorems that give sufficient conditions on the outgoing communication capacities were proven. The sufficient capacities were proven to depend on the larger-than-one singular values of the linear part of the mapping of the systems. The protocols were tested by simulations of consensus preservation in networks of logistic maps and networks of Hénon maps. For both of these types of systems, analytical bounds on the sufficient capacities in the communication channels to implement the consensus-preserving protocols were provided. To validate these protocols, the bounds were compared with the rates observed in simulations. Future extensions of this work include developing consensus-preserving protocols that constantly control the agents by means of relatively small inputs contrary to the scheme proposed in this chapter where a large input is applied only just prior to every current communication time, while the agents are left uncontrolled at all other times, as well as, considering more general situations with non-additive inputs, and considering communication networks with multi-terminal coding (see Chapter 9 of Yüksel and Başar [2013]).

Appendices

3.A Proofs of the Results from Section 3.5

We define $\lambda_i(M)$, where M is square matrix, as the solutions of

$$\det\left(M^{\intercal}PM-\lambda P\right)=0,$$

where these solutions are ranked in decreasing order and repeated according to their algebraic multiplicity. We also define $\sigma_i^P(M) \coloneqq \sqrt{\lambda_i(M)}$. Note that by decomposing P as $P = U^\intercal U$, where U is non-singular (since the matrix P is positive definite, such a decomposition always exists), the solutions of the previous equation correspond to the squares of the singular values of the operator defined by the matrix M expressed in a different coordinate basis (which is because the solutions of the previous equation are identical to those of $\det\left(U^{-\intercal}M^\intercal U^\intercal U M U^{-1} - \lambda I_n\right) = 0$). We will refer to those $\sigma_i^P(M)$ as P-generalized singular values. Note that for the P-generalized singular values of $A(x)$, we have $\sigma_i^P\left(A(x)\right) = \sqrt{\lambda_i(x)}$, where $\lambda_i(x)$ are taken from Assumption 3.8.

3.A.1 Lemmata from other papers

For the convenience of the reader, we start with formulation of three lemmata that were established in Matveev and Pogromsky [2016], Voortman et al. [2018b], Pogromsky and Matveev [2011] and will be used to prove our main results.

Lemma 3.14. *Matveev and Pogromsky [2016] For any $x \in X_\xi$ $\bar{s} \geq 1$ and $d \in \{1, \ldots, n_x\}$ we have $\prod_{i=1}^{d} \sigma_i^P\left(A^{\bar{s}}(x)\right) \leq 2^{\frac{\bar{\Lambda}\bar{s}}{2} + \bar{v}}$, where $\bar{v}$ is defined in (3.18) and $\bar{\Lambda}$ is defined in (3.19).*

Lemma 3.15. *Voortman et al. [2018b] Let Assumptions 3.1 and 3.8 hold. Then, for any $\epsilon > 0$, there exists δ_1^* such that for all $\delta : 0 < \delta \leq \delta_1^*$, $x_i, x_j \in X_\xi$ for which $\|x_i - x_j\|_P \leq \delta$, we have $\|\varphi^q(x_i) - \varphi^q(x_j)\|_P \leq \epsilon$, $\forall\ 0 < q \leq \frac{2\log_2(\epsilon) - 2\log_2(\delta) - \bar{v}}{\Lambda_1}$.*

Lemma 3.16. *Pogromsky and Matveev [2011] For any non-negative real numbers $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq 0$ enumerated in descending order, the following is true $\prod_{i=1}^{n} \lceil \alpha_i \rceil \leq \max\left\{2^n \max_{d \in [1:n]} \prod_{i=1}^{n} \alpha_i; 1\right\}$.*

3.A.2 Auxiliary Results

Lemma 3.17. *For any $\bar{s} \geq 1$, $\eta > 0$, and $P = P^\intercal \succ 0$, there exists a $\delta_2^* > 0$ such that for all δ for which $0 < \delta \leq \delta_2^*$, $x \in X_\xi$, and $x_0 \in X_\xi$ for which $\|x - x_0\|_P \leq \delta$, $\|\varphi^{\bar{s}}(x) - \varphi^{\bar{s}}(x_0) - A^{\bar{s}}(x)(x - x_0)\|_P \leq \eta\delta$ holds.*

Proof: The proof of this lemma simply follows from the continuous differentiability of $f(\cdot, \cdot)$ from Assumption 3.1. □

In what follows, we use the term *orthotope* to refer to a P-orthotope, an orthotope in $\|\cdot\|_P$.

Lemma 3.18. *For any $\bar{s} \geq 1$, $\eta > 0$, and $P = P^\intercal \succ 0$, there exists $\delta_3^* > 0$ such that for all $\delta : 0 < \delta \leq \delta_3^*$ and $x \in X$, $\varphi^{\bar{s}}(B_\delta(x))$ is inscribed in an orthotope, centered in $\varphi^{\bar{s}}(x)$ and with semi-axes of length $\sigma_i^P\left(A^{\bar{s}}(x)\right)\delta + \eta\delta$, $i \in \{1, \ldots, n_x\}$.*

Proof: Let δ_3^* be chosen smaller than or equal to δ_2^* from Lemma 3.17. The proof then simply follows from the definition of the generalized singular values σ_i^P as well as Lemma 3.17. □

Lemma 3.19. *For any $\bar{s} \geq 1$ and $\eta > 0$, there exists $\delta_4^* > 0$ and a set V satisfying Definition 3.4 such that for all $\delta : 0 < \delta \leq \delta_4^*$ and for each ball $B_\delta(v_q)$ with $v_q \in V$, its image I_q can be covered by $m_q = |V_q|$ balls $B_\delta(v_1^q), B_\delta(v_2^q), \ldots, B_\delta(v_{|V_q|}^q)$, where $|V_q| \leq \prod_{i=1}^{n_x} \left\lceil 2\sqrt{n_x}\sigma_i^P\left(A^{\bar{s}}(v_q)\right) + 2\sqrt{n_x}\eta \right\rceil$.*

Proof: Let δ_4^* be chosen smaller than or equal to δ_3^* from Lemma 3.18 and smaller than ξ. Lemma 3.18 holds for all δ smaller than δ_3^* and hence also for δ_4^*. Next, construct V as follows: take n_x-dimensional grid, with a distance in $\|\cdot\|_P$ of $\delta_4^*/2\sqrt{n_x}$ between neighbouring grid elements. Next, select all grid points within a distance of $\delta_4^*/2$ or less from X. Let the set of these points be V_-. Finally map each grid point to (one of) the closest element(s) in X to that point. Let V be a set constructed in the aforementioned manner, then $X \subseteq \bigcup_{q=1}^{|V|} B_{\delta_4^*/2}(v_q)$, which we now prove by contradiction. Let there be $x \in X$ such that $x \notin \bigcup_{q=1}^{|V|} B_{\delta_4^*/2}(v_q)$. x is at a distance of at most $\delta_4^*/4$ to the nearest point in V_- (this is a property of an n_x-dimensional grid with distance $\delta_4^*/2\sqrt{n_x}$ between neighbouring grid points). Let y be one of these points in V_- at a distance of at most $\delta_4^*/4$ of x. $x \notin \bigcup_{q=1}^{|V|} B_{\delta_4^*/2}(v_q)$ implies that there exists $z \in X$, to which y was shifted. Moreover, z was at most at a distance of $\delta_4^*/4$ to y, or else y would have been shifted to x (because it would have been closer). Since the maximum distance between x and y is at most $\delta_4^*/4$, and the maximum distance between y and z is also at most $\delta_4^*/4$, the total distance between x and z is at most $\delta_4^*/2$, which implies that $x \in \bigcup_{q=1}^{|V|} B_{\delta_4^*/2}(v_q)$, we thus have a contradiction. V constructed as above also satisfies Definition 3.4 (because for Definition 3.4, balls of radius δ_4^* are used). From Lemma 3.18, the image I_q of the balls with centres in V is inscribed in a orthotope with semi-axes of length $\sigma_i^P\left(A^{\bar{s}}(v_q)\right)\delta_4^* + \eta\delta_4^*$. Next, consider an n_x-dimensional grid with distance $\delta_4^*/\sqrt{n_x}$ between neighbouring grid elements, oriented such that the grid lines are parallel with the sides of the orthotope. Keep all the points in this grid which are inside of the orthotope. Evidently, there are at most $\prod_{i=1}^{n_x} \left\lceil \frac{2\sigma_i^P\left(A^{\bar{s}}(v_q)\right)\delta_4^* + 2\eta\delta_4^*}{\delta_4^*/\sqrt{n_x}} \right\rceil$ such points (the length of the axes of the orthotope divided by the spacing between neighbouring gridpoints). For each of these points, select the closest element in V. Let the sets constructed in the aforementioned manner be V_q. We prove, again by contradiction, that $I_q \subseteq \bigcup_{p=1}^{|V_q|} B_{\delta_4^*}(v_p^q)$. Let there be $\bar{x} \in I_q$ such that $\bar{x} \notin \bigcup_{p=1}^{|V_q|} B_{\delta_4^*}(v_p^q)$. Since $\bar{x}$ is in I_q, it is also inside of the orthotope inscribed around I_q. Let $\bar{y}$ be the closest point in the grid that was placed over the orthotope. Because the distance between the neighbouring grid points was $\delta_4^*/\sqrt{n_x}$, the distance from $\bar{x}$ to $\bar{y}$ is at most $\delta_4^*/2$. Let $\bar{z}$ be the point to which $\bar{y}$ was mapped (when V_- was

mapped to V). Since the points in the grid are mapped to the closest point in V and the points in V form a covering of balls of radius $\delta_4^*/2$ of X, the distance between $\bar{y}$ and $\bar{z}$ is at most $\delta_4^*/2$. This implies that the distance between $\bar{x}$ and $\bar{z}$ is at most δ_4^* and hence $\bar{x} \in \bigcup_{p=1}^{|V_q|} B_{\delta_4^*}(v_p^q)$. We set V_q constructed as above thus satisfies Definition 3.5 and the image I_q can thus be covered by at most $\prod_{i=1}^{n_x} \left\lceil 2\sqrt{n_x}\sigma_i^P\left(A^{\bar{s}}(v_q)\right) + 2\sqrt{n_x}\eta \right\rceil$ balls. Evidently, the above can be repeated for $\delta : 0 < \delta \leq \delta_4^*$. □

Lemma 3.20. *Let Assumptions 3.1 and 3.8, and (3.2) hold for system (3.11). There exists a $\delta_5^* > 0$ such that for all $\delta : 0 < \delta \leq \delta_5^*$, the number of elements in the sets V_q as defined in Definition 3.5 is upper bounded as follows $|V_q| \leq 8^{n_x} n_x^{n_x/2} 2^{\frac{\bar{\Lambda}\bar{s}}{2}+\bar{v}}$, where $\bar{v}$ is defined in (3.18) and $\bar{\Lambda}$ is defined in (3.19).*

Proof: For any $\bar{s} \geq 2$ and $\eta = 1/2$, Lemma 3.19 guarantees that there exists δ_4^* such that for all $\delta : 0 < \delta \leq \delta_4^*$ the number of balls required to cover the image of $B_\delta(v_q)$ is $|V_q| \leq \prod_{i=1}^{n_x} \left\lceil 2\sqrt{n_x}\sigma_i^P\left(A^{\bar{s}}(v_q)\right) + \sqrt{n_x} \right\rceil$. Let δ_5^* be equal to this δ_4^*. Next, we use Lemma 3.16, to obtain

$$|V_q| \leq \max\left\{ 2^{n_x} \max_{d\in[1,\ldots,n_x]} \prod_{i=1}^{d} \left(2\sqrt{n_x}\sigma_i^P\left(A^{\bar{s}}(v_q)\right) + \sqrt{n_x}\right), 1 \right\},$$

which implies

$$|V_q| \leq \max\left\{ 4^{n_x} n_x^{n_x/2} \max_{d\in[1,\ldots,n_x]} \prod_{i=1}^{d} \left(\sigma_i^P\left(A^{\bar{s}}(v_q)\right) + \frac{1}{2}\right), 1 \right\}. \tag{3.23}$$

Here, $\max_{d\in[1,\ldots,n_x]}$ implies that we only consider i such that $\sigma_i^P\left(A^{\bar{s}}(v_q)\right) + \frac{1}{2} > 1$. We thus have $\sigma_i^P\left(A^{\bar{s}}(v_q)\right) > \frac{1}{2}$ and (3.23) can be rewritten as

$$|V_q| \leq \max\Big\{1,$$
$$4^{n_x} n_x^{n_x/2} \max_{d\in[1,\ldots,n_x]} \prod_{i=1}^{d} \left(1 + \frac{1}{2\sigma_i^P\left(A^{\bar{s}}(v_q)\right)}\right) \sigma_i^P\left(A^{\bar{s}}(v_q)\right)\Big\}.$$

Since $\sigma_i^P\left(A^{\bar{s}}(v_q)\right) > \frac{1}{2}$, this implies that

$$|V_q| \leq \max\left\{ 8^{n_x} n_x^{n_x/2} \max_{d\in[1,\ldots,n_x]} \prod_{i=1}^{d} \sigma_i^P\left(A^{\bar{s}}(v_q)\right), 1 \right\}.$$

Applying Lemma 1 to this inequality yields

$$|V_q| \leq \max\left\{ 8^{n_x} n_x^{n_x/2} \max_{d\in[1,\ldots,n_x]} 2^{\frac{\bar{\Lambda}\bar{s}}{2}+\bar{v}}, 1 \right\}.$$

which, given that $\bar{\Lambda} \geq 0$ implies $|V_q| \leq 8^{n_x} n_x^{n_x/2} 2^{\frac{\bar{\Lambda}\bar{s}}{2}+\bar{v}}$. Since Lemma 3.19 also applies to $\delta : 0 < \delta \leq \delta_4^* = \delta_5^*$ and the rest of the proof is independent of δ, the proof is complete. □

Lemma 3.21. *Let Assumptions3.1 and 3.8, and (3.2) hold for system (3.11). For any $\bar{s} \geq 1$, there exists a δ_6^* such that for all $\delta : 0 < \delta \leq \delta_6^*$, if the states of all systems are within the same ball of radius δ at $t \in S = \{0, \bar{s}, 2\bar{s}, \ldots\}$ then $G = 2^{\frac{\Lambda_1 \bar{s}+\bar{v}+2}{2}}$ is a consensus factor.*

Proof: Take $\delta_6^* = \delta_1^*/2$ from Lemma 3.15 with $\epsilon = 2^{\frac{\Lambda_1 \bar{s}+\bar{v}}{2}}\delta_1^*$. If the states of all systems are within the same ball of radius $\delta : 0 < \delta \leq \delta_6^*$ (i.e., if $\|x_i(t) - x_j(t)\|_P \leq 2\delta$), then, by Lemma 3.15, we have that $\|x_i(t) - x_j(t)\|_P \leq 2\delta$ implies $\|\varphi^q(x_i(t)) - \varphi^q(x_j(t))\|_P \leq 2^{\frac{\Lambda_1 \bar{s}+\bar{v}}{2}} 2\delta$, $\forall x_i(t), x_j(t) \in \mathbb{R}^{n_x}$, $i, j \in \{1, \ldots, k\}, q \in \{1, \ldots, \bar{s}\}$, $t \in S$. For all consensus protocols $u_i(t) = 0$, $\forall t : t+1 \notin S$, which implies that $\forall t : t+1 \notin S$, $x_i(t+1) = \varphi(x_i(t))$. Combining this property with the previous inequality is equivalent to stating that $\|x_i(t+q) - x_j(t+q)\|_P \leq 2^{\frac{\Lambda_1 \bar{s}+\bar{v}+2}{2}}\delta$, $\forall i, j \in \{1, \ldots, k\}$, $q \in \{1, \ldots, \bar{s}-1\}$, $t \in S$. This implies that $\|x_i(t) - x_j(t)\|_P \leq 2^{\frac{\Lambda_1 \bar{s}+\bar{v}+2}{2}}\delta$, $\forall i, j \in \{1, \ldots, k\}$, $\forall t \geq 0$. From the definition of the consensus factor, we thus have that $G = 2^{\frac{\Lambda_1 \bar{s}+\bar{v}+2}{2}}$ is a consensus factor. □

Lemma 3.22. *For any communication channel $(i, j) \in J$, $c_i > \frac{\Lambda}{2}$ implies $\mathfrak{c}_{ij} > \frac{\Lambda}{2\sum_{q=1:k} \mathfrak{a}_{iq}}$.*

Proof: We have $c_i > \alpha$ which implies $\min_{j:\mathfrak{a}_{ij}=1} \mathfrak{c}_{ij} \sum_{q=1}^{k} \mathfrak{a}_{iq} > \alpha$ from (3.9). The proof is completed by noticing that $\mathfrak{c}_{ij} \geq \min_{j:\mathfrak{a}_{ij}=1} \mathfrak{c}_{ij}$.

□

Lemma 3.23. *If, for all $(i, j) \in J$, the length of the alphabet verifies $l_{ij}(t) \leq 8^{n_x} n_x^{n_x/2} 2^{\frac{\bar{\Lambda}\bar{s}}{2}+\bar{v}}$, and the channel capacities verify $\mathfrak{c}_{ij} > \frac{\bar{\Lambda}}{2}$, there exists $\bar{s}$ such that the channel rate constraint (3.6) holds.*

Proof: For all $(i, j) \in J$, we have $\frac{\bar{\Lambda}}{2} < \mathfrak{c}_{ij} := \liminf_{\bar{s}\to\infty} \frac{b_{ij}^+(\bar{s})}{\bar{s}}$. This implies that there exist $\mu_{ij} > 0$ such that $\liminf_{\bar{s}\to\infty} \frac{b_{ij}^+(\bar{s})}{\bar{s}} = \frac{\bar{\Lambda}}{2} + \mu_{ij}$. From the definition of the inferior limit, for any $\epsilon_{ij} > 0$, there exist $\bar{s}_{ij}^*$ such that $\forall \bar{s}_{ij} \geq \bar{s}_{ij}^*$, $\frac{b_{ij}^+(\bar{s}_{ij})}{\bar{s}_{ij}} > \frac{\bar{\Lambda}}{2} + \mu_{ij} - \epsilon_{ij}$. This implies that for $\epsilon_{ij} = \mu_{ij}/2$, there exists $\bar{s}_{ij}^*$ such that for all $\bar{s}_{ij} \geq \bar{s}_{ij}^*$

$$\frac{b_{ij}^+(\bar{s}_{ij})}{\bar{s}_{ij}} > \frac{\bar{\Lambda}}{2} + \mu_{ij}. \tag{3.24}$$

Meanwhile, $\log_2 l_{ij}(t) \leq \log_2 \left(8^{n_x} n_x^{n_x/2} 2^{\frac{\bar{\Lambda}\bar{s}}{2}+\bar{v}}\right) = \frac{\bar{\Lambda}\bar{s}}{2} + \bar{v} + 3n_x + \log_2 n_x^{n_x/2}$. Since $\lim_{\bar{s}\to\infty} \frac{\frac{\bar{\Lambda}\bar{s}}{2}+\bar{v}+3n_x+\log_2 n_x^{n_x/2}}{\bar{s}} = \frac{\bar{\Lambda}}{2}$, there exists $\bar{s}^\dagger$ such that $\forall \bar{s} \geq \bar{s}^\dagger$, $\frac{\frac{\bar{\Lambda}\bar{s}}{2}+\bar{v}+3n_x+\log_2 n_x^{n_x/2}}{\bar{s}} \leq \frac{\bar{\Lambda}}{2} + \mu_{ij}$, $\forall \mu_{ij}$. Let $\bar{s} = \max\{\bar{s}_{ij}^*, \bar{s}^\dagger\}$, starting again from

(3.24), we have

$$\frac{b_{ij}^{+}(\bar{s})}{\bar{s}} > \frac{\bar{\Lambda}}{2} + \mu_i \geq \frac{\frac{\bar{\Lambda}\bar{s}}{2} + \bar{v} + 3n_x + \log_2 n_x^{n_x/2}}{\bar{s}} \geq \frac{\log_2 l_{ij}(t)}{\bar{s}},$$

which evidently implies that the channel rate constraints (3.6) are verified. □

3.A.3 Proof of Theorem 3.9

For the protocol to preserve consensus, it needs to verify both conditions specified in Definition 3.3. Let δ^* be chosen smaller than ξ and smaller or equal to δ_6^* from Lemma 3.21. Then for any $\delta : 0 < \delta \leq \delta^*$, the following reasoning can be applied.

Regarding Definition 3.3, (i): If at the communication instants the states of all agents are within the same ball of radius δ, (i) follows from Lemma 3.21. We prove by induction that the states of all agents are indeed within the same ball of radius δ at $t \in S$. For the first time instant, this property holds, due to (3.2). Assume that it holds for t, we prove that it holds for $t + \bar{s}$. If a zero control input is applied at $t + \bar{s} - 1$, the states of the systems end up in the balls with centres $v_{m_{ii}(t)}^{l_t}$ (because the messages $m_{ii}(t)$ correspond to the indices of the balls in which the states end up if they were left unactuated). This implies that $\|\varphi(x_i(t + \bar{s} - 1)) - v_{m_{ii}(t)}^{l_t}\|_P \leq \delta$, $\forall i$, if $u(t + \bar{s} - 1) = 0$. However, the additive control input $u(t + \bar{s} - 1) = r(t + \bar{s}) - v_{m_{ii}(t)}^{l_t}$ is applied, $\forall i$, implying that instead we have $x_i(t + \bar{s}) = \varphi(x_i(t + \bar{s} - 1)) + r(t + \bar{s}) - v_{m_{ii}(t)}^{l_t}$, $\forall i$, hence $\varphi(x_i(t + \bar{s} - 1)) = x_i(t + \bar{s}) - r(t + \bar{s}) + v_{m_{ii}(t)}^{l_t}$ which implies that we have $\|x_i(t + \bar{s}) - r(t + \bar{s}) + v_{m_{ii}(t)}^{l_t} - v_{m_{ii}(t)}^{l_t}\|_P = \|x_i(t + \bar{s}) - r(t + \bar{s})\|_P \leq \delta$, $\forall i$, which implies that all the states x_i of the systems are within the same ball centred in $r(t + \bar{s})$ of radius δ at $t + \bar{s}$.

Regarding Definition 3.3, (ii): The first half of item (ii) trivially holds from the alphabet function (3.12) and encoder function (3.13). For the second half of item (ii) to hold, the channel rate constraints (3.6) must hold. From the theorem statement we have $\mathfrak{a}_{ii} = 1$, $\forall i \in \{1, \ldots, k\}$. From Lemma 3.22 and since $c_i > \frac{\bar{\Lambda}}{2}$ and $\sum_{j=1}^{k} \mathfrak{a}_{ij} = 1$, we have $\frac{\bar{\Lambda}}{2} < \mathfrak{c}_{ii}$. From Lemma 3.20, we have $|V_{q_t}| \leq 8^{n_x} n_x^{n_x/2} 2^{\frac{\bar{\Lambda}\bar{s}}{2} + \bar{v}}$. From (3.12), we have $l_{ij}(t) = |V_{q_t}|$. We thus use Lemma 3.23, which implies that there exists $\bar{s}$ such that channel rate constraints hold.□

3.A.4 Proof of Theorem 3.10

We follow the same structure for the proof of this theorem as for the proof Theorem 3.9. Let δ^* be chosen smaller than ξ and smaller or equal to δ_6^* from Lemma 3.21. Then for any $\delta : 0 < \delta \leq \delta^*$, the following reasoning can be applied.

Regarding Definition 3.3, (i): If at the communication instants the states of all agents are within the same ball of radius δ, (i) follows from Lemma 3.21. We prove by induction that the states of all agents are indeed within the same ball of radius δ at $t \in S$. For the first time instant, this property holds, due to (3.2). Assume that it holds for t, we prove that it holds for $t+\bar{s}$. If a zero control input is applied at $t+\bar{s}-1$, the states of the systems end up in the balls with centres $v^{l_t}_{m_{ii}(t)}$ (because the messages $m_{ii}(t)$ correspond to the indices of the balls in which the states end up if they were left unactuated). This implies that $\|\varphi(x_i(t+\bar{s}-1)) - v^{l_t}_{m_{ii}(t)}\|_P \leq \delta$, $\forall i$, if $u(t+\bar{s}-1) = 0$. For system 1, the control input is zero at all times, which implies that we have $\|\varphi(x_i(t+\bar{s}-1)) - v^{l_t}_{m_{ii}(t)}\|_P \leq \delta$. For all other systems, i the additive control input $u(t+\bar{s}-1) = v^{l_t}_{m_{11}(t)} - v^{l_t}_{m_{ii}(t)}$ is applied, implying that instead we have $x_i(t+\bar{s}) = \varphi(x_i(t+\bar{s}-1)) + v^{l_t}_{m_{11}(t)} - v^{l_t}_{m_{ii}(t)}$, $\forall i$, hence $\varphi(x_i(t+\bar{s}-1)) = x_i(t+\bar{s}) - v^{l_t}_{m_{11}(t)} + v^{l_t}_{m_{ii}(t)}$ which implies that we have $\|x_i(t+\bar{s}) - v^{l_t}_{m_{11}(t)} + v^{l_t}_{m_{ii}(t)} - v^{l_t}_{m_{ii}(t)}\|_P = \|x_i(t+\bar{s}) - v^{l_t}_{m_{11}(t)}\|_P \leq \delta$, $\forall i$, which implies that all the states x_i of the systems are within the same ball centred in $v^{l_t}_{m_{11}(t)}$ and of radius δ at $t+\bar{s}$.

Regarding Definition 3.3, (ii): The first half of item (ii) trivially holds from the alphabet function (3.12) and encoder function (3.13). For the second half of item (ii) to hold, the channel rate constraints (3.6) must hold. From the theorem statement we have $\mathfrak{a}_{1i} = 1$, $\forall i \in \{2, \ldots, k\}$. From Lemma 3.22 and since $c_i > \frac{(k-1)\bar{\Lambda}}{2}$ and $\sum_{j=1}^{k} \mathfrak{a}_{1j} = k-1$, we have $\frac{\bar{\Lambda}}{2} < \mathfrak{c}_{1j}$ $\forall j \in \{2, \ldots, k\}$. From the theorem statement we also have $\mathfrak{a}_{ii} = 1$, $\forall i \in \{2, \ldots, k\}$. Again from Lemma 3.22 and since $c_i > \frac{\bar{\Lambda}}{2}$ and $\sum_{j=1}^{k} \mathfrak{a}_{ij} = 1$ $\forall i \in \{2, \ldots, k\}$, we have $\frac{\bar{\Lambda}}{2} < \mathfrak{c}_{1i}$ $\forall i \in \{2, \ldots, k\}$. From Lemma 3.20, we have $|V_{q_t}| \leq 8^{n_x} n_x^{n_x/2} 2^{\frac{\bar{\Lambda}\bar{s}}{2}+\bar{v}}$. From (3.12), we have $l_{ij}(t) = |V_{q_t}|$. We thus use Lemma 3.23, which implies that there exists $\bar{s}$ such that channel rate constraints hold. □

3.A.5 Proof of Theorem 3.11

We follow the same structure for the proof of this theorem as for the proof Theorem 3.9. Let δ^* be chosen smaller than ξ and smaller or equal to δ_6^* from Lemma 3.21. Then for any $\delta : 0 < \delta \leq \delta^*$, the following reasoning can be applied.

Regarding Definition 3.3, (i): This part of the proof is identical to the equivalent part of the proof in Theorem 3.9 with the exception that $r(t)$ is replaced by $\arg\min_{v \in V_{q_{\bar{t}}}} \left\| \sum_{j=1}^{k_m} \frac{v_{m_{ji}(\bar{t})}}{k_m} - v \right\|_P$. For brevity, we thus omit this part of the proof.

Regarding Definition 3.3, (ii): The first half of item (ii) trivially holds from the alphabet function (3.12) and encoder function (3.13). For the second half of item (ii) to hold, the channel rate constraints (3.6) must hold. From

the theorem statement we have $\mathfrak{a}_{ij} = 1$, $\forall i \in \{1, \dots, k_m\}, \forall j \in \{1, \dots, k\}$. From Lemma 3.22 and since $c_i > \frac{k\bar{\Lambda}}{2}$ and $\sum_{j=1}^{k} \mathfrak{a}_{ij} = k$, we have $\frac{\bar{\Lambda}}{2} < \mathfrak{c}_{ij}$, $\forall i \in \{1, \dots, k_m\}, \forall j \in \{1, \dots, k\}$. From the theorem statement we also have $\mathfrak{a}_{ii} = 1$, $\forall i \in \{k_m + 1, \dots, k\}$. Again from Lemma 3.22 and since $c_i > \frac{\bar{\Lambda}}{2}$ and $\sum_{j=1}^{k} \mathfrak{a}_{ij} = 1$ $\forall i \in \{k_m + 1, \dots, k\}$, we have $\frac{\bar{\Lambda}}{2} < \mathfrak{c}_{ii}$ $\forall i \in \{k_m + 1, \dots, k\}$. From Lemma 3.20, we have $|V_{q_t}| \leq 8^{n_x} n_x^{n_x/2} 2^{\frac{\bar{\Lambda}\bar{s}}{2} + \bar{v}}$. From (3.12), we have $l_{ij}(t) = |V_{q_t}|$. We thus use Lemma 3.23, which implies that there exists $\bar{s}$ such that channel rate constraints hold. $\square$

Chapter 4

An Event-Triggered Observation Scheme for Systems with Perturbations and Data Rate Constraints

In this chapter, an event-triggered observation scheme is considered for a perturbed nonlinear dynamical system connected to a remote location via a communication channel, which can only transmit a limited amount of data per unit of time. The dynamical system, which is supposed to be globally Lipschitz, is subject to bounded state perturbations. Moreover, at the system's location, the output is measured with some bounded errors. The objective is to calculate estimates of the state at the remote location in real-time with maximum given error, whilst using the communication channel as little as possible. An event-triggered communication strategy is proposed in order to reduce the average number of communications. An important feature of this strategy is to provide an estimation of the relation between the observation error and the communication rate. The observation scheme's efficiency is demonstrated through simulations of unicycle-type robots.

4.1 Introduction

Efficiency has always played a central role in the field of system dynamics and control. In the past twenty years, with the appearance of wireless technologies, efficiency has gained a new meaning. It is not sufficient enough to observe and control systems optimally. These tasks should in addition be carried out in a way that is efficient in terms of data rates. This quest for efficiency has lead to the birth of an entire sub-field in the domain: control and estimation over data rate constrained communication channels ([Matveev and Savkin, 2009], [Yüksel and

Başar, 2013]). The problems in this sub-field all share some common ingredients: one or several dynamical systems, one or several communication channels, several other devices such as controllers, actuators, sensors, that interact by exchanging messages over these communication channels, and a source of uncertainty. This source of uncertainty can be in the form of noise, perturbations, parametric uncertainty, or deviations in the initial conditions. The incertitude can be understood as information in the sense of Shannon's information theory (see [Shannon, 1948]). This information needs to be transmitted via the communication channels, which are generally limited either in the frequency at which they can send messages or in the number of bits they can transmit per unit of time and can be subject to losses or noise themselves.

The earliest work in this sub-field focused on linear systems, which naturally have a simpler structure. For example, these early results include [Wong and Brockett, 1997], where the problem of state estimation for a stochastic plant over data rate constrained communication channel is investigated, and [Elia and Mitter, 2001], where the problem of stabilization of a linear plant with limited information is considered. Many more results were obtained for linear systems and broad surveys of these results are available in [Nair et al., 2007], [Baillieul and Antsaklis, 2007] and [Andrievsky et al., 2010].

For nonlinear systems, results followed soon after. These include [De Persis, 2003], where the problem of the stabilization of a nonlinear system via a data rate constrained channel is posed, and [Baillieul, 2004] which investigates the data rate requirements for feedback control. Both of these early works assumed a particular structure on the nonlinear system. Results for nonlinear systems with more general structures were obtained in [Nair et al., 2004] and [Liberzon and Hespanha, 2005] which adapted techniques for linear systems from [Nair and Evans, 2003] and [Liberzon, 2003a] to nonlinear ones. We put the emphasis on [Nair et al., 2004] because it is among the first papers to introduce a notion of entropy (in this case, topological feedback entropy) to describe the minimum sufficient data rate to stabilize a system. Several other notions of entropy have since then been used to provide bounds on the sufficient and/or necessary data rates allowing for constrained control and/or observation of unperturbed systems. These results include invariance entropy ([Kawan, 2013]), topological entropy ([Liberzon and Mitra, 2016], [Matveev and Pogromsky, 2016], [Matveev and Savkin, 2009], [Sibai and Mitra, 2018], and [Voortman et al., 2019]), estimation entropy/α-entropy ([Kawan, 2018] and [Sibai and Mitra, 2017]), and restoration entropy ([Matveev and Pogromsky, 2019]). As an alternative to notions of entropy, some works such as [Fradkov et al., 2008a] relied on passivity-based methods to provide bit-rate bounds. The aforementioned results on entropy are limited to unperturbed systems as the entropy of a perturbed system is, generally speaking, infinite and, as such, entropy is not a useful mathematical tool to analyze perturbed systems.

Around the same time, and for similar reasons, another topic appeared in

the world of dynamical systems and control: event-based control. Some of the earliest works on this topic include [Åarzén, 1999], where an event-triggered PID controller is presented and [Åström and Bernhardsson, 1999], where the effects of event-based sampling are compared to periodic ones. An introduction to event-based control can be found in [Heemels et al., 2012] and an overview of sampling-related results in [Hetel et al., 2017].

One possible approach to obtain constructive bounds for the case of sampled-data systems is to rely on LMI-based techniques. Early results making use of this technique include [Fridman et al., 2004] which provides sufficient conditions for the robust sampled-data stabilization of linear systems with delayed input and [Fu and Xie, 2005] which considers several quantized feedback design problems for linear systems. Recently, LMI-based techniques have been employed for nonlinear systems with specific structures such as Lur'e-type systems ([Seifullaev and Fradkov, 2016, Zhang et al., 2017]), nonlinear systems with cone-bounded nonlinearities [Tarbouriech et al., 2017] and with cone-bounded nonlinear inputs ([Moreira et al., 2019]).

In some recent works, both concepts (data rate constraints and event-based) have explicitly been used together for control and observation purposes. Among them are [Han et al., 2015], which uses an event-triggered sensor schedule for remote estimation for a linear system, [Shi et al., 2016], designing a remote estimator for a linear system with unknown exogenous inputs, [Huang et al., 2017], where a remote estimator for a system with an energy harvesting sensor is developed, [Trimpe, 2017], which tackles distributed state estimation with data rate constraints [Xia et al., 2017], which considers networked state estimation with a shared communication medium, [Muehlebach and Trimpe, 2018], where an LMI approach is used for the networked state estimation problem over a shared communication medium, and [Abdelrahim et al., 2019] where output-based stabilization of linear time-invariant systems affected by unknown external disturbances is studied.

Focusing on the observation, the problem statement in this chapter is motivated by the following practical situation: a unicycle-type robot needs to communicate its position and orientation to a remote location by using Wi-Fi, whilst measuring only its position and using limited computation capacities. Because wireless networks can't transmit infinite amounts of data, it is necessary to develop an observation scheme that minimizes the data rate usage. The simplest equations describing a unicycle-type robot are

$$\begin{aligned} \dot{x}_1(t) &= v_l(1+\bar{d}_1(t))\cos(x_3(t)) \\ \dot{x}_2(t) &= v_l(1+\bar{d}_1(t))\sin(x_3(t)) \\ \dot{x}_3(t) &= v_\theta(1+\bar{d}_2(t)) \\ y(t) &= \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} \bar{w}_1(t) \\ \bar{w}_2(t) \end{bmatrix} \end{aligned} \tag{4.1}$$

with $x(t) = [x_1(t)\ x_1(t)\ x_3(t)]^\intercal$ the state-space vector, $y(t)$ the measured output, where $x_1(t) \in \mathbb{R}$, $x_2(t) \in \mathbb{R}$ are the coordinates of the robot in the xy-plane, $x_3(t) \in [0, 2\pi)$ is the angular orientation of the robot, v_l and v_θ are linear and angular velocities respectively, $\bar{d}_1(t)$ and $\bar{d}_2(t)$ are perturbations such that $\bar{d}_{\max} \geq \bar{d}_i(t) \geq \bar{d}_{\min} > -1$ which correspond to an actuation mismatch, and $\bar{w}(t) \in \mathbb{R}^2$ is a measurement error. Note that the bound $\bar{d}_{\min} > -1$ implies that the actual velocity $v_l(1 + \bar{d}_1(t))$ has the same sign as v_l, which is a natural condition stemming from experiments (see [Guerra et al., 2014] and [Guerra et al., 2017] for more details about this formulation).

In this chapter, an event-triggered observation scheme is developed for the remote observation of dynamical systems with Lipschitz nonlinearities, state perturbations, and measurement noise (as in (4.1)) via data rate constrained communication channels. The first original feature of the observation scheme is that the minimum duration between two consecutive messages that are sent via the communication channel can be chosen. The second original feature is that the precision of the estimates can be tuned. The third feature is that the observation scheme functions on an event-triggered basis, by using the knowledge of the remote estimate to only communicate a new estimate when the precision of the current estimate is not sufficient anymore. The combination of these features leads to an observation scheme, which is very efficient in terms of the transmitted number of bits, which is the main contribution of this chapter. This chapter is an extension of [Voortman et al., 2020b]. This chapter, considers continuous-time systems, whereas [Voortman et al., 2020b] studies discrete-time systems. Moreover, [Voortman et al., 2020b] deals with linear systems while this chapter extends the class of systems to Lipschitz-nonlinear ones. An observer specific for unicycle-time robots which utilizes a similar communication protocol as this chapter has been experimentally validated on Turtlebots. The results have been submitted for publication in [Voortman et al., 2021b].

The structure of this work is as follows. First, in Section 4.2 we specify the problem statement. In order to solve this problem, an observation scheme is developed in Section 4.3. In Section 4.4, two results about this observation scheme are exposed. The first one is a proposition that provides a bound on the maximum observation error. The second one is a theorem that evaluates a bound on the so-called "channel transmission capacity" which is sufficient to implement the observation scheme. Finally, in Section 4.5 simulations of the observation scheme are provided for the motivating example (4.1). These simulations illustrate why the observation scheme is particularly efficient and how its different parameters can be tuned to fit the user's preferences in terms of performance.

4.2 Problem Statement

We consider continuous-time systems of the following form

$$\begin{aligned} \dot{x}(t) &= Ax(t) + S\varphi(Hx(t)) + d(t), \\ y(t) &= Cx(t) + w(t), \end{aligned} \tag{4.2}$$

where $x(t) \in \mathbb{R}^n$ is the state, $A \in \mathbb{R}^{n\times n}$, $S \in \mathbb{R}^{n\times p}$, $\varphi : \mathbb{R}^p \to \mathbb{R}^p$ is a vector field, $H \in \mathbb{R}^{p\times n}$, $d(t) \in \mathbb{R}^n$ is an unknown state perturbation, $y(t) \in \mathbb{R}^m$ is the output, $C \in \mathbb{R}^{m\times n}$, and $w(t) \in \mathbb{R}^m$ is a measurement error. We make the following assumptions about the external signals:

Assumption 4.1. *For the state perturbation $d(t)$ and the measurement error $w(t)$,*

$$\|d(t)\|_2 \leq \delta, \ \|w(t)\|_2 \leq \omega, \qquad \forall t \geq 0,$$

where $\|\cdot\|_2$ is the Euclidean norm, δ is the maximum state perturbation, and ω is the maximum measurement error.

We make the following regularity assumption about the right-hand side of (4.2).

Assumption 4.2. *The vector field φ is globally Lipschitz with Lipschitz constant L with respect to the Euclidean norm.*

Note that the previous assumption can be relaxed to local Lipschitz continuity if the perturbations from Assumption 4.1 are such that the solutions of (4.2) are uniformly ultimately bounded ([Khalil, 2002]).

The system is equipped with a smart sensor (a sensor admitting some computational capacities, which allows it to perform additional computations on the measured data) and it is connected to a remote location via a data rate constrained communication channel, which can only send messages that are of finite size. For any time interval $\bar{t}$ between two consecutive transmission, the channel can transmit at most $b^+(\bar{t})$ bits. The objective is to provide estimates $\hat{x}(t)$ of $x(t)$ at the remote location by sending messages over this communication channel. The sensor and the remote location are aware of an initial estimate $\hat{x}(0)$ which verifies

$$\|x(0) - \hat{x}(0)\|_2 \leq \epsilon_0, \tag{4.3}$$

where ϵ_0 is a user-specified parameter corresponding to the error of initial conditions. The reason why (4.3) is assumed, will be discussed further in this chapter in Remark 4.7.

In order to generate the estimates, messages $m(t_j)$, where t_j are the transmission times, are sent. Four ingredients interact with these messages: a sampler $\mathcal{S}$,

a coder $\mathcal{C}$, an alphabet function $\mathcal{A}$, and a decoder $\mathcal{D}$. The four devices together form a communication protocol. The following constants/parameters are known by all devices: the system matrices A and C, the vector field φ, the maximum state perturbation δ, the maximum measurement error ω, the discretization error ϵ (which is induced by coding/decoding operation), and the initial estimate $\hat{x}(0)$ with its accuracy ϵ_0. At the system side, the sampler $\mathcal{S}$ generates the instants of transmission in the following way

$$t_{j+1} = \mathcal{S}(t_j, \{y(s)\}_{t_j \geq s \geq 0}, m(t_1), \ldots, m(t_j)), \tag{4.4}$$

$t_0 = 0$. The coder then generates the messages in the following way

$$m(t_j) = \mathcal{C}(\hat{x}(0), \{y(s)\}_{t_j \geq s \geq 0}, m(t_1), \ldots, m(t_{j-1})), \tag{4.5}$$

$\forall t_j : j > 0$. At each communication instant, the different possible messages are encoded into a finite-sized alphabet (the finiteness being due to the data rate constraints). The alphabet function $\mathcal{A}$ determines the last index of the messages l_j in the following way

$$l_j = \mathcal{A}(\hat{x}(0), m(t_1), \ldots, m(t_{j-1})), \qquad \forall t_j : j > 0. \tag{4.6}$$

The restriction on the choice of messages is then

$$m(t_j) \in \{1, \ldots, l_j\}, \qquad \forall t_j : j > 0.$$

After encoding the messages into sequences of bits, the number of transmitted bits should not exceed the maximum number of bits that can be sent during the communication interval. This implies the following constraint on the alphabet length:

$$\log_2 l_j \leq b^+(t_{j+1} - t_j) \qquad \forall t_j : j > 0. \tag{4.7}$$

At the remote location, the decoder $\mathcal{D}$ receives the messages and interprets them to generate a deterministic estimate of the state $\hat{x}(t)$ in the following way

$$\hat{x}(t) = \mathcal{D}(\hat{x}(0), m(t_1), \ldots, m(t_j)), \qquad \forall t \in [t_j, t_{j+1}), \tag{4.8}$$

$\forall j \geq 0$. Because of the perturbation, measurement error and finite data rate, it is impossible to provide exact estimates at the remote location.

Definition 4.3. *Let (4.2) and (4.8) define respectively the state and its estimate. Then, the quantity*

$$\xi := \sup_{t \geq 0} \|x(t) - \hat{x}(t)\|_2 \tag{4.9}$$

is called the ***maximum observation error****.*

To have $\xi = 0$ would require infinite data rates, as was proven in Theorem 2.3.17 of [Matveev and Savkin, 2009]. We thus instead define the following goals for the chapter:

1. To design the observation scheme (4.4), (4.5), (4.6), and (4.8) such that $\xi < \infty$.

2. To design the observation scheme such that its performance in terms of data rate is better when the perturbations are not the worst-case realizations every time.

3. To investigate the relationship of the time interval between subsequent communications $\bar{t}_j := t_{j+1} - t_j$, the maximum number of bits per time interval $b^+(\cdot)$, and the maximum observation error ξ for the proposed communication scheme.

4.3 Designing the devices

In this section, we introduce the different devices of the communication protocol. The main mechanism can be described as follows: the sensor emulates the dynamics of the remote estimate on the last sent estimate and forwards a new local estimate whenever such value is "far away" from the local measurement. More specifically, at the sensor side, a local observer transforms the output into estimates of the state $\bar{x}(t)$. A copy of the decoder is also simulated by the computational capacity of the sensor so that the sensor knows the current estimate $\hat{x}(t)$ the decoder currently has. This 'copy' of the remote estimate which is provided by the smart sensor will be denoted $\hat{x}_c(t)$. Starting at the initial estimate $\hat{x}(0)$ and in the absence of messages, the decoder computes real-time estimates as solutions of (4.2) without perturbations. When the distance between $\bar{x}(t)$ and $\hat{x}_c(t) = \hat{x}(t)$ becomes larger than the prescribed maximum error (including a margin for the local observation error $\bar{e}(t) := x(t) - \bar{x}(t)$), the sampler decides to communicate and the coder sends a message to the decoder to provide a new estimate $\hat{x}(t)$. Fig. 4.1 depicts how the different elements interact. Below, each of these algorithms is presented in detail.

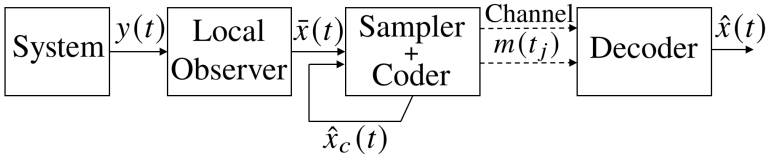

Figure 4.1: Structure of setup.

4.3.1 The Local Observer

The local estimate $\bar{x}(t)$ has the following dynamics

$$\dot{\bar{x}}(t) = A\bar{x}(t) + S\varphi(H\bar{x}(t)) + K(y(t) - C\bar{x}(t)), \tag{4.10}$$

where $K \in \mathbb{R}^{n\times m}$ is a gain matrix. The dynamics of $\bar{e}(t)$ are thus

$$\begin{aligned}\dot{\bar{e}}(t) =&(A + KC)\bar{e}(t) + S(\varphi(Hx(t)) - \varphi(H\bar{x}(t))) + d(t)\\ &+ Kw(t).\end{aligned} \tag{4.11}$$

The local observer uses $\bar{x}(0) = \hat{x}(0)$ as an initial point, which implies that $\|\bar{e}(0)\| \leq \epsilon_0$. The gain K is computed by using the solutions of an LMI program:

$$\begin{aligned}&\arg \min_{M,P,Y,\gamma_i} \quad -\gamma_1 + \delta^2\gamma_2 + \omega^2\gamma_3,\\ &\text{s.t.} \begin{bmatrix} Q_{11} & PS & P & Y \\ S^\intercal P & \text{-}S^\intercal MS & 0_{p\times n} & 0_{p\times m} \\ P & 0_{n\times p} & \text{-}\gamma_2 I_n & 0_{n\times m} \\ Y^\intercal & 0_{m\times p} & 0_{m\times n} & \text{-}\gamma_3 I_m \end{bmatrix} \preceq 0,\\ &P \succ 0,\\ &M \succ 0,\\ &\gamma_4 I_p \succeq S^\intercal MS,\\ &\gamma_1 > 0,\\ &\gamma_2, \gamma_3, \gamma_4 \geq 0,\end{aligned} \tag{4.12}$$

with $Q_{11} = A^\intercal P + C^\intercal Y^\intercal + PA + YC + \gamma_1 I_n + \gamma_4 L^2 H^\intercal H$, and where I_n is the $n\times n$ identity matrix and $0_{n\times m}$ is the $n \times m$ zero matrix. This LMI program is then used to compute K as a function of Y and P, which in turns provides a bound on the norm of the local observation error, as proven in the next proposition.

Proposition 4.4. *Let (4.12) have a solution $(M_*, P_*, Y_*, \gamma_i^*)$ and ϵ_0 be chosen smaller than $\frac{\sqrt{\gamma_2^*}\delta + \sqrt{\gamma_3^*}\omega}{\sqrt{\gamma_1^*}}$. Then by choosing the gain of the local observer as $K = P_*^{-1}Y_*$, local observation error satisfies:*

$$\|\bar{e}(t)\|_2 \leq \eta := \frac{\sqrt{\gamma_2^*}\delta + \sqrt{\gamma_3^*}\omega}{\sqrt{\gamma_1^*}}, \qquad \forall t \geq 0. \tag{4.13}$$

The proof of this proposition is provided in Appendix 4.A.

Remark 4.5. *The objective function of the LMI (4.12) simply aims to minimize the size of the attractive region (by maximizing γ_1^* and minimizing γ_2^* and γ_3^*). This, in turn, minimizes the bound on the local observation error. Ideally, one would want to minimize η but this leads to a - generally intractable - nonlinear matrix inequality problem.*

4.3.2 The Protocol Description

We now present the communication procedure, which we will further reference as Procedure 2. It is composed of a sampler, alphabet function, coder, and decoder as described below. For this particular communication procedure, a minimum time interval between communications is going to be employed. This quantity, denoted as $\bar{t}$, is known by all devices. It is a user-specified parameter, which is to be chosen finite and it directly influences the upper bound on the estimation error. How one might choose $\bar{t}$ and how it influences the error will be discussed further in this chapter.

To properly describe the communication instants, we will need several quantities. The indexes j of the communication instants are inherently known by all devices. The quantity $\bar{j}(t)$ refers to the index of the last instant of communication (initially, $j(0) = 0$). This quantity is always known by the sampler (because it knows how many communication instants it defined), the coder (because it knows how many messages it sent), as well as the decoder (because it knows how many messages it received). In between messages (i.e. for $t \in [t_j, t_{j+1})$), estimates $\hat{x}(t)$ and $\hat{x}_c(t)$ are computed as solutions of the system

$$\dot{\hat{x}}(t) = A\hat{x}(t) + S\varphi(H\hat{x}(t)), \tag{4.14}$$

with the initial conditions $\hat{x}(t_j)$ coming from the messages $m(t_j)$.

Before we can define the communication protocol, a final lemma is necessary. The alphabet relies on the assumption that the estimate $\bar{x}(t)$ will lie within a known set V_j when the communications occur. This assumption guarantees that $\|\bar{x}(t) - \hat{x}(t)\|_2 \leq \epsilon$ after each communication instant, which makes the procedure repeatable. The following lemma proves this property.

Lemma 4.6. *For any $t_j \geq 0$, $x(t_j), \hat{x}(t_j) \in \mathbb{R}^n$, $\epsilon > 0$, $\eta > 0$ such that $\|x(t_j) - \hat{x}(t_j)\|_2 \leq \epsilon + \eta$, and $\bar{t} > 0$, the following holds*

$$\|x(t) - \hat{x}(t)\|_2 \leq e^{\frac{\mu_1^* t}{2}}(\epsilon + \eta) + \delta\sqrt{\mu_2^*}\sqrt{\frac{e^{\mu_1^* t} - 1}{\mu_1^*}}, \tag{4.15}$$

$\forall t \in [t_j, t_j + \bar{t}]$, where $x(t)$ is the solution of (4.2) with $x(t_j)$ as an initial condition, $\hat{x}(t)$ is the solution of (4.14) with $\hat{x}(t_j)$ as an initial condition and

$$\begin{aligned}
&(\mu_1^*, \mu_2^*, \hat{M}_*) := \arg\min_{\mu_i, \hat{M}} \quad \mu_1 + \delta^2\mu_2, \\
&\text{s. t.} \begin{bmatrix} A^\intercal + A - \mu_1 I_n + L^2\mu_3 H^\intercal H & I_n & I_n \\ I_n & -\mu_2 I_n & 0_{n\times n} \\ I_n & 0_{n\times n} & -S^\intercal \hat{M} S \end{bmatrix} \preceq 0, \\
&\mu_3 I_p \geq S^\intercal \hat{M} S, \\
&\mu_i \geq 0.
\end{aligned} \tag{4.16}$$

The proof of this lemma is provided in Appendix 4.B.

Note that (4.16) is an LMI program. Because of its formulation, (4.16) always admits a solution (this can be seen from the fact that μ_i can be chosen arbitrarily large, which makes the first inequality to hold for some μ_i sufficiently large).

Procedure 2.
The Sampler: For all $t \geq t_{\bar{j}(t)} + \bar{t}$, the sampler verifies whether the following condition is satisfied

$$\|\bar{x}(t) - \hat{x}_c(t)\| \leq e^{\frac{\mu_1^* \bar{t}}{2}}(\epsilon + \eta) + \delta\sqrt{\mu_2^*}\sqrt{\frac{e^{\mu_1^* \bar{t}} - 1}{\mu_1^*}} - \eta, \tag{4.17}$$

where μ_1^* and μ_2^* are solutions of (4.16). If the condition is not met, a message must be sent to provide a new estimate. The sampler thus updates $\bar{j}(t)$ and $t_{\bar{j}(t)} = t$ (j increases by one and $\bar{j}(t) = j$).

The Alphabet Function: If $t = t_{\bar{j}(t)}$, the coder and decoder build a covering of the set V_j, where V_j is defined as

$$V_j \coloneqq \left\{ x \in \mathbb{R}^n \;\middle|\; \|x\text{-}\hat{x}(t_j)\|_2 \leq e^{\frac{\mu_1^* \bar{t}}{2}}(\epsilon + \eta) + \delta\sqrt{\mu_2^*}\sqrt{\frac{e^{\mu_1^* \bar{t}}\text{-}1}{\mu_1^*}} \right\}, \tag{4.18}$$

with balls of size ϵ. The balls in the covering are numbered from 1 till $l_{\bar{j}(t)}$, where $l_{\bar{j}(t)}$ is the length of the alphabet and hence the output of the alphabet function.

The Coder: At the communication instants, the coder function finds the index of the ball in the covering whose center is the closest to $\bar{x}(t_j)$ and sends this index over the communication channel. To compute $x_c(t)$, the coder computes the solutions of (4.14) with the center of that ball as an initial condition.

The Decoder: When the decoder receives a message, it computes the solutions of (4.14) with the center of that ball as an initial condition. This solution is then used as the estimate $\hat{x}(t)$.

The strategy to build the alphabet is based on the following idea. As was previously mentioned, in the absence of messages, new estimates are obtained by computing the solutions of (4.14). After receiving a message, the state of the system $x(t)$ is contained in a ball of a certain radius whose center is the estimate $\hat{x}(t)$. In the absence of any messages, this "ball" of uncertainty is gradually deformed into a larger/small uncertain set. The uncertain set evolves due to three factors: first of all, the unknown state perturbation $d(t)$ increases its radius, secondly, the uncertainty set is stretched/compressed by the action of the dynamics of the system (the deformations are proportional to the eigenvalues A), and thirdly, its radius is increased due to the measurement noise. Given that the communication intervals are chosen to be finite, this uncertain set remains of finite size, estimated by Lemma 4.6 with $t = \bar{t}$, in between communications. It can thus be covered by a finite number of balls of size $\epsilon > 0$. The balls in

the covering can be indexed from 1 till $l_{\max} < \infty$. In order to produce such a covering, the only information needed is the initial ball and the different upper bounds on the uncertainties/errors, which implies that both the coder and the decoder can build the set. In order to transmit a new estimate, one can simply send the index of one of the balls whose center then serves as a new estimate with a precision that will depend on ϵ. The cost of communicating in that fashion is dependent on how many balls of size ϵ are required to cover the uncertain set.

Remark 4.7.

- *The coordinates of the centers of the balls used in the covering are always relative to the previous estimate. By communicating in a relative fashion, it is possible to keep the size of the messages limited even if the system is unstable. If the system is unstable, state-space trajectories can drift arbitrarily far away from the origin, which implies that sending an estimate in an absolute fashion (that is, in a coordinate system that is with respect to a fixed point, e.g., the origin), requires to cover all of $\mathbb{R}^n$, which needs infinitely many balls and hence infinitely many bits to be sent. The main drawback of communicating in this fashion is that the channel has to be lossless since the loss of a single message would put the communication protocol to a halt. It is possible to make the communication protocol robust towards losses in the communication channel (see e.g. [Voortman et al., 2019] for more information on communication protocols that are robust towards losses). This option was not explored as robustness towards losses lies outside of the scope of the current work.*

- *Since $\hat{x}_c(t) = \hat{x}(t)$, both devices can build the set I_j according to its definition (4.18). The covering procedure which determines the alphabet is not demanding from a computational point of view since it consists of covering a set that always has the same shape except the whole set is shifted by a certain vector from the origin. Moreover, since this set is centered around the previous estimate, both the coder and decoder can build a covering for it and thus have access to the alphabet.*

- *The existence of $\bar{t}$, the minimum time interval between two consecutive communications, implies that Zeno behaviour is automatically avoided since at least $\bar{t}$ time has to elapse between two triggering instants and $\bar{t}$ is a strictly positive parameter that does not change during the execution of the communication protocol.*

- *The assumption that (4.3) holds is made in order to avoid unnecessarily complicating the communication protocol. If (4.3) does not hold, then an additional initialization step would be required, during which an initial estimate is provided. Because this step does not change the rest of the*

communication protocol and would have little impact on the overall communication rate, it is omitted and replaced by the assumption that (4.3) holds, to facilitate the understandability of the procedure.

4.4 Rate and Errors

With the observation scheme and its devices fully introduced, we now focus on determining what minimum number of bits per time interval is sufficient to implement the observation scheme. The first result we present provides a closed-form expression for maximum observation error ξ.

Proposition 4.8. *The observation scheme described in Procedure 2 ensures that (4.9) holds with*

$$\xi \leq e^{\frac{\mu_1^* \bar{t}}{2}}(\epsilon + \eta) + \delta\sqrt{\mu_2^*}\sqrt{\frac{e^{\mu_1^* \bar{t}} - 1}{\mu_1^*}}. \tag{4.19}$$

Proof: At the times of communication, $\|\hat{x}(t) - \bar{x}(t)\|_2$ is at most ϵ, since $\hat{x}(t)$ is the center of a ball that contains $\bar{x}(t)$. From Proposition 4.4, we have that $\|x(t) - \bar{x}(t)\|_2 \leq \eta$. At the times of communication, we consequently have that $\|x(t) - \hat{x}(t)\|_2 \leq \eta + \epsilon$. We thus apply Lemma 4.6, to obtain that following a communication,

$$\|x(t+s) - \hat{x}(t+s)\|_2 \leq e^{\frac{\mu_1^* \bar{t}}{2}}(\epsilon + \eta) + \delta\sqrt{\mu_2^*}\sqrt{\frac{e^{\mu_1^* \bar{t}} - 1}{\mu_1^*}} \tag{4.20}$$

for $s \in [0, \bar{t}]$. Since after $\bar{t}$, the sampler checks whether the distance is going to be exceeded, and a communication resets the distance to $\epsilon + \eta$ should this bound be reached, then (4.20) holds $\forall t \geq 0$. □

The next result of this section, which is also the main result of the chapter, aims to provide a lower bound on $b^+(\cdot)$ for the designed communication scheme. In the following theorem, the notation $\lceil \cdot \rceil$ refers to the ceiling function (aka the smallest integer that is greater than the argument of the function).

Theorem 4.9. *The observation scheme described in Procedure 2 with $\bar{t} > 0$ is implementable on any channel with*

$$b^+(\bar{t}) \geq n \log_2 \left\lceil \sqrt{n} \frac{e^{\frac{\mu_1^* \bar{t}}{2}}(\epsilon + \eta) + \delta\sqrt{\mu_2^*}\sqrt{\frac{e^{\mu_1^* \bar{t}} - 1}{\mu_1^*}}}{\epsilon} \right\rceil. \tag{4.21}$$

Proof: In order to implement Procedure 2, (4.7) should be verified for all j. The size of the alphabet is equal to the number of balls of radius ϵ required to

cover V_j. Since the radius of the set is

$$e^{\frac{\mu_1^* \bar{t}}{2}}(\epsilon + \eta) + \delta\sqrt{\mu_2^*}\sqrt{\frac{e^{\mu_1^* \bar{t}}\text{-}1}{\mu_1^*}},$$

it can evidently be covered by no more than

$$\left\lceil \frac{e^{\frac{\mu_1^* \bar{t}}{2}}(\epsilon + \eta) + \delta\sqrt{\mu_2^*}\sqrt{\frac{e^{\mu_1^* \bar{t}}\text{-}1}{\mu_1^*}}}{\frac{\epsilon}{\sqrt{n}}} \right\rceil^n$$

hypercubes of radius $\frac{\epsilon}{\sqrt{n}}$. These hypercubes are themselves contained in a sphere of radius ϵ. In order to verify (4.7), it thus suffices to take the $\log_2$ of this last quantity, which leads to (4.21) and completes the proof.

Remark 4.10. *As it will be demonstrated in the simulations section, both error bounds presented in this section may be conservative. This is due to several factors:*

1. *In the LMI formulation, the Lipschitz nonlinearity is modelled as a perturbation (that has to be compensated). For systems where the Lipschitz nonlinear term plays the most important part in the dynamics, this can be a source of conservatism.*

2. *The proof of the observation error relies on a Lyapunov-like function. As is always the case with Lyapunov functions, better functions can lead to tighter error bounds.*

3. *The objective function of (4.12) is suboptimal in the sense that it is a linear formulation that minimizes the size of the convergence region, which depends on a quotient of γ_i's, which is an inherently nonlinear function.*

4. *The covering of the set I_j is simple but not optimal. A better covering procedure would result in fewer balls being necessary and would thus improve the bound of Theorem 4.9.*

Regarding point *(3)* of the above remark, one could alternatively take the objective function $\gamma_1^{-1} + \delta^2\gamma_2 + \omega^2\gamma_3$ and use Schur's lemma to transform (4.12) into an LMI again (this approach is used in e.g. [Moreira et al., 2019]). After trying both objective functions out on the simulations which will be presented in the next section, the authors notice no improvement compared with the previous objective function. This different formulation thus wasn't used.

Proposition 4.4 and Theorem 4.9 can be used to choose ϵ and $\bar{t}$. The choice of ϵ provides a trade-off between the error (Proposition 4.4) and the requirements on $b^+(\cdot)$ (Theorem 4.9). By choosing ϵ small, the maximum observation error is small but the maximum number of bits that can be sent becomes large and vice

versa. Regarding $\bar{t}$, the smaller it is chosen, the lower the error and requirement on $b^+(\cdot)$ but the higher the frequency at which messages are sent, hence, the more often the communication channel is used.

4.5 Simulations

In this section, we apply the previously developed observation scheme to the motivating example that was presented in the introduction: a unicycle-type robot with data rate constraints. The goals of this section are

- To illustrate the validity of the theoretical upper bound (4.21), but also to show that, in a real situation, our observation strategy can be much more efficient than expected;
- To illustrate how the choices of ϵ and $\bar{t}$ affect the number of communications;
- To show that with an improved local observer, the performance of the observation scheme becomes much better.

The unicycle-type model that we consider is of the form (4.1) with $v_l = 0.15$, $\left\|[\bar{d}_1(t)\ \bar{d}_2(t)]^\intercal\right\|_2 \leq 0.1$, $\|[\bar{w}_1(t)\ \bar{w}_2(t)]^\intercal\|_2 \leq 0.05$, and $v_\theta = 0.2$. We rewrite this system such that it fits in the form (4.2):

$$\begin{aligned} \dot{x}(t) &= \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & v_l \\ 0 & 0 & 0 \end{bmatrix}}_{A} x(t) + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{S} \varphi(Hx(t)) + d(t), \\ y(t) &= \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}_{C} x(t) + w(t) \end{aligned} \tag{4.22}$$

with

$$H = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \varphi : \mathbb{R}^3 \to \mathbb{R}^3 : \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \to \begin{bmatrix} v_l \cos(r_1) \\ v_l \sin(r_1) - v_l \\ v_\theta \end{bmatrix},$$
$$d(t) = \begin{bmatrix} v_l d_1(t) \cos(x_3(t)) \\ v_l d_1(t) \sin(x_3(t)) \\ v_\theta d_2(t) \end{bmatrix}, w(t) = \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}.$$

Assumption 4.1 holds with $\delta = 0.1$, $v_l = 0.015$ and $\omega = 0.05$. In general, Assumption 4.2 holds with $L = 2v_l$. However, (4.12) is not feasible with $L = 2v_l$. We can adapt the formulation and solve this issue by using the fact that if the observation error is small, the Lipschitz constant is smaller than $2v_l$ (for

$\|x(t) - \bar{x}(t)\|_2 \leq 1$, $L \approx v_l \|x(t) - \bar{x}(t)\|_2$). In order to solve the LMI, we thus follow the following steps:

1. Pose $L = \bar{L}$, where $\bar{L}$ is some arbitrary initial value to be used in solving of the LMI (4.12);

2. Compute γ_i and K by solving the LMI (4.12) with $L = \bar{L}$;

3. Compute η from equation (4.13) and find Lipschitz constant L over the interval $[-\eta, \eta]$;

4. If the Lipschitz constant that we compute is smaller than $\bar{L}$, the problem is feasible, if not, we adapt $\bar{L}$ to a different value and start at step (1) again.

We begin by computing the gain of the local observer K by solving the LMI program (4.12). For the matrix A as in (4.22), following the aforementioned steps, we pose $L = \bar{L} = 0.38$ and obtain

$$K = \begin{bmatrix} -0.2447 & 0 \\ 0 & -1.2581 \\ 0 & -1.8184 \end{bmatrix}.$$

To solve the LMI, the MATLAB package YALMIP ([Lofberg, 2004]) was used, together with the MOSEK solver ([MOSEK ApS, 2019]).

With these values, we have $\eta = 0.370964$ and for that range, the Lipschitz constant of the system over $[-0.370964, 0.370964]$ is smaller than 0.38. The value for $\bar{L}$ is thus validated. We then used a Monte Carlo method with 10000 different simulations of the communication scheme from $t = 0$ till $t = 100$ with initial conditions $(0, 0, 0)$, an initial estimate randomly chosen in a ball of radius η centred around the origin, and random perturbations. In order to compare the actual number of transmitted bits with the theoretical maximum number of bits and to show the influence of the choices of ϵ and $\bar{t}$, we run several simulations. First, we set $\bar{t} = 0.1$ and simulate for various choices of ϵ. In Table 4.1, we expose the results of the experiments in term of maximum observation error ξ, number of communications N_{com} and number of bits per communication N_{bits}. We make several observations

1. The minimum time interval between two communications is set to 0.1, which implies that we could theoretically communicate 1000 times in 100 seconds. Clearly, the actual number of communications is much lower than that since it reaches 14.39 when the most precise estimates are sent. In that case, 8 bits need to be sent every time we communicate, on average every 7 seconds. This shows that the scheme is much more efficient in terms of the actual number of transmitted bits per unit of time, compared to the theoretical sufficient maximum number of bits.

2. As the precision of the estimates increases, the total error decreases, but the number of transmitted bits increases (more balls are required to cover V_j). There is thus a trade-off between precision and the number of transmitted bits. Since ξ is smaller, we also communicate more often.

3. Decreasing ϵ can only affect the precision up to a certain limit, which is largely dictated by the precision local observer. Since the bound on the local observation error is $\eta = 0.370964$, it is impossible to reach this bound for ξ. Moreover, even approaching this bound requires us to decreases ϵ drastically.

For the next set of simulations, we use only one value of $\epsilon = 0.05$ and simulate the observation scheme for various choices of $\bar{t}$. The results are displayed in Table 4.2. We make the following observations about the results:

1. Increasing $\bar{t}$ also increases ξ, which is logical: communicating more often leads to better precision.

2. As $\bar{t}$ increases, the average time between two consecutive communications drastically increases as well. Even with $\bar{t} = 0.5$, we communicate on average less than once every 100 time instants. This is mostly due to the fact that the state perturbations and measurement noise are relatively low.

3. For almost all choices of $\bar{t}$, the number of bits that need to be transmitted is the same. This is due to the fact that the number of bits is rounded upwards. The unrounded number of bits does increase as $\bar{t}$ increases.

4. The limiting effect of the local observation error is again present. Even when choosing $\bar{t} = 0.01$, which corresponds to a sampling time of 10 ms, ξ remains large.

ϵ	0.1	0.05	0.02	0.01	0.005
ξ	0.5128	0.4625	0.4323	0.4222	0.4172
N_{com}	2.288	5.606	9.993	12.635	14.392
N_{bits}	4	5	6	7	8

Table 4.1: Results for various ϵ with $\bar{t} = 0.1$

In both sets of simulations, the main limiting factor is that the local observer has a limited precision ($\eta = 0.371$) and this greatly influences the total error. By using a better local observer (e.g., a nonlinear observer specifically designed for unicycle-type robots), the performance of the observation scheme would be improved. In order to illustrate this fact, we consider the situation where all

$\bar{t}$	0.01	0.1	0.2	0.5	1	2
ξ	0.433	0.462	0.485	0.522	0.568	0.639
N_{com}	11.719	5.60	1.660	0.493	0.108	0.008
N_{bits}	4	5	5	5	5	5

Table 4.2: Results for various $\bar{t}$ with $\epsilon = 0.05$

states of the unicycle robot are observed locally and a local observer is not necessary. We thus consider the following system

$$\begin{aligned} \dot{x}(t) &= \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & v_l \\ 0 & 0 & 0 \end{bmatrix}}_{A} x(t) + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{S} \varphi(Hx(t)) + d(t), \\ y(t) &= x(t) + w(t). \end{aligned} \tag{4.23}$$

with

$$H = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \varphi : \mathbb{R}^3 \to \mathbb{R}^3 \qquad : \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \to \begin{bmatrix} v_l \cos(r_1) \\ v_l \sin(r_1) - v_l \\ v_\theta \end{bmatrix},$$

$v_l = 0.1$, $v_\theta = 0.2$, $\delta = 0.099$ and $\omega = 0.05$. This value for δ implies that very large perturbations are possible in the state. In particular, the linear velocity of the robot ranges from 0.001 m/s to 0.199 m/s. Since all states are observed, $\eta = \omega = 0.05$. We again used a Monte-Carlo method with 1000 different simulations of the observation scheme for 100 seconds.

Fig. 4.2 shows how the actual observation error evolves over time, together with $\|\bar{x}(t) - \hat{x}(t)\|_2$, which is used for the triggering condition. Note that the distance between both horizontal lines is equal to $\omega = 0.1$. We observe that communication is triggered each time the triggering condition is met. The observation error always remains below ξ. There is some conservatism in the triggering condition, the reasons for which have been discussed in Remark 4.10. Towards the end, it can be seen that the actual observation error is larger than $\|\bar{x}(t) - \hat{x}(t)\|_2$ and the triggering condition, which is due to the measurement noise, and also proves that the protocol is not too conservative.

Fig. 4.3 shows the trajectory of the unicycle robot in the x_1-x_2 plane for one particular simulation with $\epsilon = 0.01$ and $\bar{t} = 0.5$. We can see that the observation scheme follows the actual trajectory of the system but, due to the state perturbations and the measurement noise, the observation scheme regularly resets to a point close to the current local estimate. The full results of the Monte-Carlo simulations for relevant pairs of ϵ and δ are displayed in Table 4.3. We make the following observations about these results.

1. Although large state perturbations are present, the observation scheme is still more efficient than the theoretical maximum. Even in the case of $\epsilon = \bar{t} = 0.01$, we only communicate 477 times on average, as opposed to the theoretical maximum of 10000 times.

2. The effect of ϵ and $\bar{t}$ on ξ and N_{com} are similar as in the previous example. It is always possible to trade precision for the number of communicated bits and vice-versa.

3. The choice of $\bar{t}$ has more impact in the error, as well as the average number of communications than ϵ.

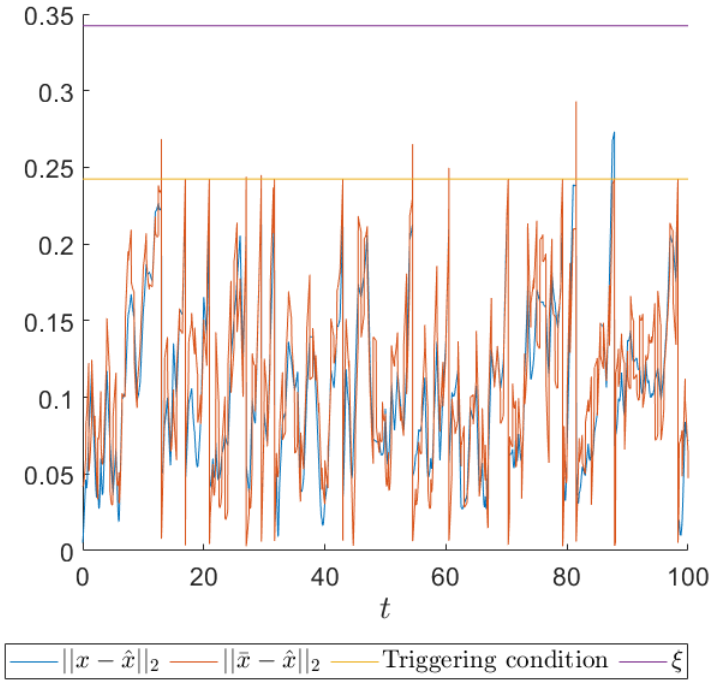

Figure 4.2: Evolution over time of the observation error and quantity used in the triggering function, together with the respective bounds for one particular simulation with full state measurement, $\omega = 0.1$, $\epsilon = 0.01$, and $\bar{t} = 0.5$.

4.6 Conclusion

In this chapter we presented an event-triggered, data rate constrained observation scheme for continuous-time linear systems with perturbations. After posing

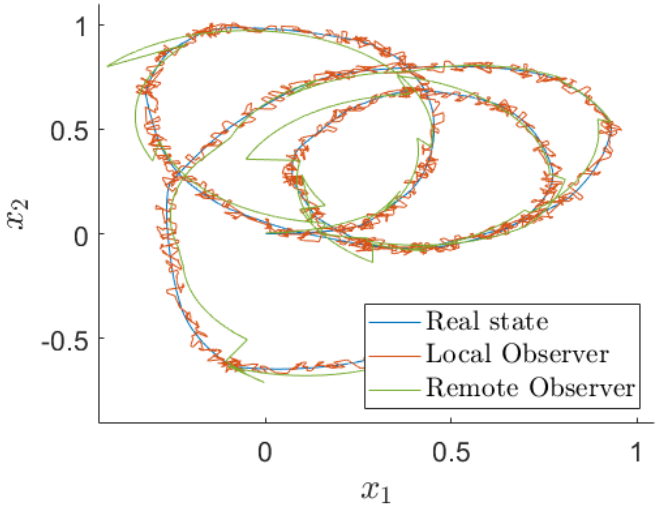

Figure 4.3: State-space trajectory of the unicycle robot in the x_1-x_2 plane for one particular simulation with $v_l = 0.1$, $v_\theta = 0.2$, $\epsilon = 0.01$, and $\bar{t} = 0.5$.

ϵ	0.05		0.01		
$\bar{t}$	0.5	0.1	0.5	0.05	0.01
ξ	0.331	0.200	0.290	0.130	0.0915
N_{com}	16.15	47.97	21.12	120.26	477.517
N_{bits}	4	3	6	5	4

Table 4.3: Results for different choices of ϵ and $\bar{t}$ with locally fully observed states.

the problem statement, the design of the devices that form the communication scheme was explained. A theorem evaluating the upper bound for the minimum bit-rate required to implement this communication protocol was then presented. The protocol was tested via simulations of unicycle-type robots. Through these simulations, the following properties of the observation scheme have been highlighted.

- On average, it was observed in simulations that the number of communications is much lower than the theoretical maximum. This is due to some conservatism in the estimates as well as the fact that the observation scheme functions on an event-triggered basis;

- By properly choosing the parameters of the observation scheme, it is possible to trade-off accuracy for a lower number of bits sent and vice-versa;
- One of the limiting factors of the observation scheme is the usage of a generic local observer. The accuracy of the local observer greatly influences the general precision of the data rate constrained observation scheme.

The continuation of this work includes:

- Using a local observer more adapted to the structure of the system to decrease the total observation error and hence the maximum observation error;
- Using the data rate constrained observation scheme on several mobile robots;
- Adapting the observation scheme to a larger class of nonlinear systems.

Appendices

4.A Proof of Proposition 4.4

Proof: For brevity, we will drop the dependency on time in the notations of this proof (e.g., $\bar{e}$ refers to $\bar{e}(t)$). We start by defining the Lyapunov function $V(\bar{e}) = \bar{e}^\intercal P_* \bar{e}$. The derivative with respect to time of this function is

$$\dot{V}(\bar{e}) = \dot{\bar{e}}^\intercal P_* \bar{e} + \bar{e}^\intercal P_* \dot{\bar{e}}.$$

From (4.11), we have

$$\begin{aligned}\dot{V}(\bar{e}) = &((A+KC)\bar{e} + S(\varphi(Hx) - \varphi(H\bar{x})) + d + Kw)^\intercal P_* \bar{e} \\ &+ \bar{e}^\intercal P_* ((A+KC)\bar{e} + S(H\varphi(x) - \varphi(H\bar{x})) + d + Kw).\end{aligned}$$

We add and subtract the following terms to the right-hand side of the previous equation: $\gamma_1^* \bar{e}^\intercal \bar{e}$, $\gamma_4^* L^2 \bar{e}^\intercal H^\intercal H \bar{e}$, $(\varphi(x) - \varphi(\bar{x}))^\intercal S^\intercal M_* S(\varphi(Hx) - \varphi(H\bar{x}))$, $\gamma_2^* d^\intercal d$, and $\gamma_3^* w^\intercal w$. We then have

$$\begin{aligned}&\dot{V}(\bar{e}) = \bar{e}^\intercal (A+KC)^\intercal P_* \bar{e} + \bar{e}^\intercal P_* (A+KC)\bar{e} + \gamma_4^* L^2 \bar{e}^\intercal H^\intercal H \bar{e} \\ &+ (\varphi(Hx) - \varphi(H\bar{x}))^\intercal S^\intercal P_* \bar{e} + \bar{e}^\intercal P_* S(\varphi(Hx) - \varphi(H\bar{x})) \\ &- (\varphi(Hx) - \varphi(H\bar{x}))^\intercal S^\intercal M_* S(\varphi(Hx) - \varphi(H\bar{x})) + d^\intercal P_* \bar{e} \\ &+ \bar{e}^\intercal P_* d - \gamma_2^* d^\intercal d + w^\intercal K^\intercal P_* \bar{e} + \bar{e}^\intercal P_* K w - \gamma_3^* w^\intercal w + \gamma_1^* \bar{e}^\intercal \bar{e} \\ &- \gamma_1^* \bar{e}^\intercal \bar{e} - \gamma_4^* L^2 \bar{e}^\intercal H^\intercal H \bar{e} + \gamma_2^* d^\intercal d + \gamma_3^* w^\intercal w \\ &+ (\varphi(Hx) - \varphi(H\bar{x}))^\intercal S^\intercal M_* S(\varphi(Hx) - \varphi(H\bar{x})).\end{aligned}$$

This can be rewritten as

$$\begin{aligned}\dot{V}(\bar{e}) =& z^\intercal Q z - \gamma_1^* \bar{e}^\intercal \bar{e} - \gamma_4^* L^2 \bar{e}^\intercal H^\intercal H \bar{e} + \gamma_2^* d^\intercal d + \gamma_3^* w^\intercal w \\ &+ (\varphi(Hx) - \varphi(H\bar{x}))^\intercal S^\intercal M_* S(\varphi(Hx) - \varphi(H\bar{x})),\end{aligned}$$

where $z = [\bar{e}^\intercal \ (\varphi(x) - \varphi(\bar{x}))^\intercal \ d^\intercal \ w^\intercal]^\intercal$ and

$$Q = \begin{bmatrix} Q_{11} & P_* S & P_* & P_* K \\ S^\intercal P_* & -S^\intercal M_* S & 0_{p\times n} & 0_{p\times m} \\ P_* & 0_{n\times p} & -\gamma_2^* I_n & 0_{n\times m} \\ K^\intercal P_* & 0_{m\times p} & 0_{m\times n} & -\gamma_3^* I_m \end{bmatrix},$$

with $Q_{11} = (A+KC)^\intercal P_* + P_*(A+KC) + \gamma_1^* I_n + \gamma_4^* L^2 H^\intercal H$. Clearly, (4.12) implies that $Q \preceq 0$, and we thus have that

$$\begin{aligned}\dot{V}(\bar{e}) \leq& -\gamma_1^* \bar{e}^\intercal \bar{e} - \gamma_4^* L^2 \bar{e}^\intercal H^\intercal H \bar{e} + \gamma_2^* d^\intercal d + \gamma_3^* w^\intercal w \\ &+ (\varphi(Hx) - \varphi(H\bar{x}))^\intercal S^\intercal M_* S(\varphi(Hx) - \varphi(H\bar{x})).\end{aligned}$$

From (4.12), $\gamma_4^* I_p \succeq S^\intercal M_* S$, which implies that

$$\begin{aligned}\dot{V}(\bar{e}) \leq& -\gamma_1^* \bar{e}^\intercal \bar{e} - \gamma_4^* L^2 \bar{e}^\intercal H^\intercal H \bar{e} + \gamma_2^* d^\intercal d + \gamma_3^* w^\intercal w \\ &+ \gamma_4^* (\varphi(Hx) - \varphi(H\bar{x}))^\intercal (\varphi(Hx) - \varphi(H\bar{x})).\end{aligned}$$

Assumption 4.2 implies that

$$(\varphi(Hx) - \varphi(H\bar{x}))^\intercal (\varphi(Hx) - \varphi(H\bar{x})) \leq L^2 \bar{e}^\intercal H^\intercal H \bar{e}$$

and thus

$$\dot{V}(\bar{e}) \leq -\gamma_1^* \bar{e}^\intercal \bar{e} + \gamma_2^* d^\intercal d + \gamma_3^* w^\intercal w.$$

By traditional Lyapunov arguments, we thus have

$$\gamma_1^* \bar{e}^\intercal \bar{e} \leq \gamma_2^* d^\intercal d + \gamma_3^* w^\intercal w,$$

which implies that

$$\sqrt{\gamma_1^*}\, \|\bar{e}\|_2 \leq \sqrt{\gamma_2^*}\, \|d\|_2 + \sqrt{\gamma_3^*}\, \|w\|_2 .$$

By using Assumption 4.1, we thus have the following globally attractive region

$$\|\bar{e}\|_2 \leq \frac{\sqrt{\gamma_2^*}\delta + \sqrt{\gamma_3^*}\omega}{\sqrt{\gamma_1^*}}.$$

By choosing ϵ_0 smaller than $\frac{\sqrt{\gamma_2^*}\delta + \sqrt{\gamma_3^*}\omega}{\sqrt{\gamma_1^*}}$, we guarantee that the $\bar{e}(0)$ starts within that region and hence, (4.13) holds for all $t \geq 0$. □.

4.B Proof of Lemma 4.6

Proof: Starting from t_j, the error $\hat{e}(t) \coloneqq x(t) - \hat{x}(t)$ has the following dynamics.

$$\begin{aligned}\dot{\hat{e}}(t) &= \dot{x}(t) - \dot{\hat{x}}(t)\\ &= Ax(t) + d(t) + S\varphi(Hx(t)) - A\hat{x}(t) + S\varphi(H\hat{x}(t))\\ &= A\hat{e}(t) + d(t) + S(\varphi(Hx(t)) - \varphi(H\hat{x}(t))).\end{aligned} \tag{4.24}$$

We define the Lyapunov function $\hat{V}(\hat{e}(t)) \coloneqq \hat{e}(t)^\intercal \hat{e}(t)$. The derivative with respect to time of this function is

$$\begin{aligned}\dot{\hat{V}}(\hat{e}(t)) =& \dot{\hat{e}}(t)^\intercal \hat{e} + \hat{e}^\intercal \dot{\hat{e}}(t)\\ =& (\hat{e}(t)^\intercal A^\intercal + d(t)^\intercal + \varphi(Hx(t))^\intercal S^\intercal - \varphi(H\hat{x})^\intercal S^\intercal)\hat{e}(t)\\ &+ \hat{e}(t)^\intercal (A\hat{e}(t) + d(t) + S\varphi(Hx(t)) - S\varphi(H\hat{x}(t)))\end{aligned}$$

We add and subtract the following terms to the right-hand-side of the previous equation: $\mu_1^* \hat{e}(t)^\intercal \hat{e}(t)$, $\mu_2^* d(t)^\intercal d(t)$, $(\varphi(Hx(t)) - \varphi(H\hat{x}(t)))^\intercal S^\intercal \hat{M}_* S(\varphi(Hx(t)) - \varphi(H\hat{x}(t)))$, and $L^2 \mu_3^* \hat{e}(t)^\intercal H^\intercal H \hat{e}(t)$, where $\hat{M}_*$ is the solution of (4.16).

$$\begin{aligned}&\dot{\hat{V}}(\hat{e}(t)) = \hat{e}(t)^\intercal (A^\intercal + A)\hat{e}(t) - \mu_1^* \hat{e}(t)^\intercal \hat{e}(t) + d(t)^\intercal \hat{e}(t)\\ &+ \hat{e}(t)^\intercal d(t) - \mu_2^* d(t)^\intercal d(t) + (\varphi(Hx(t))^\intercal - \varphi(H\hat{x})^\intercal) S^\intercal \hat{e}(t)\\ &+ \hat{e}(t)^\intercal S(\varphi(Hx(t)) - \varphi(H\hat{x}(t))) + L^2 \mu_3^* \hat{e}(t)^\intercal H^\intercal H \hat{e}(t)\\ &- (\varphi(Hx(t)) - \varphi(H\hat{x}(t)))^\intercal S^\intercal \hat{M}_* S(\varphi(Hx(t)) - \varphi(H\hat{x}(t)))\\ &+ \mu_1^* \hat{e}(t)^\intercal \hat{e}(t) + \mu_2^* d(t)^\intercal d(t) - L^2 \mu_3^* \hat{e}(t)^\intercal H^\intercal H \hat{e}(t)\\ &+ (\varphi(Hx(t)) - \varphi(H\hat{x}(t)))^\intercal S^\intercal \hat{M}_* S(\varphi(Hx(t)) - \varphi(H\hat{x}(t)))\end{aligned}$$

By defining $\hat{z}(t) = [\hat{e}(t)^\intercal \ d(t)^\intercal \ \varphi(Hx(t)) - \varphi(H\hat{x}(t))]^\intercal$, the previous equation can be rewritten as

$$\begin{aligned}&\dot{\hat{V}}(\hat{e}(t)) = \hat{z}(t)^\intercal \hat{Q} \hat{z}(t) + \mu_1^* \hat{e}(t)^\intercal \hat{e}(t) + \mu_2^* d(t)^\intercal d(t)\\ &+ (\varphi(Hx(t)) - \varphi(H\hat{x}(t)))^\intercal S^\intercal \hat{M}_* S(\varphi(Hx(t)) - \varphi(H\hat{x}(t)))\\ &- L^2 \mu_3^* \hat{e}(t)^\intercal H^\intercal H \hat{e}(t).\end{aligned}$$

with

$$\hat{Q} = \begin{bmatrix} A^\intercal + A - \mu_1^* I_n + L^2 \mu_3^* H^\intercal H & I_n & I_n \\ I_n & -\mu_2^* I_n & 0_{n\times n} \\ I_n & 0_{n\times n} & -S^\intercal \hat{M}_* S \end{bmatrix}$$

Evidently, from (4.16), we have $\hat{z}(t)^\intercal \hat{Q} \hat{z}(t) \leq 0$ and thus

$$\begin{aligned}&\dot{\hat{V}}(\hat{e}(t)) \leq \mu_1^* \hat{e}(t)^\intercal \hat{e}(t) + \mu_2^* d(t)^\intercal d(t)\\ &+ (\varphi(Hx(t)) - \varphi(H\hat{x}(t)))^\intercal S^\intercal \hat{M}_* S(\varphi(Hx(t)) - \varphi(H\hat{x}(t)))\\ &- L^2 \mu_3^* \hat{e}(t)^\intercal H^\intercal H \hat{e}(t).\end{aligned}$$

Since $\mu_3^* I_p \geq S^\intercal \hat{M}_* S$, we have

$$\begin{aligned}\dot{\hat{V}}(\hat{e}(t)) \leq\ & \mu_1^* \hat{e}(t)^\intercal \hat{e}(t) + \mu_2^* d(t)^\intercal d(t) \\ & + \mu_3^* (\varphi(Hx(t)) - \varphi(H\hat{x}(t)))^\intercal (\varphi(Hx(t)) - \varphi(H\hat{x}(t))) \\ & - L^2 \mu_3^* \hat{e}(t)^\intercal H^\intercal H \hat{e}(t).\end{aligned}$$

By using Assumptions 4.1 and 4.2, this can further be transformed to

$$\begin{aligned}\dot{\hat{V}}(\hat{e}(t)) \leq & \mu_1^* \hat{e}(t)^\intercal \hat{e}(t) + \mu_2^* \delta^2 + L^2 \mu_3^* \hat{e}(t)^\intercal H^\intercal H \hat{e}(t) \\ & - L^2 \mu_3^* \hat{e}(t)^\intercal H^\intercal H \hat{e}(t).\end{aligned}$$

and thus

$$\dot{\hat{V}}(\hat{e}(t)) \leq \mu_1^* \hat{V}(\hat{e}(t)) + \mu_2^* \delta^2.$$

This implies that

$$\hat{V}(\hat{e}(t)) \leq e^{\mu_1^* t} \hat{V}(e(t_j)) + \mu_2^* \delta^2 \frac{e^{\mu_1^* t} - 1}{\mu_1}.$$

From the definition of $\hat{V}$, and by taking a square root on both sides of the equations, this yields the following inequality

$$\|\hat{e}(t)\|_2 \leq e^{\frac{\mu_1^* t}{2}} (\epsilon + \eta) + \delta \sqrt{\mu_2^*} \sqrt{\frac{e^{\mu_1^* t} - 1}{\mu_1^*}}.$$

□

Chapter 5

Remote State Estimation of Steered Systems with Limited Communications: an Event-Triggered Approach

In this chapter, an approach is proposed for the remote observation of a dynamical system through a data-rate constrained communication channel. The focus is put on discrete-time systems with a Lipschitz nonlinearity, driven by an external signal, and subject to bounded state perturbation and measurement error. The problem at hand is providing estimates of system's state at a remote location, which is connected via a channel which can only sent limited numbers of bits per unit of time. A solution, named observation scheme, is proposed in the form of several interacting agents. This solution is designed such that the maximum observation error is upper-bounded by a computable quantity dependent on system constants and selectable parameters. The scheme is designed in an event-triggered fashion, such that the actual communication rate is sometimes much lower than the theoretically evaluated maximum one, as it is demonstrated through simulations.

5.1 Introduction

Whether it is collective cruise control of connected cars, formation control for drones through Wi-Fi or Bluetooth, connected smart sensors, or another form of cyber-physical system technology, wireless communications are omnipresent in the modern industry. Since it is a booming application domain, new performance requirements are imposed, and the field of dynamics and control has to come up with new solutions to deal with these new problems and challenges. The problems related to the interactions between dynamical systems and com-

munication technologies are numerous, and have led to the creation of an entire sub-field within the topic: control and estimation over communication channels. All problems share some common features: one or several dynamical systems, sometimes paired with sensors, actuators and controllers, are placed at remote locations from one another. To communicate, they have to employ communication channels which are limited, either in terms of frequency of communications, size of the messages, or are subject to losses and lags. When this setting is combined with noise, parametric uncertainty, perturbations, or deviations in initial conditions, there is a need to design efficient communication strategies to deal with these uncertainties, which can be understood as additional sources of information in the sense of Shannon's information theory [Shannon, 1948].

Among the earliest works in this sub-field, one finds: [Wong and Brockett, 1997], which considered state estimation under data-rate constraints for linear noisy systems and [Elia and Mitter, 2001], which considered stabilization of a linear system under quantized state feedback. Many more results on observation, state estimation and control have been obtained for linear systems and broad overviews of these results can be found in [Nair et al., 2007], [Baillieul and Antsaklis, 2007] and [Andrievsky et al., 2010].

For nonlinear systems, important results are [Nair et al., 2004], [Liberzon and Hespanha, 2005] and [Savkin, 2006], which are based on entropy concepts. Since those three papers, many different notions of entropy have been used to provide bounds on the sufficient and/or necessary data-rates to observe and/or control nonlinear dynamical systems over constrained channels (see [Kawan, 2013], [Matveev and Savkin, 2009], [Kawan, 2018], [Sibai and Mitra, 2017], [Liberzon and Mitra, 2016], [Matveev and Pogromsky, 2016], [Sibai and Mitra, 2018], [Voortman et al., 2019] and [Matveev and Pogromsky, 2019]).

Approximately at the same time as the research on control with data-rate constraints started, the topic of event-triggered control appeared as well in the dynamics and control community. Two early papers are: [Åarzén, 1999], where an event-triggered PID controller is presented and [Åström and Bernhardsson, 1999], where the effects of event-based sampling are compared to periodic sampling. For an introduction to event-based control, one can refer to [Heemels et al., 2012]. For an overview of many sampling-related results, one can refer to [Hetel et al., 2017].

Both control with data-rate constraints and event-based control have been used together for control and observation purposes. These works include [Han et al., 2015], which uses an event-triggered sensor schedule for remote estimation for a linear system, [Shi et al., 2016], designing a remote estimator for a linear system with unknown exogenous inputs, [Huang et al., 2017], where a remote estimator for a system with an energy harvesting sensor is developed, [Trimpe, 2017], which tackles distributed state estimation with data-rate constraints, [Xia et al., 2017], which considers networked state estimation with a shared communication medium and, [Muehlebach and Trimpe, 2018], where an LMI approach

is used for the networked state estimation problem over a shared communication medium.

A particular class of dynamical systems is the class of Lipschitz-nonlinear systems which are typically found when modeling mechanical systems. This class includes systems with trigonometric nonlinearities, which are globally Lipschitz. Square or cubic nonlinearities which are also sometimes encountered with mechanical systems are locally Lipschitz and, since they often occur on physical systems, they are often paired with a saturation function (due to having restricted movement in space), which makes them globally Lipschitz. Many results have been obtained for Lipschitz-nonlinear systems, among which we note [Raghavan and Hedrick, 1994] and [Rajamani, 1998] which both develop observers for Lipschitz-noninear systems.

In this chapter, we develop a communication scheme to remotely observe a discrete-time Lipschitz-nonlinear system with state perturbations and bounded measurement error. The system is connected to a remote location by means of a data-rate constrained communication channel. It is also steered by a driving signal which is not measured at the remote location. The challenge is to design the communication protocol such that through the messages that are received from the system, it is possible to reconstruct estimates of the state at the remote location. Moreover, this should be achieved while using limited communication data-rates. The novelty of this chapter in comparison with the aforementioned works, is that we consider an event-trigger communication protocol which often requires much less than the theoretical maximum channel rate.

The chapter is structured as follows. In Section 5.3, we expose the details of the problem statement. Next, in Section 5.4, we develop the communication scheme. Section 5.5 is then dedicated to analytical bounds on the maximum observation error and communication rate. Finally, we conclude with simulations in Section 5.6, to insight on how the communication scheme functions and performs in practice.

5.2 Notations

- I_n: an $n \times n$ identity matrix;
- $\star$: a zero matrix of appropriate dimension;
- $\|v\|_2$, v is a vector: the Euclidean norm;
- $\|M\|_2$, M is a matrix: the operator norm induced by the pair $(\|\cdot\|_2, \|\cdot\|_2)$;
- $\sigma_i(M)$, M is a matrix: the singular values of M ranked in non-increasing order ($\sigma_1(M) = \|M\|_2$);
- $\mathrm{vec}\,(M)$, M is a matrix: the vectorization of M;

- $|S|$, S is a set: the cardinality of S;
- $B_\epsilon(x)$: the ball of radius ϵ, centred in x.

5.3 Problem Statement

We consider discrete-time systems of the following form

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + \varphi(Hx(k), u(k)) + d(k), \\ y(k) &= x(k) + w(k), \end{aligned} \tag{5.1}$$

where $x(k) \in \mathbb{R}^n$ is the state, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $u(k) \in \mathbb{R}^m$ is the driving signal, $\varphi : \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^n$ is a mapping, $H \in \mathbb{R}^{p\times n}$, $d(k) \in \mathbb{R}^n$ is an unknown state perturbation, $y(k) \in \mathbb{R}^n$ is the output, and $w(k) \in \mathbb{R}^n$ is a measurement error.

Remark 5.1. *The case of full state measurements is considered as this is equivalent to assuming observability for the system with bounded measurement noise. Since the main objective of the chapter is to design a remote observer, this technicality is left out.*

We assume the following about the driving signal.

Assumption 5.2. *The driving signal $u(k)$ is measured exactly and $u_i(k) \in [\underline{u}_i, \bar{u}_i]$, $i \in \{1, \dots, m\}$ where $[u_1(k) \ \dots \ u_m(k)]^\intercal = u(k) \in U \coloneqq \prod_{i=1}^m [\underline{u}_i, \bar{u}_i]$.*

The considered class of systems will be restricted to Lipschitz-nonlinear systems, which are commonly found structures of robotic and/or mechanical systems. We thus make the following assumption about the mapping φ.

Assumption 5.3. *There exist L such that*

$$\|\varphi(x_1, u) - \varphi(x_2, u)\| \leq L \|x_1 - x_2\|, \qquad \forall x_1, x_2 \in \mathbb{R}^p,$$

uniformly in $u \in \mathbb{R}^m$ and there exists $\bar{\varphi}$ such that

$$\|\varphi(x, u_1) - \varphi(x, u_2)\| \leq \bar{\varphi} \|u_1 - u_2\|, \qquad \forall u_1, u_2 \in U,$$

uniformly in $x \in \mathbb{R}^n$ (i.e., the mapping ϕ is uniformly globally Lipschitz with respect to both arguments).

Note that the previous assumption can be substituted, if the system is bounded-input bounded-state stable, by an assumption that the system is locally Lipschitz with respect to the state.

We make the following assumptions about the perturbations $d(k)$ and the errors $w(k)$.

Assumption 5.4. *There exists a maximum state perturbation $\delta > 0$ and maximum measurement error $\omega > 0$ such that*

$$\|d(k)\|_2 \leq \delta, \ \|w(k)\|_2 \leq \omega, \qquad \forall k \geq 0.$$

The system is equipped with a smart sensor (a sensor admitting some computational capacities, which allows it to perform additional computations on the measured data) and it is connected to a remote location via a data-rate constrained communication channel, which can only send messages that are of finite size. The objective is to provide estimates $\hat{x}(k)$ of $x(k)$ at the remote location by sending messages over this communication channel. Note that the measurements of input $u(k)$ are available at the plant side and also have to be communicated to the remote location. The sensor and the remote location are aware of an initial estimate $\hat{x}(0)$ which verifies

$$\|x(0) - \hat{x}(0)\|_2 \leq \epsilon_0, \tag{5.2}$$

where ϵ_0 is a selectable parameter corresponding to the error of initial conditions.

In order to generate the estimates, messages $m(k_j)$ are sent, where k_j are the transmission times and $j = \{0, 1, \ldots\}$ is the index of communication. Four ingredients interact with these messages: a sampler $\mathcal{S}$, a coder $\mathcal{C}$, an alphabet size function $\mathcal{A}$, and a decoder $\mathcal{D}$. The four devices together form a communication protocol. The following constants/parameters are known by all devices: the system matrices A, B H, the mapping φ, the constants L and $\bar{\varphi}$, the maximum state perturbation δ, the maximum measurement error ω, and the initial estimate $\hat{x}(0)$ with its accuracy ϵ_0. At the system side, the sampler $\mathcal{S}$ generates the instants of transmission in the following way

$$k_{j+1} = \mathcal{S}(\{u(k)\}_{k<k_{j+1}}, k_j, \{y(k)\}_{k<k_{j+1}}, \{m(k_i)\}_{i\leq j}), \tag{5.3}$$

$k_0 = 0$. The coder then generates the messages in the following way

$$\begin{aligned} m(k_j) = \mathcal{C}(\{u(k)\}_{k\leq k_j}, k_j, \hat{x}(0), \{y(k)\}_{k\leq k_j}, \\ \{m(k_i)\}_{i\leq j-1}), \end{aligned} \tag{5.4}$$

$\forall k_j : j > 0$. At each communication instant, the different possible messages are encoded into a finite-sized alphabet (the finite-sizedness being due to the data-rate constraints). The alphabet size function $\mathcal{A}$ determines the number of different messages l_j in the following way

$$l_j = \mathcal{A}(\hat{x}(0), \{m(k_i)\}_{i\leq j-1}), \qquad \forall j > 0. \tag{5.5}$$

The restriction on the choice of messages is then

$$m(k_j) \in \{1, \ldots, l_j\}, \qquad \forall j > 0.$$

When a message is sent, it is encoded by using bits. The number of bits b_j required to encode a message at communication instant k_j is defined as

$$b_j \coloneqq \lceil \log_2 l_j \rceil \qquad\qquad j > 0. \tag{5.6}$$

At the remote location, the decoder $\mathcal{D}$ receives the messages and interprets them to generate an estimate of the state $\hat{x}(k)$. For simplicity, we assume that there is no transmission delay, i.e., that the messages are received at the same time as they are sent. The decoder functions in the following way

$$\hat{x}(k) = \mathcal{D}(\hat{x}(0), m(k_1), \ldots, m(k_j)), \qquad \forall k \in \{k_j, \ldots, k_{j+1} - 1\},$$

$\forall j \geq 0$. Because of the perturbation, measurement error and finite data-rate, it is impossible to provide exact estimates at the remote location. Instead, the design of the communication protocol should ensure that the estimation error $\|x(k) - \hat{x}(k)\|$ is bounded by a quantity which we call the **maximum observation error** ξ and is defined as

$$\xi \coloneqq \sup_{k \geq 0} \|x(k) - \hat{x}(k)\|_2 . \tag{5.7}$$

In order to properly define the goal of the chapter, we need a quantity that evaluates the rate at which bits are sent through the communication channel. This quantity depends on both b_j, which can vary from one communication instant to another, and on $k_{j+1} - k_j$, which is fluctuating as well. We thus define the **maximum communication rate** R as

$$R \coloneqq \limsup_{\bar{j} \to \infty} \frac{\sum_{j=0}^{\bar{j}} b_j}{\bar{j} + 1}, \tag{5.8}$$

which can be commented as follows: first we define the average number of bits sent per unit of time during a window of $\bar{j}$ consecutive communications, next we start counting at time instant j such that this quantity is the largest possible, and finally, we take $\bar{j}$ such that the quantity is the smallest possible. This quantity is called maximum communication rate because we will provide results that guarantee that this rate is not exceeded.

The first objective of the chapter is to design a data-rate constrained observation scheme, to investigate what maximum observation error ξ it guarantees, and to determine the maximum communication rate R required to implement it. The second objective, is to ensure that this data-rate constrained observation scheme requires on average a lower communication rate than R when the perturbations occur in a favourable way. This will be achieved by using an even-triggered operation.

Remark 5.5. *The problem statement allows for an observation scheme that uses all previous inputs and outputs in order to generate estimates. Therefore,*

it is necessary to store these input-output data locally in memory, which can be costly. The solution that is presented in this work does not rely on storing all previous inputs and outputs. However, we decided to leave this possibility, to keep the problem statement as general as possible.

5.4 The Communication Scheme

In this section, we describe the different components that form the observation scheme: sampler, coder, alphabet, and decoder. The observation scheme needs to provide accurate estimates of $x(k)$ for all k. The naive solution is to simply send an estimate of $x(k)$ at every time instant. This solution is extremely inefficient in terms of data-rate, and a way to decrease the transmission rate is to only occasionally send estimates of the states and to utilize the system's dynamics on the decoder side in order to complement the estimates between the communications. This is the solution that will be used in this chapter.

The problem is that the decoder has no information about the driving signal $u(k)$ so it cannot reconstruct estimates based on the system's dynamics. The solution we are going to explore in this chapter, is to communicate a **driving signal of the estimate** $\hat{u}(k)$ via the messages at every time instant (i.e. $k_j = k$) and to sometimes send an estimate of the state $\hat{x}(k)$ in addition to $\hat{u}(k)$. The reasoning behind this idea is that the state and the driving signal have different dimensions (respectively n and m) and hence, their communication produces different loads on the channel. Typically, $m \leq n$ and it is less "expensive" to transmit an estimate of the driving signal than the state of the system.

Therefore, the observation scheme is going to send two different types of messages: messages which only contain the driving signal of the estimate and messages which contain both the driving signal of the estimate and an estimate of the state. We denote j_x is the index of the last communication instant when an estimate of the state was transmitted. The following dynamics are used to generate new estimates at the remote site in between messages containing $\hat{x}$:

$$\hat{x}(k) = A\hat{x}(k-1) + B\hat{u}(k-1) + \varphi(H\hat{x}(k-1), \hat{u}(k-1)), \tag{5.9}$$

$\forall k : k \neq k_{j_x}$, that is, for all time instants when the current message does not contain an estimate of the state.

Since these estimates are based entirely on the messages, the sampler and coder also maintain local copies of these estimates $\hat{x}_c(k) = \hat{x}(k)$ and $\hat{u}_c(k) = \hat{u}(k)$. These copies are used by the smart sensor to determine when to trigger communications, as will be explained further in this chapter. Figure 5.1 displays how the different agents interact.

When are messages sent? It is at this point that the event-triggered mechanism comes into play. Since the sampler knows $\hat{x}_c(k)$ and $\hat{u}_c(k)$, it will communicate new estimates $\hat{x}(k)$ only when the distance between $x(k)$ and $\hat{x}(k)$ becomes too large (including a margin of error to account for the measurement

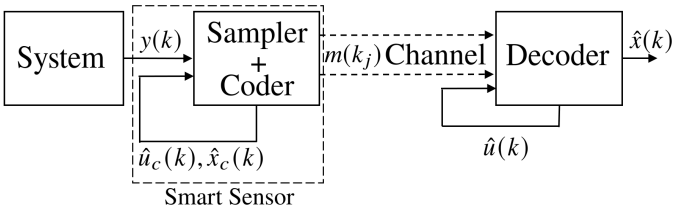

Figure 5.1: Structure of setup.

noise). Also, in order to reduce the amount of communications, the sampler will space out subsequent communications of estimates of $x(k)$ by at least $\bar{k}$ time instants. The quantity $\bar{k}$ is a tunable integer constant larger or equal to one, which allows one to tune the error and maximum communication rate, as will be proven later in the chapter. We assume that, as a part of the observation scheme, this quantity is known by all agents.

In order to properly define the observation scheme, we will need several additional notions.

First of all, the sets of points $\mathcal{V}$, $\mathcal{W}$ and $\mathcal{W}_x$. The set $\mathcal{V} = \{v_l\} \in \mathbb{R}^m$ is a constant set of points indexed from 1 till $|\mathcal{V}|$. The set $\mathcal{W} = \{w_l\} \in \mathbb{R}^n$ is a constant set of points indexed from 1 till $|\mathcal{W}|$. The set $\mathcal{W}_x$ is equal to the set $\mathcal{W}$, shifted by the vector x. Its elements $\{w_l^x\} \in \mathbb{R}^n$ are indexed from 1 till $|\mathcal{W}|$ (since the set is a shift of $\mathcal{W}$, it has the same number of elements).

Next, the notation $A^\frown B$ refers to the concatenation of the text A and B (i.e. "12"$^\frown$"34" = "1234"), $K \in \mathbb{R}^{m\times n}$ is a tunable constant gain matrix, α is a tunable constant. Since we communicate at every time instant, we have $k = k_j = j$. For the sake of correctness and to highlight their different meaning, we keep referring to each of these quantities separately, despite the fact that they are always equal. In any message $m(k_j)$, $m_u(k_j)$ refers to the part of the message that encodes information necessary to reconstruct $\hat{u}(k)$ and $m_x(k_j)$ to the part that encodes information necessary to reconstruct $\hat{x}(k)$. Finally, we define the provisional estimate $\hat{x}^-(k) \coloneqq A\hat{x}(k-1)+B\hat{u}(k-1)+\varphi(H\hat{x}(k-1), \hat{u}(k-1))$ and its local copy $\hat{x}_c^-(k) = \hat{x}^-(k)$ (which are maintained by the sampler and coder). Note that $\hat{x}^-(k) = \hat{x}(k)$ only if $k \neq k_{j_x}$. It is a provisional estimate and will coincide with the actual estimate $\hat{x}(k)$ only if no message is sent. If a message containing information to reconstruct $\hat{x}(k)$ is sent, $\hat{x}^-(k)$ is discarded and that message is used to generate the estimate $\hat{x}(k)$.

Procedure 3.
Sampler $\mathcal{S}$*:*

1: ***if*** $k = 0$ ***then***
2: $\quad j \leftarrow 0$
3: $\quad j_x \leftarrow 0$
4: $\quad k_j \leftarrow 0$

5: ***else***
6: $\quad j \leftarrow k_j \leftarrow k$
7: ***end if***

Alphabet Size Function $\mathcal{A}$***:***

1: $l_j \leftarrow |\mathcal{V}|$
2: ***if*** $k_j \geq k_{j_x} + \bar{k}$ ***and*** $\|y(k) - \hat{x}_c^-(k)\|_2 \geq \alpha$ ***then***
3: $\quad l_j \leftarrow l_j\,|\mathcal{W}|$
4: ***end if***

Coder $\mathcal{C}$***:***

1: $\hat{x}_c^-(k) = A\hat{x}_c(k-1) + B\hat{u}_c(k-1) + \varphi(H\hat{x}_c(k-1), \hat{u}_c(k-1))$
2: $m(k_j) \leftarrow \arg \min_{l \in \{1,\dots,|\mathcal{V}|\}} \|u(k) - K(y(k) - \hat{x}(k)) - v_l\|_2$
3: $\hat{u}_c(k) \leftarrow v^u_{m(k_j)}$
4: ***if*** $k_j \geq k_{j_x} + \bar{k}$ ***and*** $\|y(k) - \hat{x}_c^-(k)\|_2 \geq \alpha$ ***then***
5: $\quad m(k_j) \leftarrow m(k_j)^\frown \arg \min_{l \in \{1,\dots,|\mathcal{W}|\}} \left\| y(k) - w_l^{\hat{x}_c^-(k)} \right\|_2$
6: $\quad \hat{x}_c(k) \leftarrow w^{\hat{x}_c^-(k)}_{m_x(k_j)}$
7: $\quad j_x \leftarrow k_j$
8: ***else***
9: $\quad \hat{x}_c(k) \leftarrow \hat{x}_c^-(k)$
10: ***end if***

Decoder $\mathcal{D}$***:***

1: $\hat{x}^-(k) = A\hat{x}(k-1) + B\hat{u}(k-1) + \varphi(H\hat{x}(k-1), \hat{u}(k-1))$
2: $\hat{u}(k) \leftarrow v_{m_u(k_j)}$
3: ***if*** $m_x(k_j) \neq \emptyset$ ***then***
4: $\quad \hat{x}(k) \leftarrow w^{\hat{x}^-(k)}_{m_x(k_j)}$
5: ***else***
6: $\quad \hat{x}(k) \leftarrow \hat{x}^-(k)$
7: ***end if***

The Sampler $\mathcal{S}$ is relatively simple: there is an initialization step (lines 2-4) and an incremental step (lines 6), where j and k_j are simply updated to the current time instant (since we communicate $\hat{u}$ at every time instant).

The Alphabet Size Function $\mathcal{A}$ starts (line 1) by being equal to the cardinality of $\mathcal{V}$ (since $\mathcal{V}$ is used to transmit $\hat{u}$ and we communicate $\hat{u}$ at every time instant). Next (line 2), it checks whether two conditions are verified: firstly that enough time has elapsed since the last communication of $\hat{x}$ ($k_j \geq k_{j_x} + \bar{k}$) and secondly that the triggering condition is verified ($\|y(k) - \hat{x}_c^-(k)\|_2 \geq \alpha$). If both condition are true, then an estimate $\hat{x}$ will be communicated in addition to $\hat{u}$ so we increase the length of the alphabet by multiplying the previous value (which was set to the cardinality of $\mathcal{V}$ at line 1) with the cardinality of $\mathcal{W}$ (line 3).

The Coder $\mathcal{C}$ starts by computing the local copy of the provisional estimate

$\hat{x}_c^-$ (line 1). It then computes the point in $\mathcal{V}$ closest to $u(k) + K(y(k) - \hat{x}(k))$, uses its index as the message (line 2), and updates the local copy of $\hat{u}_c$ (line 3). Next, it verifies whether the triggering condition is verified (line 4). If it is, then it computes the point in $\mathcal{W}_{\hat{x}_c^-(k)}$ closest to $y(k)$, concatenates its index to the previous part of the message (line 5), updates the local copy of $\hat{x}$ accordingly (line 6), and updates j_x (line 7). If the triggering condition is not verified then the provisional estimate is used for $\hat{x}_c(k)$ (line 9).

The Decoder $\mathcal{D}$ starts by computing the provisional estimate $\hat{x}^-$ (line 1) and uses the index it received via the message to set the current driving signal of the estimate to the point in $\mathcal{V}$ to which this index corresponds (line 2). Next, if it received a message containing information on $\hat{x}$ (line 3), then it uses the point in $W_{\hat{x}_c^-(k)}$ corresponding to the received index (line 4), otherwise, it updates $\hat{x}$ to be equal to $\hat{x}^-$ (line 6). Since $|\mathcal{V}|$ and $|\mathcal{W}|$ are constant, the length of the alphabet indicates whether a message contains information about $\hat{u}(k)$, or both $\hat{u}(k)$ and $\hat{x}(k)$. By looking at the number of bits it received, the decoder can thus clearly distinguish between the different types of messages. The decoder can also easily distinguish which bits encode information about the driving signal and which bits encode information about the state, when estimates of both are sent at the same time (the first bits encode for the driving signal, the last bits for the state).

As it follows from this procedure, the driving signal of the estimate $\hat{u}(k)$ includes a state correction term:

$$\hat{u}(k) = u(k) - K(y(k) - \hat{x}(k)).$$

Further in this chapter, we will prove if Assumption 5.2 holds, there exists η_u such that $\hat{u}(k) \in B_{\eta_u}(\frac{\underline{u}+\bar{u}}{2})$, i.e., that the driving signal of the estimate is bounded. We conclude this section with several remarks regarding the observation scheme.

Remark 5.6.

- *The observation scheme relies on choices of the quantities $\bar{k}$, ϵ_u, $\mathcal{V}$, ϵ_x, $\mathcal{W}$, K, α. At this stage, there is no guarantee that, no matter the choices for these quantities, the observation scheme will always be implementable. Moreover, the choices of these quantities greatly impact the resulting maximum observation error and maximum communication rate. This will be discussed extensively in the next section.*

- *The sets $\mathcal{V}$ and $\mathcal{W}$ play an essential role in the observation scheme. They will be used to make a covering of the sets containing $\hat{u}(k)$ and $\hat{x}(k_{j_x})$. The observation scheme relies on them having the following properties:*

  - *For any possible $\hat{u}(k)$, there should exists $v_l \in \mathcal{V}$ such that $\|\hat{u}(k) - v_l\| \leq \epsilon_u$ (precision of the covering).*

- *For any possible* $\hat{x}(k_{j_x})$*, there should exists* $w_l^{\hat{x}^-(k_{j_x})} \in \mathcal{W}_{\hat{x}^-(k_{j_x})}$ *such that* $\left\| \hat{x}(k_{j_x}) - w_l^{\hat{x}^-(k)} \right\| \leq \epsilon_x$ *(precision of the covering).*
- *The cardinalities* $|\mathcal{V}|$ *and* $|\mathcal{W}|$ *are finite (size of the covering).*

5.5 Choices, Error and Rates

With the observation scheme and its agents fully introduced, this section aims to answer the following questions: "What is the resulting maximum error ξ?" and "What is the resulting maximum communication rate R?". The answers come in the form of two theorems: one for each quantity.

Before the theorems can be presented, there is a need to provide proper choices for the constant α, gain matrix K, and sets $\mathcal{V}$ and $\mathcal{W}$.

To this end, we first introduce the following Bilinear Matrix Inequality program.

$$\begin{aligned}
&(\mu_i^*, N^*, Q_i^*, S^*) \coloneqq \arg \min_{\mu_i, N, Q_i, S} (\delta^2 \mu_4 + \epsilon_u^2 \mu_5 + \omega^2 \mu_6), \\
&\text{subject to: (5.21) and} \\
&\mu_i I_n \succeq Q_i, \ \forall i \in \{2, \ldots, 4\}, \\
&\mu_i I_m \succeq Q_i, \ \forall i \in \{5, 6\}, \\
&\mu_1 \geq 0, \\
&N \succ 0, \\
&Q_i \succeq 0, \ \forall i \in \{2, \ldots, 6\},
\end{aligned} \tag{5.10}$$

Based on the solution of this program $(\mu_i^*, N^*, Q_i^*, S^*)$ (due to the structure of this particular BMI, such a solution always exists), we define

$$P^* \coloneqq (N^*)^{-1} \tag{5.11}$$

and the matrix T which is a matrix verifying $T^\intercal T = P^*$. Note that since P^* is symmetric and positive definite, this decomposition always exists. The gain K used in the observation scheme is defined as follows:

$$K^* \coloneqq P^* S^*. \tag{5.12}$$

As was previously mentioned, an important part of the observation scheme is the sets $\mathcal{V}$ and $\mathcal{W}$. We will use the concept of covering of a set which is defined as follows:

Definition 5.7. *A set* S_1 *with elements* s_l ***generates a covering*** *of radius* ϵ_s *of the set* S_2 *if the following holds*

$$S_2 \subseteq \bigcup_{l=1}^{|S_1|} B_{\epsilon_s}(s_l).$$

The set $\mathcal{V}$ is then determined as follows: first we compute η_u such that the distance between $\hat{u}(k)$ and $\frac{\underline{u}+\bar{u}}{2}$ is at most η_u (that is $\hat{u}(k) \in B_{\eta_u}(\frac{\underline{u}+\bar{u}}{2})$), next we define $\mathcal{V}$ such that it generates a covering of radius ϵ_u of $B_{\eta_u}(\frac{\underline{u}+\bar{u}}{2})$. This necessarily implies that $\min_{l\in\{1,\dots,|\mathcal{V}|\}} \|\hat{u}(k_j) - v_l\|_2 \leq \epsilon_u$.

For the set $\mathcal{W}$, we proceed similarly by computing η_x such that $\hat{x}(k_{j_x}) \in B_{\eta_x}(\hat{x}^-(k_{j_x}))$, then we define $\mathcal{W}$ such that it is a covering of radius ϵ_x of $B_{\eta_x}(0)$. The sets $\mathcal{W}_{\hat{x}^-(k_{j_x})}$ are then constructed by simply shifting the set $\mathcal{W}$ by a vector $\hat{x}^-(k_{j_x})$.

The aforementioned quantities η_u and η_x are chosen as :

$$\begin{aligned}\eta_u \coloneqq & \left\| \frac{\underline{u}-\bar{u}}{2} \right\|_2 + \sigma_1(K^*)\omega + \epsilon_u \\ & + \sigma_1(K^*) \max \left\{ \epsilon_x + \omega, \frac{(\mu_1^*)^{\frac{\bar{k}-1}{2}} \sigma_1(T)}{\sigma_n(T)} (\epsilon_x + \omega) \right. \\ & \left. + \frac{1}{\sigma_n(T)} (\mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 + \mu_6^* \sigma_1(K^*)^2 \omega^2)^{\frac{1}{2}} \sum_{i=0}^{\bar{k}-2} (\mu_1^*)^{\frac{i}{2}} \right\},\end{aligned} \tag{5.13}$$

$$\begin{aligned}\eta_x \coloneqq & \frac{(\mu_1^*)^{\frac{\bar{k}}{2}} \sigma_1(T)}{\sigma_n(T)} (\epsilon_x + \omega) + \omega \\ & + \frac{1}{\sigma_n(T)} \left(\mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 + \mu_6^* \sigma_1(T)^2 \omega^2 \right)^{\frac{1}{2}} \sum_{i=0}^{\bar{k}-1} (\mu_1^*)^{\frac{i}{2}}\end{aligned}. \tag{5.14}$$

In the appendix, Lemmata 5.13 and 5.14 prove that by choosing them as above, we indeed have $\hat{u}(k) \in B_{\eta_u}(\frac{\underline{u}+\bar{u}}{2})$ and $\hat{x}(k_{j_x}) \in B_{\eta_x}(\hat{x}^-(k_{j_x}))$.

In order to implement Procedure 3, it is also necessary to define α, which is used in the triggering condition of the observation scheme and it is defined as $\alpha \coloneqq$

$$\begin{aligned}\min \Bigg\{ & \frac{\sigma_n(T)[\eta_x - \omega - \left(\mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 + \mu_6^* \sigma_1(K^*)^2 \omega^2 \right)^{\frac{1}{2}}]}{(\mu_1^*)^{\frac{1}{2}} \sigma_1(T)} - \omega, \\ & \max \Bigg\{ \epsilon_x + \omega, \frac{(\mu_1^*)^{\frac{\bar{k}-1}{2}} \sigma_1(T)}{\sigma_n(T)} (\epsilon_x + \omega) \\ & \frac{1}{\sigma_n(T)} (\mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 + \mu_6^* \sigma_1(K^*)^2 \omega^2)^{\frac{1}{2}} \sum_{i=0}^{\bar{k}-2} (\mu_1^*)^{\frac{i}{2}} \Bigg\} \Bigg\}.\end{aligned} \tag{5.15}$$

We now present the two main results of the chapter: first a theorem that provides a bound on the maximum observation error and then a theorem that gives a bound on the maximum communication rate.

In the following theorem, $\bar{k}$, ϵ_x and ϵ_u are tunable constants.

Theorem 5.8. *Let Assumptions 5.2 to 5.4 hold for the system (5.1). Then for any $\bar{k} \geq 1$, $\epsilon_x > 0$, $\epsilon_0 \leq \epsilon_x + \omega$ and $\epsilon_u > 0$, Procedure 3 guarantees the following bound on the* ***maximum observation error****:*

$$\xi \leq \max \left\{ \epsilon_x + \omega, \frac{(\mu_1^*)^{\frac{\bar{k}-1}{2}} \sigma_1(T)}{\sigma_n(T)} (\epsilon_x + \omega) \right. \\ \left. + \sum_{i=0}^{\bar{k}-2} (\mu_1^*)^{\frac{i}{2}} \frac{1}{\sigma_n(T)} (\mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 + \mu_6^* \sigma_1(K^*)^2 \omega^2)^{\frac{1}{2}} \right\} \tag{5.16}$$

The proof of this theorem is provided in Appendix 5.A.1.

In the following theorem, the notation $\lceil \cdot \rceil$ refers to the ceiling function (i.e. the function that rounds up to the nearest integer).

Theorem 5.9. *Let Assumptions 5.2 to 5.4 hold for the system (5.1). For any $\bar{k} \geq 1$, $\epsilon_x > 0$, $\epsilon_0 \leq \epsilon_x + \omega$ and $\epsilon_u > 0$, Procedure 3 results in the following bound on the* ***maximum communication rate****:*

$$R \leq \left\lceil m \log_2 \frac{2\eta_u \sqrt{m}}{\epsilon_u} \right\rceil + \frac{\left\lceil n \log_2 \frac{2\eta_x \sqrt{n}}{\epsilon_x} \right\rceil}{\bar{k}} \tag{5.17}$$

The proof of this theorem is presented in Appendix 5.A.2.

5.6 Simulations

In this section, various systems are simulated under the observation scheme. The objectives are:

- To compare the bound on R with the actual rate observed in simulations;
- To show how the Lipschitz-nonlinear term of the system's dynamics affects R, the actual rate, and ξ;
- To show how R, the actual rate, and ξ are influenced by the choices of $\bar{k}$, ϵ_x, and ϵ_u;
- To show how R, the actual rate, and ξ are influenced by the perturbations $d(k)$ and $w(k)$.

We will consider two examples of Lipschitz-nonlinear systems. First, a simple two-dimensional system with a trigonometric nonlinearity, which is a structure typical of mechanical systems. Secondly, a model for a flexible joint robot, which has already been studied in [Raghavan and Hedrick, 1994].

5.6.1 Example 1

We consider the discretization, by using an Euler method of the following continuous-time system (inspired by Example 2 of [Grandvallet et al., 2013]):

$$\begin{aligned}\dot{x}(t) &= \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} x(t) + \begin{bmatrix}1\\ 0\end{bmatrix} u(t) + \begin{bmatrix}\gamma \sin(x_2(t))\\ 0\end{bmatrix} + d(t),\\ y(t) &= x(t) + w(t),\end{aligned}$$

where $\gamma > 0$ is a parameter, with a discretization step of 10 ms. This yields the following discrete-time system:

$$\begin{aligned}x(k+1) = &\begin{bmatrix}1 & 0.01\\ 0 & 1\end{bmatrix} x(k) + \begin{bmatrix}0.01\\ 0\end{bmatrix} u(k) + \begin{bmatrix}0.01\gamma \sin(x_2(k))\\ 0\end{bmatrix}\\ &+ d(k),\\ y(k) = &x(k) + w(k),\end{aligned} \tag{5.18}$$

For this system, Assumption 5.3 holds with $L = 0.01\gamma$ and $\bar{\varphi} = 0$. We first consider $\gamma = 2$, Assumption 5.2 with $\underline{u}_1 = -1$ and $\bar{u}_1 = 1$, Assumption 5.4 with $\delta = 0.01$ and $\omega = 0.01$, and the following choices for the observer constants

$$\epsilon_u = 0.01,\ \epsilon_x = 0.01,\ \bar{k} = 2.$$

Solving (5.10) and using Theorems 5.8 and 5.9, we obtain

$$\begin{aligned}\xi &\leq 0.0519\\ R &\leq 17.\end{aligned}$$

We used Monte-Carlo methods to simulate the observation scheme for $10,000$ iterations from $k = 0$ to $k = 1000$ each. The simulated communication rate R^* is $R^* = 12.0068$. The number of bits N_u used to transmit $\hat{u}$ is $N_u = 12$ and the number of bits N_x required to transmit $\hat{x}$ is $N_x = 10$. Since we communicate $\hat{u}$ at every instant, we sent 12 bits at each time instant. Since the communications of $\hat{x}$ are spaced out by at least $\bar{k}$ instants, we at most send 10 bits every $\bar{k}$ instants. When decomposing the rate between the rate that is used to send estimates of u (R_u^*) and the rate that is used to send estimates of x (R_x^*), we observe $R_u^* = 12$ and $R_x^* = 0.0068$. This means that the coder extremely rarely sends estimates of the state and instead relies on the driving signal of the estimate to keep the error low. We compared these quantities for varying γ and the same values for the other system parameters and observer constants. The results are displayed in Table 5.1. We observe that:

- The nonlinearity influences ξ: the bigger γ, the bigger the maximum error. Since the Lipschitz nonlinearity is modelled as a perturbation, it is natural that a larger nonlinearity implies a larger ξ;

- The communication rate is mostly due to communications of $\hat{u}$;
- In terms of communication rate, we observe that a significant decrease happens, when γ is increased. This is due to the fact that α grows with L and hence with γ, which implies that the triggering condition is verified less often. For $\gamma = 10$, there are simply no communications of the state at all whilst ξ remains small which implies that our observer makes good usage of the driving signal of the estimate.

γ	1	2	5	10
ξ	0.0479	0.0519	0.0607	0.0706
N_u	12	12	12	12
N_x	10	10	11	11
R^*	12.0279	12.0068	12.0003	12
R_u^*	8	8	8	8
R_x^*	0.0279	0.0068	0.0003	0

Table 5.1: Example 1: Results for various values γ.

Next, we simulated the observation scheme, using the same process, for $\gamma = 2$ and various choices of $\bar{k}$. The results are displayed in Table 5.2. We observe that ξ is greatly influenced by $\bar{k}$. The error is multiplied by 5 while $\bar{k}$ increases from 2 to 5. The same effect as for the nonlinearity happens here: the increase in $\bar{k}$ also increases α which means that the triggering condition is verified less and less often. The event-triggering mechanism is very useful in this case as it completely removes the need to communicate estimates of the state.

$\bar{k}$	1	2	3	5
ξ	0.0200	0.0519	0.0958	0.2172
N_u	11	12	12	13
N_x	8	10	11	13
R^*	18.992	12.0068	12	13
R_u^*	11	12	12	13
R_x^*	7.992	0.0068	0	0

Table 5.2: Example 1: Results for various choices $\bar{k}$.

5.6.2 Example 2

Now that the effects of L and $\bar{k}$ have been illustrated on a simple system, we turn to a higher order system to illustrate the effects of changing $\bar{k}$, ϵ_u, ϵ_x, δ, and ω.

We consider the discretization of the following continuous-time system, which was first introduced in [Raghavan and Hedrick, 1994] *(Flexible joint robot)*:

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -48.6 & -1.25 & 48.6 & 0 \\ 0 & 1 & 0 & 0 \\ 19.5 & 0 & -19.5 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 21.6u(t) \\ 0 \\ -3.33\sin(x_3(t)) \end{bmatrix}$$

We consider a discretization step of 10 ms. The system matrices and mapping are then

$$A = \begin{bmatrix} 1 & 0.01 & 0 & 0 \\ -0.4860 & 0.9875 & 0.486 & 0 \\ 0 & 0 & 1 & 0.01 \\ 0.195 & 0 & -0.195 & 1 \end{bmatrix}, \; B = \begin{bmatrix} 0 \\ 0.216 \\ 0 \\ 0 \end{bmatrix},$$

$$H = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}, \varphi : \zeta \to \begin{bmatrix} 0 \\ 0 \\ 0 \\ -0.033\sin(\zeta) \end{bmatrix}$$

For this system, Assumption 5.3 holds with $L = 0.033$ and $\bar{\varphi} = 0$. For the simulations, the default values for the different parameters and constants are as follow: we assume Assumption 5.2 with $\underline{u}_1 = -10$ and $\bar{u}_1 = 10$, Assumption 5.4 with $\delta = 0.5$ and $\omega = 0.2$, and the following choices for the observer constants

$$\epsilon_u = 0.1, \; \epsilon_x = 0.05, \; \bar{k} = 2.$$

For all constants that are not mentioned in the upcoming description of the simulations, the reader should assume that the constant takes the aforementioned default value. The first simulations investigated the impact of changes in $\bar{k}$. The results are displayed in Table 5.3. We observe the following effects:

- The error is greatly influenced by the choice of $\bar{k}$;
- In terms of transmitted number of bits bit, both N_u and N_x increase with $\bar{k}$;
- For the case $\bar{k} = 1$, we can see that the maximum error is extremely close to the sum of the measurement error and the discretization error of $\hat{x}$, which is why the event-triggering condition is triggered almost at every communication instant and hence why the theoretical rate of 37 is almost equal to the actual rate. For larger $\bar{k}$, the error increases drastically and hence the need for communications of $\hat{x}$ decreases also.

We then analysed the impact of the choices of ϵ_u and ϵ_x. The results of the simulations are displayed in Tables 5.4 and 5.5. Several observations are to be made:

k	1	2	3	5
ξ	0.2678	1.2164	2.7549	7.6805
N_u	9	10	10	11
N_x	28	32	35	40
R^*	36.5411	12.6822	11.0757	11.468
R_u^*	9	10	10	11
R_x^*	27.5411	2.6822	1.0757	0.468

Table 5.3: Example 2: Results for various choices $\bar{k}$.

- ϵ_u has a smaller impact on the error than ϵ_x proportionally;
- Augmenting both ϵ_u and ϵ_x increases the discretization error, leads to a larger error, and a smaller communication rate;
- The impact in terms of rate is greater for ϵ_u than for ϵ_x;
- The conclusion of these simulations is to send less precise estimates of the estimate driving signal and more precise estimates of the state rather than the other way around.

ϵ_u	0.01	0.1	1
ξ	1.2087	1.2164	1.3382
N_u	13	10	6
N_x	32	32	32
R^*	15.7028	12.6822	8.3312
R_u^*	13	10	6
R_x^*	2.7028	2.6822	2.3312

Table 5.4: Example 2: Results for various choices ϵ_u.

ϵ_x	0.025	0.05	0.1
ξ	1.1641	1.2164	1.3195
N_u	10	10	10
N_x	36	32	28
R^*	13.2434	12.6822	12.0589
R_u^*	10	10	10
R_x^*	3.2434	2.6822	2.0589

Table 5.5: Example 2: Results for various choices ϵ_x.

Finally, we analysed the impact of the size of the state perturbations and measurement error, through the bounds on their maximum norm δ and ω. The results for various values of δ and ω are displayed in Tables 5.6 and 5.7. We make the following observations about these results:

- The impact of the state perturbation on the error and on the communication rate is greater than the impact of the measurement error;
- Even in situation with high perturbations, the number of communications stays well below the theoretical maximum (for $\delta = 1$, the theoretical maximum is $10 + 35/2 = 27.5$);

- The measurement noise only affects the rate up to a certain point. When it becomes too large, α becomes large as well and hence the triggering condition is more rarely verified, which explains the reduction in communication rate.

δ	0.25	0.5	1
ξ	0.9137	1.2164	1.8088
N_u	9	10	10
N_x	30	32	35
R^*	10.5982	12.6822	14.4613
R_u^*	9	10	10
R_x^*	1.5982	2.6822	4.4613

Table 5.6: Example 2: Results for various values of δ.

ω	0.01	0.1	1
ξ	0.9823	1.2164	1.6688
N_u	9	10	10
N_x	31	32	33
R^*	12.0663	12.6822	12.0865
R_u^*	9	10	10
R_x^*	2.0663	2.6822	2.0865

Table 5.7: Example 2: Results for various choices ω.

Based on these simulations, we make the following concluding observations:

- The minimum duration between subsequent communications $\bar{k}$ is the constant that has the biggest impact on ξ;
- It is better to choose ϵ_u large rather than ϵ_x, both in terms of error and in terms of communication rate;
- The event-triggering mechanism greatly reduces the actual communication rate compared to the theoretical maximum;
- Most of the communication rate is due to transmitting the driving signal of the estimate.

5.7 Conclusion

In this chapter, an approach was proposed for the remote observation of a dynamical system with a Lipschitz nonlinearity through a data-rate constrained communication channel. A solution, named observation scheme, in the form of several interacting agents was proposed and evaluated. The main features of the observation scheme are:

- The maximum observation error is upper-bounded by a quantity that can be computed from the system's features as well as selectable parameters;
- The maximum communication rate is also upper bounded by a quantity that can be computed from the system's features as well as selectable parameters;

- The scheme uses an event-triggering mechanism to reduce the overall required communication rate, without reducing the performance in terms of maximum observation error.

As was demonstrated through simulations, the actual communication rate is much lower than the theoretically evaluated maximum one, which is due to two factors. First of all, the error bounds are conservative, which is due to the usage of a basic quadratic Lyapunov function. Secondly, the event-triggered communication protocol, which greatly helps in reducing the resulting communication rate.

The continuation of this work includes:

- Improving the error bounds through finding a better Lyapunov function;
- Extending the observer to deal with a larger class of nonlinear systems;
- Using this observation scheme for consensus problems in networks of perturbed dynamical systems with data-rate constraints.

Appendices

5.A Proofs of Section 5.5

We first provide several auxiliary lemmata. We define $e(k) \coloneqq x(k) - \hat{x}(k)$ and $V(e(k)) \coloneqq e(k)^\intercal P^* e(k)$, where P^* is defined in (5.11).

The following lemma provides a bound on the one-step evolution of $V(e(k))$, provided that $\hat{u}(k) = u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)$, where $e_u(k)$ is the quantization error on $\hat{u}(k)$ at time instant k which verifies $\|e_u(k)\|_2 \leq \epsilon_u$, ϵ_u being the maximum quantization error.

Lemma 5.10. *Let Assumptions 5.2 to 5.4 hold for the system (5.1). For any $k \geq 0$, any $e_u(k)$ such that $\|e_u(k)\|_2 \leq \epsilon_u$, any $x(k), \hat{x}(k) \in \mathbb{R}^n$, any $w(k), d(k) \in \mathbb{R}^n$ satisfying Assumption 5.4, any $u(k)$ satisfying Assumption 5.2, and for $y(k)$, $\hat{u}(k)$, $e(k+1)$, $x(k+1)$, and $\hat{x}(k+1)$ such that $y(k) = x(k) + w(k)$, $\hat{u}(k) = u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)$, $e(k+1) = x(k+1) - \hat{x}(k+1)$, $x(k+1) = Ax(k) + Bu(k) + \varphi(Hx(k), u(k)) + d(k)$, and $\hat{x}(k+1) = A\hat{x}(k) + B\hat{u}(k) + \varphi(H\hat{x}(k), \hat{u}(k)))$ the following holds*

$$V(e(k+1)) \leq \mu_1^* V(e(k)) + \mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 + \mu_6^* \sigma_1(K^*)^2 \omega^2. \tag{5.19}$$

Proof. From

$$x(k+1) = Ax(k) + Bu(k) + \varphi(Hx(k), u(k)) + d(k),$$

$\hat{u}(k) = u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)$, and

$$\begin{aligned}\hat{x}(k+1) =& A\hat{x}(k) + B\hat{u}(k) + \varphi(H\hat{x}(k), \hat{u}(k)) \\ =& A\hat{x}(k) + Bu(k) - BK^*(y(k) - \hat{x}(k)) + Be_u(k) \\ &+ \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)) \\ =& A\hat{x}(k) + Bu(k) - BK^*(x(k) - \hat{x}(k)) + Be_u(k) \\ &+ \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)) \\ &- BK^*w(k),\end{aligned}$$

we have

$$\begin{aligned}&e(k+1) = Ax(k) + Bu(k) + \varphi(Hx(k), u(k)) + d(k) - A\hat{x}(k) \\ &- Bu(k) + BK^*(x(k) - \hat{x}(k)) - Be_u(k) + BK^*w(k) \\ &- \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)) \\ &= (A+BK^*)e(k) + \varphi(Hx(k), u(k)) + d(k) - Be_u(k) + BK^*w(k) \\ &- \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)).\end{aligned}$$

We add and subtract $\varphi(H\hat{x}(k), u(k))$ from the previous equation to obtain

$$\begin{aligned}&e(k+1) = (A + BK^*)e(k) + \varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k)) \\ &+ \varphi(H\hat{x}(k), u(k)) - \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)) \\ &+ d(k) - Be_u(k) + BK^*w(k).\end{aligned}$$

We then have

$$\begin{aligned}&V(e(k+1)) = e(k+1)^\intercal P^* e(k+1) \\ &= [(A + BK^*)e(k) + \varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k)) \\ &+ \varphi(H\hat{x}(k), u(k)) - \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)) \\ &+ d(k) - Be_u(k) + BK^*w(k)]^\intercal P^* \\ &[(A + BK^*)e(k) + \varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k)) \\ &+ \varphi(H\hat{x}(k), u(k)) - \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)) \\ &+ d(k) - Be_u(k) + BK^*w(k)]\,.\end{aligned}$$

By adding an subtracting the following terms from the previous equation

- $\mu_1 e(k)^\intercal P e(k)$;
- $L^2\mu_2^* e(k)^\intercal H^\intercal H e(k)$;
- $[\varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k))]^\intercal Q_2^* [\varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k))]$;
- $[\varphi(H\hat{x}(k), u(k)) - \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k))]^\intercal Q_3^*$ $[\varphi(H\hat{x}(k), u(k)) - \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k))]$;

- $d(k)^\intercal Q_4^* d(k)$;
- $e_u(k)^\intercal Q_5^* e_u(k)$;
- $w(k)^\intercal (K^*)^\intercal Q_6^* K^* w(k)$;

it can be rewritten as

$$\begin{aligned}
&V(e(k+1)) = z(k)^\intercal M z(k) + \mu_1^* e(k)^\intercal P^* e(k) \\
&- L^2 \mu_2^* e(k)^\intercal H^\intercal H e(k) + [\varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k))]^\intercal Q_2^* \\
&[\varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k))] + [\varphi(H\hat{x}(k), u(k)) \\
&- \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k))]^\intercal Q_3^* [\varphi(H\hat{x}(k), u(k)) \\
&- \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k))] + d(k)^\intercal Q_4^* d_u(k) \\
&+ e_u(k)^\intercal Q_5^* e_u(k) + w(k)^\intercal (K^*)^\intercal Q_6^* K^* w(k)
\end{aligned}$$

where $z(k) =$

$$\begin{bmatrix}
e(k) \\
\varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k)) \\
\varphi(Hx(k), u(k)) - \varphi(Hx(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)) \\
d(k) \\
e_u(k) \\
K^* w(k)
\end{bmatrix}$$

and M as in (5.22). From (5.10), we have $\mu_2^* I_n \succeq Q_2^*$ $\mu_3^* I_n \succeq Q_3^*$, $\mu_4^* I_n \succeq Q_4^*$, $\mu_5^* I_m \succeq Q_5^*$, $\mu_6^* I_m \succeq Q_6^*$ and hence

$$\begin{aligned}
&V(e(k+1)) = z(k)^\intercal M z(k) + \mu_1^* e(k)^\intercal P^* e(k) \\
&- L^2 \mu_2^* e(k)^\intercal H^\intercal H e(k) + \mu_2^* [\varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k))]^\intercal \\
&[\varphi(Hx(k), u(k)) - \varphi(H\hat{x}(k), u(k))] + \mu_3^* [\varphi(H\hat{x}(k), u(k)) \\
&- \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k))]^\intercal [\varphi(H\hat{x}(k), u(k)) \\
&- \varphi(H\hat{x}(k), u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k))] + \mu_4^* d(k)^\intercal d_u(k) \\
&+ \mu_5^* e_u(k)^\intercal e_u(k) + \mu_6^* w(k)^\intercal (K^*)^\intercal K^* w(k).
\end{aligned}$$

Using Assumption 5.3, we have

$$\begin{aligned}
&V(e(k+1)) = z(k)^\intercal M z(k) + \mu_1^* e(k)^\intercal P^* e(k) \\
&- L^2 \mu_2^* e(k)^\intercal H^\intercal H e(k) + L^2 \mu_2^* [Hx(k) - H\hat{x}(k)]^\intercal [Hx(k) - H\hat{x}(k)] \\
&+ \mu_3^* \bar{\varphi}^2 [u(k) - u(k) + K^*(y(k) - \hat{x}(k)) - e_u(k)]^\intercal [u(k) - u(k) \\
&+ K^*(y(k) - \hat{x}(k)) - e_u(k)] + \mu_4^* d(k)^\intercal d_u(k) + \mu_5^* e_u(k)^\intercal e_u(k) \\
&+ \mu_6^* w(k)^\intercal (K^*)^\intercal K^* w(k).
\end{aligned}$$

Equivalently

$$\begin{aligned}
&V(e(k+1)) = z(k)^\intercal M z(k) + \mu_1^* e(k)^\intercal P^* e(k) \\
&- L^2 \mu_2^* e(k)^\intercal H^\intercal H e(k) + L^2 \mu_2^* e(k)^\intercal H^\intercal H e(k) \\
&+ \mu_3^* \bar{\varphi}^2 [K^*(y(k) - \hat{x}(k)) - e_u(k)]^\intercal [K^*(y(k) - \hat{x}(k)) - e_u(k)] \\
&+ \mu_4^* d(k)^\intercal d_u(k) + \mu_5^* e_u(k)^\intercal e_u(k) + \mu_6^* w(k)^\intercal (K^*)^\intercal K^* w(k)
\end{aligned}$$

and thus

$$\begin{aligned}
&V(e(k+1)) = z(k)^\intercal M z(k) + \mu_1^* e(k)^\intercal P^* e(k) \\
&+ \mu_3^* \bar{\varphi}^2 [K^*(y(k) - \hat{x}(k)) - e_u(k)]^\intercal [K^*(y(k) - \hat{x}(k)) - e_u(k)] \\
&+ \mu_4^* d(k)^\intercal d_u(k) + \mu_5^* e_u(k)^\intercal e_u(k) + \mu_6^* w(k)^\intercal (K^*)^\intercal K^* w(k)
\end{aligned}$$

Using (5.1), Assumption 5.4, as well as the fact that $\|Kx\| \leq \sigma_1(K) \|x\|_2$ for any vector x and matrix K, we obtain

$$\begin{aligned}
&V(e(k+1)) \leq z(k)^\intercal M z(k) + \mu_1^* V(e(k)) + \mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 \\
&+ \mu_6^* \sigma_1(K^*)^2 \omega^2 \\
&+ \mu_3^* \bar{\varphi}^2 [K^* e(k) + K^* w(k) - e_u(k)]^\intercal [K^* e(k) + K^* w(k) - e_u(k)].
\end{aligned}$$

The previous equation can be rewritten as

$$\begin{aligned}
&V(e(k+1)) \leq z(k)^\intercal \hat{M} z(k) + \mu_1^* V(e(k)) + \mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 \\
&+ \mu_6^* \sigma_1(K^*)^2 \omega^2
\end{aligned}$$

with $\hat{M}$ as in (5.23).

If $\hat{M} \preceq 0$, we have

$$V(e(k+1)) \leq \mu_1^* V(e(k)) + \mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 + \mu_6^* \sigma_1(K^*)^2 \omega^2.$$

It remains to prove that (5.10) implies that $\hat{M} \preceq 0$. Starting from $\hat{M} \preceq 0$, using Schur's complement once, yields (5.24). Applying Schur's complement twice to move the term $L^2 \mu_2^* H^\intercal H$ leads to (5.25). Using Schur's complement, one obtains (5.26), which after pre and post-multiplication with the following matrix

$$\bar{P} = \begin{bmatrix}
(P^*)^{-1} & & & & & & & \\
& I_n & & & & & & \\
& & I_n & & & & & \\
& & & I_n & & & & \\
& & & & I_n & & & \\
& & & & & I_n & & \\
& & & & & & I_n & \\
& & & & & & & I_n
\end{bmatrix}$$

gives (5.27). Since $(P^*)^{-1} K^* = S^*$, we clearly have that $\hat{M} \preceq 0$ from (5.10). □

Lemma 5.11. *For any $x_1, x_2 \in \mathbb{R}^n$, any $\beta \geq 0$, any $\gamma \geq 0$, the following inequality*

$$V(x_1) \leq \beta V(x_2) + \gamma \tag{5.20}$$

implies

$$\|x_1\|_2 \leq \frac{\sqrt{\beta}\sigma_1(T)}{\sigma_n(T)} \|x_2\|_2 + \frac{1}{\sigma_n(T)}\sqrt{\gamma}.$$

Proof. The inequality (5.20) implies that

$$\sqrt{V(x_1)} \leq \sqrt{\beta}\sqrt{V(x_2)} + \sqrt{\gamma}.$$

Since P^* is symmetric and positive definite, it defines a norm $\|x\|_P \coloneqq \sqrt{V(x)}$. Since $P^* = T^\intercal T$, we have

$$\|x\|_P = \|Tx\|_2 \leq \sigma_1(T) \|x\|_2$$

and

$$\|x\|_2 = \left\|T^{-1}x\right\|_P \leq \frac{1}{\sigma_n(T)} \|x\|_P ,$$

by the usual properties of the singular values of a matrix, and hence

$$\|x_1\|_2 \leq \frac{\sqrt{\beta}\sigma_1(T)}{\sigma_n(T)} \|x_2\|_2 + \frac{1}{\sigma_n(T)}\sqrt{\gamma}.$$

□

$$\begin{bmatrix} -\mu_1 N & \star & \star & \star & \mu_3\bar{\varphi}^2 S^\intercal & \mu_3\bar{\varphi}^2 S^\intercal & NA^\intercal + S^\intercal B^\intercal & \mu_3\bar{\varphi}^2 S^\intercal \\ \star & -Q_2 & \star & \star & \star & \star & I_n & \star \\ \star & \star & -Q_3 & \star & \star & \star & I_n & \star \\ \star & \star & \star & -Q_4 & \star & \star & I_n & \star \\ \mu_3\bar{\varphi}^2 S & \star & \star & \star & -Q_5 + \mu_3\bar{\varphi}^2 I_m & \mu_3\bar{\varphi}^2 I_m & B^\intercal & \star \\ \mu_3\bar{\varphi}^2 S & \star & \star & \star & \mu_3\bar{\varphi}^2 I_m & -Q_6 + \mu_3\bar{\varphi}^2 I_m & B^\intercal & \star \\ AN + BS & I_n & I_n & I_n & B & B & -N + L^2\mu_2 H^\intercal H & \star \\ \mu_3\bar{\varphi}^2 S & \star & \star & \star & \star & \star & \star & \text{-}I_m \end{bmatrix} \preceq 0 \tag{5.21}$$

$M=$

$$\begin{bmatrix} M_1 & (A+BK^*)^\intercal P^* & (A+BK^*)^\intercal P^* & (A+BK^*)^\intercal P^* & (A+BK^*)^\intercal P^* B & (A+BK^*)^\intercal P^* B \\ P^*(A+BK^*) & P^* - Q_2^* & P^* & P^* & P^* B & P^* B \\ P^*(A+BK^*) & P^* & P^* - Q_3^* & P^* & P^* B & P^* B \\ P^*(A+BK^*) & P^* & P^* & P^* - Q_4^* & P^* B & P^* B \\ B^\intercal P^*(A+BK^*) & B^\intercal P^* & B^\intercal P^* & B^\intercal P^* & B^\intercal P^* B - Q_5^* & B^\intercal P^* B \\ B^\intercal P^*(A+BK^*) & B^\intercal P^* & B^\intercal P^* & B^\intercal P^* & B^\intercal P^* B & B^\intercal P^* B - Q_6^* \end{bmatrix} \tag{5.22}$$

$M_1 = (A+K^*B)^\intercal P^*(A+BK^*) - \mu_1^* P^* + L^2\mu_2^* H^\intercal H$

$$\hat{M}=\begin{bmatrix} \hat{M}_1 & (A+BK^*)^\intercal P^* & (A+BK^*)^\intercal P^* & (A+BK^*)^\intercal P^* & \hat{M}_3^\intercal & \hat{M}_2^\intercal \\ P^*(A+BK^*) & P^* - Q_2^* & P^* & P^* & P^* B & P^* B \\ P^*(A+BK^*) & P^* & P^* - Q_3^* & P^* & P^* B & P^* B \\ P^*(A+BK^*) & P^* & P^* & P^* - Q_4^* & P^* B & P^* B \\ \hat{M}_3 & B^\intercal P^* & B^\intercal P^* & B^\intercal P^* & B^\intercal P^* B - Q_5^* + \mu_3^*\bar{\varphi}^2 I_m & B^\intercal P^* B - \mu_3^*\bar{\varphi}^2 I_m \\ \hat{M}_2 & B^\intercal P^* & B^\intercal P^* & B^\intercal P^* & B^\intercal P^* B - \mu_3^*\bar{\varphi}^2 I_m & B^\intercal P^* B - Q_6^* + \mu_3^*\bar{\varphi}^2 I_m \end{bmatrix}$$

$\hat{M}_1 = (A+K^*B)^\intercal P^*(A+BK^*) - \mu_1^* P^* + L^2\mu_2^* H^\intercal H + (\mu_3^*)^2\bar{\varphi}^2 (K^*)^\intercal K^*$

$\hat{M}_2 = B^\intercal P^*(A+BK^*) + \mu_3^*\bar{\varphi}^2 K^*,\ \hat{M}_3 = B^\intercal P^*(A+BK^*) - \mu_3^*\bar{\varphi}^2 K^*$

(5.23)

$$\begin{bmatrix} -\mu_1^* P^* + L^2\mu_2^* H^\intercal H + (\mu_3^*)^2\bar{\varphi}^4 (K^*)^\intercal K^* & \star & \star & \star & -\mu_3^*\bar{\varphi}^2 (K^*)^\intercal & \mu_3^*\bar{\varphi}^2 (K^*)^\intercal & (A+BK^*)^\intercal \\ \star & -Q_2^* & \star & \star & \star & \star & I_n \\ \star & \star & -Q_3^* & \star & \star & \star & I_n \\ \star & \star & \star & -Q_4^* & \star & \star & I_n \\ -\mu_3^*\bar{\varphi}^2 K^* & \star & \star & \star & -Q_5^* + \mu_3^*\bar{\varphi}^2 I_m & -\mu_3^*\bar{\varphi}^2 I_m & B^\intercal \\ \mu_3^*\bar{\varphi}^2 K^* & \star & \star & \star & -\mu_3^*\bar{\varphi}^2 I_m & -Q_6^* + \mu_3\bar{\varphi}^2 I_m & B^\intercal \\ (A+BK^*) & I_n & I_n & I_n & B & B & -(P^*)^{-1} \end{bmatrix} \preceq 0 \tag{5.24}$$

$$\begin{bmatrix} -\mu_1^* P^* + (\mu_3^*)^2\bar{\varphi}^4 (K^*)^\intercal K^* & \star & \star & \star & -\mu_3^*\bar{\varphi}^2 (K^*)^\intercal & \mu_3^*\bar{\varphi}^2 (K^*)^\intercal & A^\intercal + (K^*)^\intercal B^\intercal \\ \star & -Q_2^* & \star & \star & \star & \star & I_n \\ \star & \star & -Q_3^* & \star & \star & \star & I_n \\ \star & \star & \star & -Q_4^* & \star & \star & I_n \\ -\mu_3^*\bar{\varphi}^2 K^* & \star & \star & \star & -Q_5^* + \mu_3^*\bar{\varphi}^2 I_m & -\mu_3^*\bar{\varphi}^2 I_m & B^\intercal \\ \mu_3^*\bar{\varphi}^2 K^* & \star & \star & \star & -\mu_3^*\bar{\varphi}^2 I_m & -Q_6^* + \mu_3^*\bar{\varphi}^2 I_m & B^\intercal \\ A+BK^* & I_n & I_n & I_n & B & B & -(P^*)^{-1} + L^2\mu_2^* H^\intercal H \end{bmatrix} \preceq 0 \tag{5.25}$$

$$\begin{bmatrix} -\mu_1^* P^* & \star & \star & \star & -\mu_3^*\bar{\varphi}^2(K^*)^\intercal & \mu_3^*\bar{\varphi}^2(K^*)^\intercal & A^\intercal+(K^*)^\intercal B^\intercal & \mu_3^*\bar{\varphi}^2(K^*)^\intercal \\ \star & -Q_2^* & \star & \star & \star & \star & I_n & \star \\ \star & \star & -Q_3^* & \star & \star & \star & I_n & \star \\ \star & \star & \star & -Q_4^* & \star & \star & I_n & \star \\ -\mu_3^*\bar{\varphi}^2K^* & \star & \star & \star & -Q_5^*+\mu_3^*\bar{\varphi}^2 I_m & -\mu_3^*\bar{\varphi}^2 I_m & B^\intercal & \star \\ \mu_3^*\bar{\varphi}^2K^* & \star & \star & \star & -\mu_3^*\bar{\varphi}^2 I_m & -Q_6^*+\mu_3^*\bar{\varphi}^2 I_m & B^\intercal & \star \\ A+BK^* & I_n & I_n & I_n & B & B & -(P^*)^{-1}+L^2\mu_2^* H^\intercal H & \star \\ \mu_3^*\bar{\varphi}^2K^* & \star & \star & \star & \star & \star & \star & -I_m \end{bmatrix} \preceq 0 \quad (5.26)$$

$$\begin{bmatrix} -\mu_1^*(P^*)^{-1} & \star & \star & \star & -\mu_3^*\bar{\varphi}^2(P^*)^{-1}(K^*)^\intercal & \mu_3^*\bar{\varphi}^2(P^*)^{-1}(K^*)^\intercal & \hat{M}_4^\intercal & \mu_3^*\bar{\varphi}^2(P^*)^{-1}(K^*)^\intercal \\ \star & -Q_2^* & \star & \star & \star & \star & I_n & \star \\ \star & \star & -Q_3^* & \star & \star & \star & I_n & \star \\ \star & \star & \star & -Q_4^* & \star & \star & I_n & \star \\ -\mu_3^*\bar{\varphi}^2K^*(P^*)^{-1} & \star & \star & \star & -Q_5^*+\mu_3^*\bar{\varphi}^2 I_m & -\mu_3^*\bar{\varphi}^2 I_m & B^\intercal & \star \\ \mu_3^*\bar{\varphi}^2K^*(P^*)^{-1} & \star & \star & \star & -\mu_3^*\bar{\varphi}^2 I_m & -Q_6^*+\mu_3^*\bar{\varphi}^2 I_m & B^\intercal & \star \\ \hat{M}_4 & I_n & I_n & I_n & B & B & \hat{M}_5 & \star \\ \mu_3^*\bar{\varphi}^2K^*(P^*)^{-1} & \star & \star & \star & \star & \star & \star & -I_m \end{bmatrix} \prec 0,$$

$$\hat{M}_4 = A(P^*)^{-1}+BK^*(P^*)^{-1},$$
$$\hat{M}_5 = -(P^*)^{-1}+L^2\mu_2^* H^\intercal H. \quad (5.27)$$

Lemma 5.12. *Let there be $\alpha_0 \geq 0$, $\beta \geq 0$, $\gamma \geq 0$ $k \geq 1$. For any sequence α_i such that*

$$\alpha_i = \beta\alpha_{i-1} + \gamma, \qquad \forall i \geq 1 \tag{5.28}$$

the following holds

$$\alpha_i \leq \max\left\{\alpha_0, \beta^k\alpha_0 + \gamma\sum_{i=0}^{k-1}\beta^i\right\}, \qquad \forall i \in \{0,\dots,k\}.$$

Proof. We start by noticing that (5.28) implies that $\alpha_i = \beta^i\alpha_0 + \gamma\sum_{j=0}^{i-1}\beta^j$. First of all, for $\beta \geq 1$, the result trivially holds (because the series is strictly increasing and hence the last term upper bounds all previous ones) and hence it what follows, we assume $\beta < 1$. In that case, we have that $\alpha_k = \beta^k\alpha_0 + \gamma\frac{1-\beta^k}{1-\beta}$, $\forall k \geq 0$. We consider two cases: $\alpha_0 \leq \frac{\gamma}{1-\beta}$ and $\alpha_0 > \frac{\gamma}{1-\beta}$.

If $\alpha_0 \leq \frac{\gamma}{1-\beta}$, then, since $\beta < 1$, we have

$$\beta^i \geq \beta^k, \qquad \forall i \in \{0,\dots,k\},$$

since $\alpha_0 - \frac{\gamma}{1-\beta} \leq 0$, we have

$$\beta^i\left(\alpha_0 - \frac{\gamma}{1-\beta}\right) \leq \beta^k\left(\alpha_0 - \frac{\gamma}{1-\beta}\right), \qquad \forall i \in \{0,\dots,k\},$$

which, after adding $\frac{\gamma}{1-\beta}$ to both sides yields

$$\beta^i\alpha_0 + \gamma\frac{1-\beta^i}{1-\beta} \leq \beta^k\alpha_0 + \gamma\frac{1-\beta^k}{1-\beta}, \qquad \forall i \in \{0,\dots,k\},$$

and hence

$$\alpha_i \leq \alpha_k, \qquad \forall i \in \{0,\dots,k\}.$$

If $\alpha_0 > \frac{\gamma}{1-\beta}$, then we have

$$\alpha_i = \beta^i\alpha_0 + \gamma\frac{1-\beta^i}{1-\beta}$$

since $\alpha_0 > \frac{\gamma}{1-\beta}$, we have

$$\alpha_i \leq \beta^i\alpha_0 + \alpha_0(1-\beta^i) = \alpha_0,$$

which completes the proof □

The following lemma provides a condition on the choice of η_u such that $\hat{u}(k) \in B_{\eta_u}\left(\frac{\underline{u}+\bar{u}}{2}\right)$.

Lemma 5.13. *Let Assumptions 5.2 to 5.4 hold for the system (5.1). For any $k \geq 0$, any $e_u(k)$ such that $\|e_u(k)\|_2 \leq \epsilon_u$, any $x(k), \hat{x}(k) \in \mathbb{R}^n$, any $w(k), d(k) \in \mathbb{R}^n$ satisfying Assumption 5.4, any $u(k)$ satisfying Assumption 5.2, and for $y(k)$ and $\hat{u}(k)$, such that $y(k) = x(k)+w(k)$ and $\hat{u}(k) = u(k)-K^*(y(k)-\hat{x}(k))+e_u(k)$, choosing η_u as*

$$\eta_u \geq \left\| \frac{\underline{u} - \bar{u}}{2} \right\|_2 + \sigma_1(K^*) \left\| x(k) - \hat{x}(k) \right\|_2 + \sigma_1(K^*)\omega + \epsilon_u \tag{5.29}$$

implies

$$\hat{u}(k) \in B_{\eta_u}\left(\frac{\underline{u} + \bar{u}}{2} \right).$$

Proof. We have

$$\begin{aligned}
&\left\| \hat{u}(k) - \frac{\underline{u} + \bar{u}}{2} \right\|_2 \leq \left\| u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k) - \frac{\underline{u} + \bar{u}}{2} \right\|_2 \\
&\leq \left\| u(k) - \frac{\underline{u} + \bar{u}}{2} \right\|_2 + \left\| K^*(y(k) - \hat{x}(k)) \right\|_2 + \left\| e_u(k) \right\|_2 \\
&\leq \left\| u(k) - \frac{\underline{u} + \bar{u}}{2} \right\|_2 + \left\| K^*(x(k) - \hat{x}(k)) \right\|_2 + \left\| K^* w(k) \right\|_2 \\
&+ \left\| e_u(k) \right\|_2 \\
&\leq \left\| u(k) - \frac{\underline{u} + \bar{u}}{2} \right\|_2 + \sigma_1(K^*) \left\| (x(k) - \hat{x}(k)) \right\|_2 \\
&+ \sigma_1(K^*) \left\| w(k) \right\|_2 + \left\| e_u(k) \right\|_2 \\
&\leq \left\| u(k) - \frac{\underline{u} + \bar{u}}{2} \right\|_2 + \sigma_1(K^*) \left\| x(k) - \hat{x}(k) \right\|_2 + \sigma_1(K^*)\omega + \epsilon_u.
\end{aligned} \tag{5.30}$$

From Assumption 5.2, we have

$$\left\| u(k) - \frac{\underline{u} + \bar{u}}{2} \right\|_2 \leq \left\| \frac{\underline{u} - \bar{u}}{2} \right\|_2$$

which, when combined with (5.30), concludes the proof. □

The next lemma provides a condition on the choices of η_x and α such that $y(k_{j_x})$ will be contained in $B_{\eta_x}(\hat{x}^-(k_{j_x}))$, provided that an estimate of precision ϵ_x was provided at the last communication instant contained the state guess. To this end, we define $\bar{l}(j_x)$: the communication instant before j_x where an estimate of the state was provided.

Lemma 5.14. *Let Assumptions 5.2 to 5.4 hold for the system (5.1). For any $\bar{k} \geq 1$, any $\epsilon_x > 0$, any $\epsilon_u > 0$ such that (5.10) has a solution $(\mu_i^*, N^*, Q_i^*, S^*)$,*

for any $x(k_{\bar{l}(j_x)})$, *any* $w(k_{\bar{l}(j_x)})$, *for* $y(k_{\bar{l}(j_x)})$ *such that* $y(k_{\bar{l}(j_x)}) = x(k_{\bar{l}(j_x)}) + w(k_{\bar{l}(j_x)})$, *for* $\hat{x}(k_{\bar{l}(j_x)})$ *such that* $\left\|y(k_{\bar{l}(j_x)}) - \hat{x}(k_{\bar{l}(j_x)})\right\|_2 \leq \epsilon_x$, *for any* $u(k_{\bar{l}(j_x)} + 1), \ldots, u(k_{j_x} - 1)$, *for* $e_u(k)$, $k \in \{k_{\bar{l}(j_x)}, \ldots, k_{j_x} - 1\}$ *such that* $\|e_u(k)\|_2 \leq \epsilon_u$, *for* $\hat{u}(k)$, $k \in \{k_{\bar{l}(j_x)}, \ldots, k_{j_x} - 1\}$, *such that* $\hat{u}(k) = u(k) - K^*(y(k) - \hat{x}(k)) + e_u(k)$, $\forall k \in \{k_{\bar{l}(j_x)}, \ldots, k_{j_x} - 1\}$, *choosing* α *and* η_x *as*

$$\begin{aligned}\eta_x \geq & \frac{(\mu_1^*)^{\frac{\bar{k}}{2}}\,\sigma_1(T)}{\sigma_n(T)}(\epsilon_x + \omega) + \omega \\ & + \frac{1}{\sigma_n(T)}(\mu_4^*\delta^2 + \mu_5^*\epsilon_u^2 + \mu_6^*\sigma_1(T)^2\omega^2)^{\frac{1}{2}} \sum_{i=0}^{\bar{k}-1}(\mu_1^*)^{\frac{i}{2}}\end{aligned} \tag{5.31}$$

and

$$\begin{aligned}\alpha \geq & \\ & \min\left\{\frac{\sigma_n(T)[\eta_x - \omega - \left(\mu_4^*\delta^2 + \mu_5^*\epsilon_u^2 + \mu_6^*\sigma_1(K^*)^2\omega^2\right)^{\frac{1}{2}}]}{\sqrt{\mu_1^*}\sigma_1(T)} - \omega,\right. \\ & \max\left\{\epsilon_x + \omega, \frac{(\mu_1^*)^{\frac{\bar{k}-1}{2}}\,\sigma_1(T)}{\sigma_n(T)}(\epsilon_x + \omega)\right. \\ & \left.\left.\frac{1}{\sigma_n(T)}(\mu_4^*\delta^2 + \mu_5^*\epsilon_u^2 + \mu_6^*\sigma_1(K^*)^2\omega^2)^{\frac{1}{2}} \sum_{i=0}^{\bar{k}-2}(\mu_1^*)^{\frac{i}{2}}\right\}\right\}.\end{aligned} \tag{5.32}$$

implies

$$y(k_{j_x}) \in B_{\eta_x}(\hat{x}^-(k_{j_x})).$$

Proof. We distinguish two different cases: the first one is the situation where $j_x = \bar{l}(j_x) + \bar{k}$, the second is the situation where $j_x > \bar{l}(j_x) + \bar{k}$.

Situation 1: Let there be k_{j_x} such that $j_x = \bar{l}(j_x) + \bar{k}$. By the theorem statement, at $k_{\bar{l}(j_x)}$, we have

$$\begin{aligned}\left\|x\left(k_{\bar{l}(j_x)}\right) - \hat{x}^-\left(k_{\bar{l}(j_x)}\right)\right\|_2 &= \left\|x\left(k_{\bar{l}(j_x)}\right) - \hat{x}\left(k_{\bar{l}(j_x)}\right)\right\|_2 \\ &\leq \epsilon_x + \omega.\end{aligned}$$

Applying Lemma 5.10 $\bar{k}$ times, we obtain

$$\begin{aligned}& V(x(k_{\bar{l}(j_x)} + \bar{k}) - \hat{x}(k_{\bar{l}(j_x)} + \bar{k})) \\ & \leq (\mu_1^*)^{\bar{k}} V\left(x\left(k_{\bar{l}(j_x)}\right) - \hat{x}\left(k_{\bar{l}(j_x)}\right)\right) \\ & + (\mu_4^*\delta^2 + \mu_5^*\epsilon_u^2 + \mu_6^*\sigma_1(K^*)^2\omega^2)\sum_{i=0}^{\bar{k}}(\mu_1^*)^i.\end{aligned}$$

Applying Lemma 5.11, this implies

$$\left\|x(k_{\bar{l}(j_x)}+\bar{k})-\hat{x}(k_{\bar{l}(j_x)}+\bar{k})\right\|_2 \leq \frac{(\mu_1^*)^{\frac{\bar{k}}{2}}\sigma_1(T)}{\sigma_n(T)}(\epsilon_x+\omega)$$
$$+\frac{1}{\sigma_n(T)}(\mu_4^*\delta^2+\mu_5^*\epsilon_u^2+\mu_6^*\sigma_1(K^*)^2\omega^2)^{\frac{1}{2}}\sum_{i=0}^{\bar{k}-1}(\mu_1^*)^{\frac{i}{2}}.$$

And thus

$$\left\|y(k_{j_x})-\hat{x}^-(k_{j_x})\right\|_2 \leq \|y(k_{j_x})-x(k_{j_x})\|_2$$
$$+\left\|x(k_{j_x})-\hat{x}^-(k_{j_x})\right\|_2 \leq \omega+\frac{(\mu_1^*)^{\frac{\bar{k}}{2}}\sigma_1(T)}{\sigma_n(T)}(\epsilon_x+\omega)$$
$$+\frac{1}{\sigma_n(T)}(\mu_4^*\delta^2+\mu_5^*\epsilon_u^2+\mu_6^*\sigma_1(K^*)^2\omega^2)^{\frac{1}{2}}\sum_{i=0}^{\bar{k}-1}(\mu_1^*)^{\frac{i}{2}},$$

which implies $y(k_{j_x}) \in B_{\eta_x}(\hat{x}^-(k_{j_x}))$.

Situation 2: Since $j_x > \bar{l}(j_x)+\bar{k}$, the triggering condition

$$\|y(k_{j_x}-1)-\hat{x}(k_{j_x}-1)\|_2 \leq \alpha$$

necessarily holds (otherwise, a communication would have been triggered at the previous time instant). We have

$$\left\|x(k_{j_x}-1)-\hat{x}^-(k_{j_x}-1)\right\|_2$$
$$\leq \|y(k_{j_x}-1)-x(k_{j_x}-1)\|_2+\left\|y(k_{j_x}-1)-\hat{x}^-(k_{j_x}-1)\right\|_2$$
$$\leq \omega+\left\|y(k_{j_x}-1)-\hat{x}^-(k_{j_x}-1)\right\|_2 \leq \omega+\alpha.$$

Applying Lemma 5.10 and Lemma 5.11 together we have

$$\left\|x(k_{j_x})-\hat{x}^-(k_{j_x})\right\|_2 \leq \frac{\sigma_1(T)\sqrt{\mu_1^*}}{\sigma_n(T)}(\omega+\alpha)$$
$$+\frac{1}{\sigma_n(T)}(\mu_4^*\delta^2+\mu_5^*\epsilon_u^2+\mu_6^*\sigma_1(K^*)^2\omega^2)^{\frac{1}{2}}$$

which, after replacing α with the right-hand side of (5.32) and some computations, yields

$$\left\|x(k_{j_x})-\hat{x}^-(k_{j_x})\right\|_2 \leq \eta_x-\omega,$$

which implies $y(k_{j_x}) \in B_{\eta_x}(\hat{x}^-(k_{j_x}))$. □

5.A.1 Proof of Theorem 5.8

Proof. At first, from (5.2), we have

$$\|x(0) - \hat{x}(0)\|_2 \leq \epsilon_0,$$

and thus from the theorem statement, it immediately follows that

$$\|x(0) - \hat{x}(0)\|_2 \leq \epsilon_x + \omega.$$

Lemma 5.10 implies

$$\begin{aligned} V(e(1)) \leq & \frac{\sqrt{\mu_1^*}\sigma_1(T)}{\sigma_n(T)} V(e(0)) \\ & + \frac{1}{\sigma_2(T)}(\mu_4^*\delta^2 + \mu_5^*\epsilon_u^2 + \mu_6^*\sigma_1(K^*)^2\omega^2)^{\frac{1}{2}}. \end{aligned}$$

and 5.11 thus implies that

$$\begin{aligned} \|x(1) - \hat{x}(1)\|_2 \leq & \frac{\sqrt{\mu_1^*}\sigma_1(T)}{\sigma_n(T)}(\epsilon_x + \omega) \\ & + \frac{1}{\sigma_2(T)}(\mu_4^*\delta^2 + \mu_5^*\epsilon_u^2 + \mu_6^*\sigma_1(K^*)^2\omega^2)^{\frac{1}{2}}. \end{aligned}$$

From (5.13) and Lemma 5.13, we have that $\hat{u}(1) \in B_{\eta_u}\left(\frac{\underline{u}+\bar{u}}{2}\right)$ and hence we can apply Lemma 5.10, again. From there on, proceeding sequentially for $j \in \{2, \ldots, \bar{k}\}$. Every time, we apply Lemma 5.10, Lemma 5.11 and Lemma 5.12 to obtain that

$$\begin{aligned} \|x(l) - \hat{x}(l)\|_2 \leq \max \Bigg\{ & \epsilon_x + \omega, \frac{\left(\sqrt{\mu_1^*}\right)^j \sigma_1(T)}{\sigma_n(T)}(\epsilon_x + \omega) \\ & + \frac{1}{\sigma_n(T)}(\mu_4^*\delta^2 + \mu_5^*\epsilon_u^2 + \mu_6^*\sigma_1(K^*)^2\omega^2)^{\frac{1}{2}} \sum_{i=0}^{j-1} (\mu_1^*)^{\frac{i}{2}} \Bigg\}, \end{aligned}$$

$\forall l \in \{1, \ldots, j\}$ The parameters chosen in (5.13) and Lemma 5.13 then imply that $\hat{u}(j) \in B_{\eta_u}\left(\frac{\underline{u}+\bar{u}}{2}\right)$ which implies that the same can be repeated until $j = \bar{k} - 1$ at which point, we have

$$\begin{aligned} \|x(j) - \hat{x}(j)\|_2 \leq \max \Bigg\{ & \epsilon_x + \omega, \frac{\left(\sqrt{\mu_1^*}\right)^{\bar{k}-1} \sigma_1(T)}{\sigma_n(T)}(\epsilon_x + \omega) \\ & + \frac{1}{\sigma_n(T)}(\mu_4^*\delta^2 + \mu_5^*\epsilon_u^2 + \mu_6^*\sigma_1(K^*)^2\omega^2)^{\frac{1}{2}} \sum_{i=0}^{\bar{k}-2} (\mu_1^*)^{\frac{i}{2}} \Bigg\}. \end{aligned}$$

$\forall j \in \{0, \ldots, \bar{k}-1\}$. At $j = \bar{k}$, and at all subsequent time instants, if no communication occurs at $k = \bar{k}$, $\|x(j) - \hat{x}(j)\|_2 \leq \alpha$ is verified, otherwise a communication would occur. From (5.15), we thus have

$$\|x(j) - \hat{x}(j)\|_2 \leq \max \left\{ \epsilon_x + \omega, \frac{\left(\sqrt{\mu_1^*}\right)^{\bar{k}-1} \sigma_1(T)}{\sigma_n(T)} (\epsilon_x + \omega) \right.$$
$$\left. + \frac{1}{\sigma_n(T)} (\mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 + \mu_6^* \sigma_1 (K^*)^2 \omega^2)^{\frac{1}{2}} \sum_{i=0}^{\bar{k}-2} (\mu_1^*)^{\frac{i}{2}} \right\}$$

Finally, at $k = k_{j_x}$, we use Lemma 5.14, so that $y(k_{j_x}) \in B_{\eta_x}(\hat{x}^-(k_{j_x}))$ and thus an estimate of precision ϵ_x is transmitted, which resets the error to $\epsilon_x + \omega$. From there on, the observation scheme simply repeats the same steps in between two communications of an estimate of the state and hence

$$\|x(k) - \hat{x}(k)\|_2 \leq \max \left\{ \epsilon_x + \omega, \frac{\left(\sqrt{\mu_1^*}\right)^{\bar{k}-1} \sigma_1(T)}{\sigma_n(T)} (\epsilon_x + \omega) \right.$$
$$\left. + \frac{1}{\sigma_n(T)} (\mu_4^* \delta^2 + \mu_5^* \epsilon_u^2 + \mu_6^* \sigma_1 (K^*)^2 \omega^2)^{\frac{1}{2}} \sum_{i=0}^{\bar{k}-2} (\mu_1^*)^{\frac{i}{2}} \right\}$$

$\forall k \geq 0$. □

5.A.2 Proof of Theorem 5.9

Proof. Although the messages of the communication protocol can possibly change every instant, the length of their associated alphabets can only be two values: $|\mathcal{V}|$ or $|\mathcal{V}||\mathcal{W}|$. The first case corresponds to when only the driving signal of the estimate is communicated while the second corresponds to when both the driving signal of the estimate and an estimate of the state are communicated. Let us call the first situation (a) and the second (b). If (a) occurs for a particular k_j, then we have $b_j = \lceil \log_2 |\mathcal{V}| \rceil$. If (b) occurs for a particular k_j, then we have $b_j = \lceil \log_2 |\mathcal{V}||\mathcal{W}| \rceil$. At each time instant, either (a) or (b) occurs. Let $\#_b(\bar{j})$ be the number of times that the situation (b) has occurred up to the communication

instant $k_{\bar{j}}$. We have

$$\begin{aligned}\frac{\sum_{j=0}^{\bar{j}} b_j}{\bar{j}+1} =& \#_b(\bar{j})\frac{\lceil \log_2 |\mathcal{V}||\mathcal{W}| \rceil}{\bar{j}+1} + (\bar{j}+1-\#_b(\bar{j}))\frac{\lceil \log_2 |\mathcal{V}| \rceil}{\bar{j}+1} \\ \leq& \#_b(\bar{j})\frac{\lceil \log_2 |\mathcal{V}| \rceil}{\bar{j}+1} + \#_b(\bar{j})\frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{j}+1} \\ &+ (\bar{j}+1-\#_b(\bar{j}))\frac{\lceil \log_2 |\mathcal{V}| \rceil}{\bar{j}+1} \\ =& \lceil \log_2 |\mathcal{V}| \rceil + \#_b(\bar{j})\frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{j}+1}.\end{aligned}$$

Since at least $\bar{k}$ time instants elapse between two successive communications of $\hat{x}$ (or events (b)), we have $\#_b(\bar{j}) \leq \frac{\bar{j}+1}{\bar{k}} + 1$, $\forall \bar{j} \geq 0$. This implies that

$$\begin{aligned}\frac{\sum_{j=0}^{\bar{j}} b_j}{\bar{j}+1} &\leq \lceil \log_2 |\mathcal{V}| \rceil + \left(\frac{\bar{j}+1}{\bar{k}} + 1\right)\frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{j}+1} \\ &= \lceil \log_2 |\mathcal{V}| \rceil + \frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{k}} + \frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{j}+1}.\end{aligned}$$

Starting from the definition of R, we thus have

$$\begin{aligned}R \coloneqq& \limsup_{\bar{j}\to\infty} \frac{\sum_{j=0}^{\bar{j}} b_j}{\bar{j}+1} \\ \leq& \limsup_{\bar{j}\to\infty} \left[\lceil \log_2 |\mathcal{V}| \rceil + \frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{k}} + \frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{j}+1}\right] \\ =& \lceil \log_2 |\mathcal{V}| \rceil + \frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{k}} + \limsup_{\bar{j}\to\infty} \frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{j}+1} \\ =& \lceil \log_2 |\mathcal{V}| \rceil + \frac{\lceil \log_2 |\mathcal{W}| \rceil}{\bar{k}}.\end{aligned}$$

The set $\mathcal{V}$ is a covering of a ball of radius η_u with balls of radius ϵ_u. One possible solution to obtain such a covering is as follows. In any ball of radius ϵ_u, there is a cube of side $\epsilon_u/\sqrt{m}$ inscribed. Any ball of radius η_u is inscribed in a hypercube of side $2\eta_u$. It is thus possible to cover $B_{\eta_u}\left(\frac{u+\bar{u}}{2}\right)$ with $\left(\frac{2\eta_u}{\frac{\epsilon_u}{\sqrt{m}}}\right)^m$ balls. By an identical reasoning, at most $\left(\frac{2\eta_x}{\frac{\epsilon_x}{\sqrt{n}}}\right)^n$ balls are required to cover $B_{\eta_x}(0)$ and hence

$$R \leq \left\lceil m \log_2 \frac{2\eta_u\sqrt{m}}{\epsilon_u} \right\rceil + \frac{\left\lceil n \log_2 \frac{2\eta_x\sqrt{n}}{\epsilon_x} \right\rceil}{\bar{k}}.$$

□

Chapter 6

Observing a Unicycle Robot with Data Rate Constraints: A Case Study

In this chapter, we consider the problem of remote observation of a unicycle-type mobile robot through a data rate constrained communication channel, which can only send a limited number of bits per unit of time. The objective is to reconstruct estimates of the state of the robot at the remote location through the messages that are sent. The design of the communication protocol should ensure that the maximum observation error is bounded whilst using as few bits per unit of time as possible. An event-triggered observation scheme is developed specifically for the unicycle-type robot. This observer is tested through experiments on Turtlebots. The experiments show that the event-triggered scheme is very efficient at reducing the average number of required communications.

6.1 Introduction

As wireless communication technologies have become omnipresent in modern society, the world of dynamics and control has been taken over by them as well. There are many different applications where one or several dynamical systems or the components thereof are connected via data rate constrained communication channels. Examples include: cooperative load displacement by robots, underwater communication of autonomous vehicles, formation control for drones, cooperative cruise control, etc. All these problems share the common feature that some source of uncertainty (in the sense of Shannon [1948]) makes it necessary to communicate over a communication channel that is limited either in the transmission rate of packets, the size of packets, or both. The sources of uncertainty include perturbations, noise, parametric uncertainty, and sensitivity to initial conditions. Although the problem of communication can be seen sep-

arately from the control/observation associated with the underlying dynamical systems, a combined approach allows one to provide better results, be it in terms of control performance, error bounds, sufficient rates, etc.

Among the many works in this field, we note the pioneering contributions Wong and Brockett [1997], Elia and Mitter [2001] which provide results for linear systems. Many other results were obtained for linear systems and broad overviews of such results can be found in Baillieul and Antsaklis [2007], Hespanha et al. [2007], Andrievsky et al. [2010]. For nonlinear systems, early important results include Nair et al. [2004] and Liberzon and Hespanha [2005], which use the concept of entropy to characterize the minimum data rates. Many more papers exploited entropy-based techniques to provide constructive bounds on the sufficient/necessary data rates (see Kawan [2013], Matveev and Savkin [2009], Kawan [2018], Sibai and Mitra [2017], Liberzon and Mitra [2016], Matveev and Pogromsky [2016], Sibai and Mitra [2018], Voortman et al. [2019] and Matveev and Pogromsky [2019]).

At the same time as data rate constrained control appeared in the literature, another topic emerged: event-triggered control. Two of the earliest works in this field include Åarzén [1999] and Åström and Bernhardsson [1999]. An introduction to event-based control can be found in Heemels et al. [2012] and an overview of sampling-related results in Hetel et al. [2017]. Event-triggered control and data rate constrained control have recently been combined in some works of the literature. Examples of such works include: Han et al. [2015] (event-triggered sensor schedule for remote estimation for a linear system), Trimpe [2017] (distributed state estimation with data rate constraints), Xia et al. [2017] (networked state estimation with a shared communication medium) and, Muehlebach and Trimpe [2018] (LMI approach is used for the networked state estimation problem over a shared communication medium).

In this chapter, we present the theory behind a data rate constrained observer for a unicycle robot, which is modeled by a nonlinear dynamical system and has already been studied extensively (see e.g. Jiang and Nijmeijer [1997], Evers and Nijmeijer [2006] or Kostic et al. [2009]). The robot is equipped with a smart sensor, capable of measuring the position and orientation of the robot and performing computations. It is connected to a remote location via a communication channel. The smart sensor can send messages over this communication channel to the remote location to provide an estimate of the position of the robot at the remote location. The particularity of the communication channel is that it is restricted in terms of data rates that it can transmit. The objective is to develop a communication protocol such that it is possible to reconstruct the position of the robot at a remote location whilst using limited data rates. The novelty of the result and the main contribution of this work consists of using an event-triggered communication protocol which often greatly reduces the communication rate, as is proven through experiments on mobile robots. To the best of the authors' knowledge, no fundamental bounds on the minimum sufficient

capacity to observe unicycle-type robots have been obtained in the literature.

6.2 Problem statement

We consider a unicycle-type robot with the following dynamics:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{\theta}(t) \end{bmatrix} = \begin{bmatrix} (u_x(t) + d_x(t))\cos(\theta(t)) \\ (u_x(t) + d_x(t))\sin(\theta(t)) \\ u_\theta(t) + d_\theta(t) \end{bmatrix}, \qquad \forall t \geq 0, \tag{6.1}$$

where $x = (x_1, x_2) \in \mathbb{R}^2$ is the position of the robot, $\theta \in S^1$ is the orientation angle of the robot, $u = (u_x, u_\theta) \in \mathbb{R}^2$ is the input, $d_x \in \mathbb{R}$ and $d_\theta \in \mathbb{R}$ are time-varying input perturbations. The system's output y consists of the full state, sampled with sampling interval $\bar{t}$,

$$y(t_k) = \begin{bmatrix} x_1(t_k) \\ x_2(t_k) \\ \theta(t_k) \end{bmatrix}, \qquad \forall k \geq 0, \tag{6.2}$$

where $t_k = \bar{t}k$ are the sampling instants. We assume that the input perturbations are continuous signals that verify

$$|d_i(t)| \leq \delta_i, \tag{6.3}$$

$\forall i \in \{x, \theta\}, \forall t \geq 0$ where δ_i are the maximum input perturbations, which are known constants.

The system is connected to a remote location via a data rate constrained communication channel. Several devices interact to send messages $m(s_j)$ (where s_j are the transmission times) over this communication channel. These devices are a smart sensor (a sensor admitting some computational capacities, which allows it to perform additional computations on the measured data), an alphabet function $\mathcal{A}$, and a decoder $\mathcal{D}$. The smart sensor is further sub-divided into a sampler $\mathcal{S}$ and a coder $\mathcal{C}$. Together, the devices form a communication protocol. The smart sensor sends messages over the communication channel to the decoder which interprets the messages to generate an estimate $\hat{x}(t)$ of the position $x(t)$ of the robot. Fig. 6.1 depicts how the different components interact. Note that only an estimate of the position should be generated at the remote location.

The sensor and the decoder share the common knowledge of an initial estimate $\hat{x}_0$ which satisfies

$$\|x(0) - \hat{x}_0\|_2 \leq \epsilon_0, \tag{6.4}$$

where ϵ_0 is a user-specified parameter corresponding to the error in initial conditions and $\|\cdot\|_2$ is to the Euclidean norm in $\mathbb{R}^n$. We define the following metric

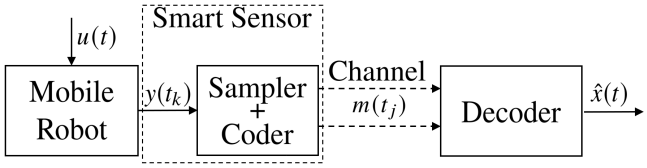

Figure 6.1: Structure of the interactions between the different components of the problem.

on S^1

$$d_S(\theta_1, \theta_2) \coloneqq \begin{cases} |\theta_1 - \theta_2|, & \text{if } |\theta_1 - \theta_2| \leq \pi, \\ 2\pi - |\theta_1 - \theta_2|, & \text{else.} \end{cases}$$

The sensor and remote location also have an initial estimate $\hat{\theta}_0$, which satisfies

$$d_S(\theta(0), \hat{\theta}_0) \leq \epsilon_{\theta_0}. \tag{6.5}$$

The knowledge of the following quantities is shared by all devices: the maximum input perturbations δ_x and δ_θ, the discretization errors ϵ_x and ϵ_θ (which are induced by coding/decoding operation and for brevity of exposition in this document the constants $\epsilon_0 = \epsilon_x$ and $\epsilon_{\theta_0} = \epsilon_\theta$), and the initial estimates $\hat{x}_0$ and $\hat{\theta}_0$.

At the system side, the sampler $\mathcal{S}$ generates the instants of transmission in the following way

$$s_{j+1} = \mathcal{S}(s_j, \{y(t_k)\}_{k:t>t_k\geq 0}, m(s_1), \ldots, m(s_j)), \tag{6.6}$$

$s_0 = 0$, $m(0) = \emptyset$, with the restriction that $s_{j+1} > s_j$. The coder then generates the messages in the following way

$$m(s_j) = \mathcal{C}(\{y(t_k)\}_{k:s_j>t_k\geq 0}, m(s_1), \ldots, m(s_{j-1})), \tag{6.7}$$

$\forall s_j : j > 0$. At each communication instant, the list of different possible messages is encoded into a finite-sized alphabet (the finite cardinality is necessary because of the data rate constraints). The alphabet function $\mathcal{A}$ determines the last index of the messages l_j in the following way

$$l_j = \mathcal{A}(m(s_1), \ldots, m(s_j)), \qquad \forall s_j : j > 0. \tag{6.8}$$

The restriction on the choice of messages is $m(s_j) \in \{1, \ldots, l_j\}$, $\forall s_j : j > 0$. At the remote location, the decoder $\mathcal{D}$ receives the messages and interprets them to generate an estimate of the state $\hat{x}(t)$ in the following way

$$\hat{x}(t) = \mathcal{D}(m(s_1), \ldots, m(s_j)), \qquad \forall t \in [s_j, s_{j+1}),$$

$\forall j \geq 0$. The number of bits b_j required to encode the messages depends on the length of the alphabet. In practice, this implies that

$$b_j \coloneqq \lceil \log_2 l_j \rceil \qquad \forall s_j : j > 0, \tag{6.9}$$

where $\lceil \cdot \rceil$ is the ceiling function (which rounds any real number to the smallest integer larger than or equal to that number). The **communication rate** R resulting from the transmission of these messages is defined as

$$R \coloneqq \lim_{j \to \infty} \frac{1}{s_j} \sum_{i=1}^{j} b_i. \tag{6.10}$$

The rate is thus defined such that it is the average overall communication instants of the number of bits that are sent. Because of the perturbation, measurement error, and finite communication rate, it is impossible to generate estimates of the state at the remote location with zero error. Instead, the design of the communication protocol should ensure that the **maximum observation error** η is bounded, where

$$\eta \coloneqq \sup_{t \geq 0} \left\| x(t) - \hat{x}(t) \right\|_2 . \tag{6.11}$$

Note that the maximum observation error only concerns the position of the robot and not its angular orientation. The first objective of this chapter is to design a communication protocol, in the form of a Sampler $\mathcal{S}$, Coder $\mathcal{C}$, Alphabet $\mathcal{A}$ and Decoder $\mathcal{D}$ such that the maximum observation error and the rate are bounded, and preferably as small as possible. The second objective is to test this communication protocol through experiments on mobile robots.

6.3 Designing the Observer

In this section, we introduce the different agents of the communication protocol. Before we describe the communication protocol and due to page number limitations, we pose the following simplifying assumption about the input.

Assumption 6.1. *The inputs $u_x(t)$ and $u_\theta(t)$ are constant, i.e. $u_x(t) = \bar{u}_x$, $u_\theta(t) = \bar{u}_\theta$, $\forall t \geq 0$ and known.*

Remark 6.2. *This hypothesis can be relaxed to the requirement that the inputs are piece-wise constant keeping their values over an interval of time, in which case the input values have to be also communicated. For more information about how to transmit estimates of the input, one can refer to Voortman et al. [2020c].*

We also introduce the system

$$\begin{bmatrix} \dot{\hat{x}}_1(t) \\ \dot{\hat{x}}_2(t) \\ \dot{\hat{\theta}}(t) \end{bmatrix} = \begin{bmatrix} \bar{u}_x \cos(\hat{\theta}(t)) \\ \bar{u}_x \sin(\hat{\theta}(t)) \\ \bar{u}_\theta \end{bmatrix} \tag{6.12}$$

which corresponds to (6.1) with zero perturbations. Note that for any initial condition $[\hat{x}_{10}\ \hat{x}_{20}\ \hat{\theta}_0]^\intercal$, (6.12) has the following exact solution

$$\begin{aligned} \hat{x}_1(t) &= \frac{\bar{u}_x}{\bar{u}_\theta}\Big(\sin(\bar{u}_\theta t + \hat{\theta}_0) - \sin(\hat{\theta}_0)\Big) + \hat{x}_{10}, \\ \hat{x}_2(t) &= \frac{\bar{u}_x}{\bar{u}_\theta}\Big(-\cos(\bar{u}_\theta t + \hat{\theta}_0) + \cos(\hat{\theta}_0)\Big) + \hat{x}_{20}, \\ \hat{\theta}(t) &= \bar{u}_\theta t + \hat{\theta}_0. \end{aligned} \tag{6.13}$$

The main mechanism of the communication protocol can be described as follows: at the sensor side, the sampled state $y(t_k)$ is measured at sampling instants t_k. The smart sensor simulates a virtual copy of the decoder, which is possible since the sensor generates the messages that the decoder receives. The sensor is thus aware of a copy of the remote estimate, which will be denoted $\hat{x}_c(t)$. The sensor sends messages which contain information necessary to reconstruct both $\hat{x}(s_j)$, as well as $\hat{\theta}(s_j)$. Starting at the estimates stemming from the last message and in the absence of messages, the decoder simply updates the estimate by computing the solution of (6.12) with $\hat{x}(s_j)$ and $\hat{\theta}(s_j)$ as an initial condition. If at some sampling instant t_k, the distance between $y(t_k)$ and $\hat{x}_c(t_k) = \hat{x}(t_k)$ becomes larger than some prescribed maximum error, the sampler decides to communicate: it sets $s_j = t_k$ and the coder then sends a message to the decoder to provide new estimates $\hat{x}(t_j)$ and $\hat{\theta}(t_j)$.

The communication procedure, which we will further reference as Procedure 4, is composed of a sampler, alphabet, coder, and decoder as described below. Although the problem statement leaves room for different choices of communication instants, our communication protocol will choose communication instants such that they coincide with sampling instants. At least N sampling instants will need to elapse between two consecutive communications, where N is a choosable parameter. This parameter is finite and a part of the communication protocol, which implies that it is known by all interacting agents. The choice of N directly influences the maximum observation error and resulting rate. How exactly one might choose N and how it influences the error will be discussed in the next section.

To properly describe the communication instants, we will need several quantities. The indexes j of the communication instants are inherently known by all agents. The quantity $\bar{j}$ refers to the index of the last instant of communication (initially, $\bar{j} = 0$). This quantity is always known by the sampler (because it knows how many communication instants it defined), the coder (because it

knows how many messages it sent), as well as the decoder (because it knows how many messages it received). Finally, the sampler and coder interact to update the knowledge of the estimate $\hat{x}(t)$ at the coder side.

Procedure 4.
The Sampler $\mathcal{S}$: At each sampling instant $t_k \geq s_{\bar{j}} + N\bar{t}$, the sampler computes $\hat{x}_c(t_k)$ and then verifies whether the following condition is satisfied

$$\|y(t_k) - \hat{x}_c(t_k)\|_2 \leq \epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t} - 2|\bar{u}_x|\bar{t} - 2\delta_x\bar{t}. \tag{6.14}$$

If the condition is not satisfied, a message must be sent to provide a new estimate. The sampler thus sets $s_{\bar{j}} = t_k$ and $\bar{j}$ increases by 1.

The Alphabet Function $\mathcal{A}$: If $t_k = s_{\bar{j}}$, the alphabet agent builds a covering of the set $I_{\bar{j}}$, where $I_{\bar{j}}$ is defined as

$$\begin{aligned} I_{\bar{j}} := \big\{ x \in \mathbb{R}^2 | \, \big\|x - \hat{x}_c(s_{\bar{j}-1})\big\|_2 \leq \epsilon_x + 2(|\bar{u}_x| + \delta_x)N\bar{t}, \\ \big\|x - \hat{x}_c(s_{\bar{j}-1})\big\|_2 \geq \epsilon_x + 2(|\bar{u}_x| + \delta_x)N\bar{t} - 2(|\bar{u}_x| + \delta_x)\bar{t} \big\}, \end{aligned} \tag{6.15}$$

with disks of radius ϵ_x. The disks in this covering are numbered from 1 till $l^x_{\bar{j}}$. The coder also divides S^1 into intervals of size $2\epsilon_\theta$. The these intervals are numbered from 1 till $l^\theta_{\bar{j}}$. The alphabet function then returns $l_{\bar{j}} = l^x_{\bar{j}} l^\theta_{\bar{j}}$.

The Coder $\mathcal{C}$: At the communication instants, the coder function finds the index of the ball in the covering made by the alphabet whose center is the closest to $x(s_{\bar{j}})$. The coder also finds the index of the interval in the covering of S^1 which contains $\theta(s_{\bar{j}})$. It then sends both indexes to the decoder. The coder also updates the local estimates $\hat{x}_c(s_{\bar{j}})$ and $\hat{\theta}_c(s_{\bar{j}})$ by setting them to be equal to the center of the ball in I_j.

The Decoder $\mathcal{D}$: In the absence of messages, the decoder computes $\hat{x}(t)$ and $\hat{\theta}(t)$ as solutions of (6.12) with $\hat{x}(t_{\bar{j}})$ and $\hat{\theta}(t_{\bar{j}})$ as an initial conditions. If a message is received, the decoder uses the center of the disk whose index it received as $\hat{x}(t_{\bar{j}})$ and the center of the interval whose index it received as $\hat{\theta}(t_{\bar{j}})$.

The alphabet is based on the following idea. As was previously mentioned, in the absence of messages, new estimates are obtained at the decoder side by solving (6.12). After receiving a message, the state of the system $x(t)$ is contained in a ball of a radius ϵ_x whose center is the estimate $\hat{x}(t)$. In the absence of any messages, the distance between the state and the estimate increases/decreases (both cases are possible). The distance between the state and estimate evolves due to two factors: first of all, the unknown state perturbation $d_x(t)$ and $d_\theta(t)$ increase it continuously (but no more than δ_x every second). Secondly, the distance set is increased/decreased by the action of the system dynamics. Given that the communication intervals are chosen to be finite (i.e., $N < \infty$), the distance remains finite in between communications. Supposing that the set I_j is defined in such a way that it always contains the current state at the

communication instants, then by covering this set with balls of radius ϵ_x, the centers of the balls of which cover I_j form an alphabet to communicate estimates $\hat{x}$.

In this case, the set I_j, which is simply a ring centered around the previous estimate, can be covered by a finite number of balls of size $\epsilon_x > 0$. The balls in the covering can be indexed from 1 till $l_{\max} < \infty$. To produce such a covering, the only information needed is the initial ball and the different upper bounds on the uncertainties/errors, which implies that both the coder as well as the decoder can build the set. To transmit a new estimate of x, one can simply send the index of one of the balls whose center then serves as a new estimate with a precision that will depend on ϵ_x. The cost of communicating in that fashion is dependent on how many balls of size ϵ_x are required to cover I_j. We finish the current section with several remarks on the different features of the proposed scheme.

Remark 6.3.

- *The copies of the estimates at the sensor side $\hat{x}_c(t)$ and $\hat{\theta}_c(t)$ are either updated by the sampler if no message is sent, or by the coder if a message is sent.*
- *The coordinates of the centers of the balls used in the covering are always relative to the previous estimates. By communicating in relative fashion, it is possible to keep the size of the messages limited. Note that since $\hat{x}_c(t) = \hat{x}(t)$, both agents can build this set according to its definition (6.15).*
- *The alphabet procedure is easy from a computational point of view since it consists of covering one set which always has the same shape except the whole set is shifted by a certain vector from the origin and another set which is simply S^1. The first of these sets is centered around the previous estimate, both the coder and decoder can build a covering for it and thus have access to the alphabet.*
- *The computational requirement on the decoder are relatively low since an exact solution (6.13) exists to (6.12).*

6.4 Rate and Errors

With the observer and its agents fully introduced, we are in a position to determine a bound on the communication rate resulting from the observer. This quantity is related to the observation error, for which we also provide a theoretical bound. The first result provides a closed-form expression of the upper bound of the total estimation error that indicates the proportionality of the different parameters/errors.

Proposition 6.4. *The observer described in Procedure 4 ensures that*

$$\eta \leq \epsilon_x + 2(|\bar{u}_x| + \delta_x)N\bar{t}, \tag{6.16}$$

where η is defined in (6.11).

The proof of this result is available in Appendix 6.A.

Remark 6.5. *The use of Proposition 6.4 is straightforward. Based on the velocities $\bar{u}_x$ and $\bar{u}_\theta$, as well as the bounds on the perturbations δ_x and δ_θ, the user of the communication scheme has an expression that links the minimum number of sampling intervals between communications, the discretization error ϵ_x and the total error. Depending on the maximum tolerable observation error, it is thus simple to find N such that the error will not exceed that bound.*

The next result of this section aims to provide an upper bound on R for the designed communication scheme.

Theorem 6.6. *The observer described in Procedure 4 with $N > 0$ results in a communication rate R such that*

$$R \leq \frac{1}{N\bar{t}} \left\lceil \log_2 \left\lceil \frac{\pi}{\epsilon_\theta} \right\rceil \left\lceil \frac{2(\epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t})\pi}{\sqrt{2}\epsilon_x} \right\rceil \left\lceil \frac{2|\bar{u}_x|\bar{t} + 2\delta_x \bar{t}}{\sqrt{2}\epsilon_x} \right\rceil \right\rceil . \quad (6.17)$$

The proof of this result is available in Appendix 6.B. Note that in the proof of Theorem 6.6 a covering procedure is presented for I_j. This covering procedure is very conservative. There exist coverings of rings with disks that require fewer disks but the objective of this theorem is simply to provide a theoretical upper bound on the required communication rate. In practice, the communication protocol requires a much lower rate than the theoretical bound, as will be shown through experiments in the next section.

6.5 Experiments

In order to test the observer, experiments were run in a lab on a Turtlebot2[1], which is a unicycle-type robot. The Turtlebot is equipped with a netbook, which utilizes ROS (Robot Operating System[2]). Four sets of experiments were run in total, each with the objective to test one particular configuration. All experiments were run with the following setting for the velocities $(\bar{u}_x, \bar{u}_\theta) = (0.1\ [\text{m/s}], -0.2\ [\text{rad/s}])$. The precision of the estimates is $(\epsilon_x, \epsilon_\theta) = (0.01\ [\text{m}], 0.01\ [\text{rad}])$. The four sets of experiments are:

1. *No perturbations:* The first set of experiments consists of steering the robot with the constant velocities without any additional perturbations;

2. *Small angular velocity perturbations:* The second set of experiments revolves around testing a configuration with perturbations only in the angular velocity;

[1] See https://www.turtlebot.com/turtlebot2/ for more information.

[2] ROS is an open-source operating system to control robots, more information available at https://www.ros.org/.

3. *Larger perturbations:* The third set of experiments involves large perturbations, both in the linear velocity and in the angular velocity;

4. *Larger perturbations, larger observation error:* The fourth and final set of experiments involves the same perturbations as the third set of experiments except N is now chosen larger.

A video of the experiments (as well as Gazebo simulations) is available at the following URL: https://www.youtube.com/watch?v=zx3Mckyj4EM.

6.5.1 First Experiment - No Perturbations

In this experiment, the dynamics of the robot are assumed to be unperturbed (i.e., $\delta_x = \delta_\theta = 0$). The sampling time for the robot is $\bar{t} = 0.5$s. We choose $N = 2$ (implying that a communication could potential occur every second). The following bounds are then obtained on η and R by applying Proposition 6.4 and Theorem 6.6:$\eta \leq 0.21$ [m], $R \leq 17$ [bits/s]. Several experiments are run, each consisting of 120 seconds. A typical trajectory in the x_1, x_2-plane is depicted on Fig. 6.2 while the observation error for the same run is depicted in Fig. 6.3.

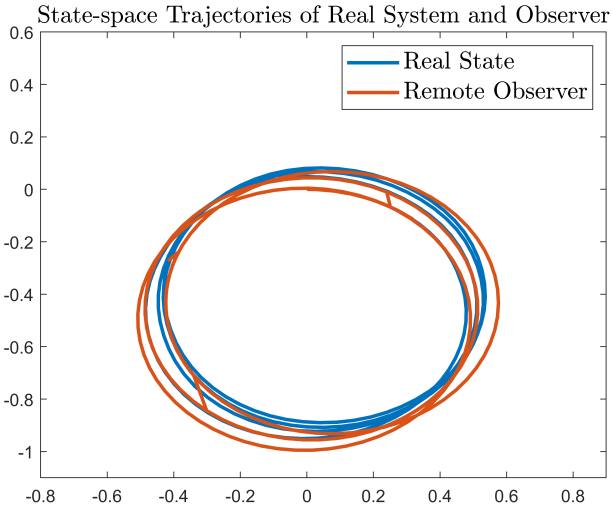

Figure 6.2: Figure depicting the state-space trajectory of the robot (blue) and the remote estimate (orange) for the unperturbed case.

As can be seen from the figures, even in the unperturbed case, several communications are required. This is due to internal frictions of the robot, imperfect actuation, frictions with the floor, ... all of which are not accounted

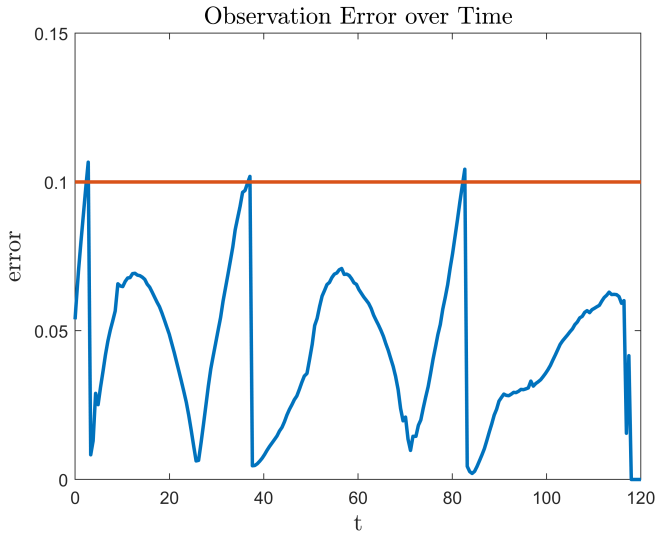

Figure 6.3: Figure depicting the observation error (blue), together with the triggering condition (orange) for the unperturbed case.

for in the model. On average, over 10 different trials, the number of communications is 4.1 per 120s. This is still much lower than the upper bound on the theoretical rate predicted, 120 communications. The resulting rate is $17 \times 4.1/120 = 0.5805$ [bits/s], which is 34 times lower than the upper bound on the theoretical rate, which proves the effectiveness of the event-triggered protocol, as well as some conservatism in the theoretical error bounds. In terms of observation error, the error remains much below the maximum observation bound.

6.5.2 Second Experiment - Small Angular Velocity Perturbations

In this experiment, the dynamics of the robot are assumed to only be perturbed in angular velocity. The sampling time for the robot is $\bar{t} = 0.5$s. We use $\delta_x = 0$ and $\delta_\theta = 0.05$ (implying up to 25% variation in the angular velocity). We again choose $N = 2$ (implying that a communication could potential occur every second). The following bounds are then obtained on η and R by applying Proposition 6.4 and Theorem 6.6: $\eta \leq 0.21$ [m], $R \leq 17$ [bits/s]. Note that since only δ_θ changed from the first experiments, the bounds are identical. Several experiments are run, each consisting of 120 seconds. A typical trajectory in the x_1, x_2-plane is depicted on Figure 6.4 while the observation error for the same run is depicted in Figure 6.5.

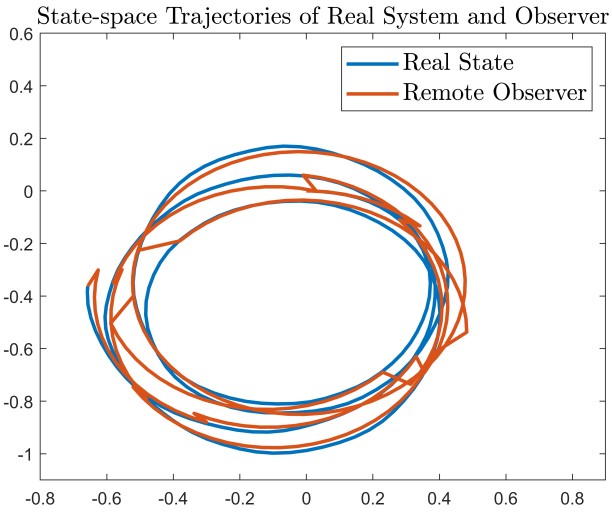

Figure 6.4: Figure depicting the state-space trajectory of the robot (blue) and the remote estimate (orange) for the case with angular velocity perturbations.

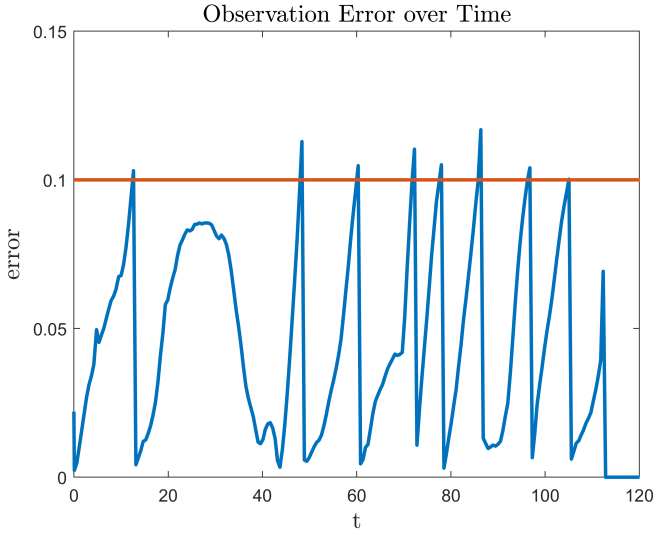

Figure 6.5: Figure depicting the observation error (blue), together with the triggering condition (orange) for the case with angular velocity perturbations.

This time, the state-space trajectories of the system deviate much more compared to the unperturbed case. The effects of the perturbation on the angular

velocity can clearly be seen. Over 10 experiments, an average of 10.5 communications are necessary, which is much lower than the upper bound on the theoretical rate of 120. The resulting rate is $17 \times 10.5/120 = 1.3125$ [bits/s]. This again validates the effectiveness of the event-triggered communication scheme.

6.5.3 Third Experiment - Large Perturbations

In this experiment, the dynamics of the robot are assumed to be perturbed by large perturbations. We use $\delta_x = 0.2$ and $\delta_\theta = 0.2$. This time, the sampling interval is assumed to be $\bar{t} = 0.1$. We choose $N = 3$ (implying that a communication could potential occur every 0.3 seconds). The following bounds are then obtained on η and R by applying Proposition 6.4 and Theorem 6.6: $\eta \leq 0.19$ [m], $R \leq 53.33$ [bits/s]. Several experiments are run, each consisting of 120 seconds. A typical trajectory in the x_1, x_2-plane is depicted on Figure 6.6.

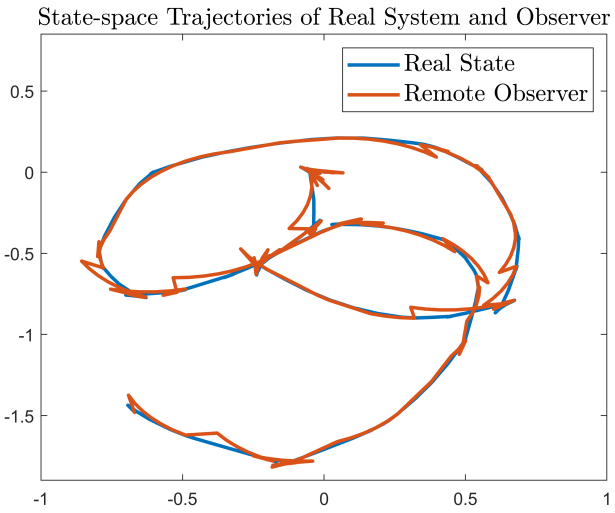

Figure 6.6: Figure depicting the state-space trajectory of the robot (blue) and the remote estimate (orange) for the large perturbations case.

The effects of these large perturbations are immediately seen on the state-space trajectories of the robot. Many more communications are therefore required: on average 53.6 over 120 seconds. This results in an effective communication rate of $16 \times 53.6/120 = 7.1467$ [bits/s]. This is again much below the upper bound on the theoretical rate.

6.5.4 Fourth Experiment - Large Perturbation Large Observation Error

For the final round of experiments, the same setting as the previous experiments is assumed, except $N = 4$ (implying that a communication could potential occur every 0.4 seconds). The following bounds are then obtained on η and R by applying Proposition 6.4 and Theorem 6.6: $\eta \leq 0.25$ [m], $R \leq 40$ [bits/s]. The different choice of N thus immediately impacts both the maximum error, which becomes larger and the upper bound on the theoretical rate, which becomes smaller. Several experiments are run, each consisting of 120 seconds. A typical trajectory in the x_1, x_2-plane is depicted on Figure 6.7.

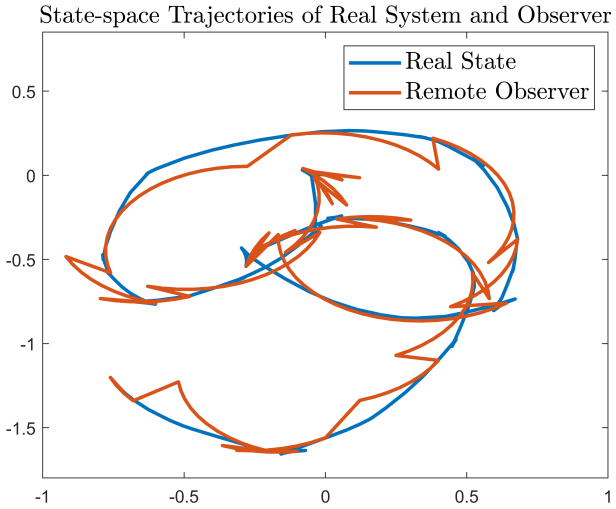

Figure 6.7: Figure depicting the state-space trajectory of the robot (blue) and the remote estimate (orange) for the large perturbations case with larger observation error.

The experiments confirm the same effect as the theoretical bounds: as can be seen in Figure 6.7, the maximum distance between the observed trajectory and the actual state-space trajectory is larger than in the previous experiment (see Figure 6.6). In terms of the number of communications, the average now sits at 24.2 communications per 120 seconds, which implies an effective rate of $16 \times 24.2/120 = 3.2267$ [bits/s], almost twice as low as in the previous experiment. This confirms the fact that there is a tradeoff between precision and rate that can be tuned through N.

6.6 Conclusion

In this document, we presented an event-triggered, data rate constrained observer for unicycle-type robots with constant velocities and time-varying perturbations. After posing the problem statement, the design of the agents that form the communication protocol was developed. Two theoretical results were presented. First, a proposition which links the minimum time interval between two consecutive communications and the maximum error. Secondly, a theorem that upper bounds the communication rate resulting from the communication protocol. The effectiveness of the proposed communication scheme was experimentally validated on Turtlebots.

We conclude with the following remarks on the communication protocol:

1. The communication protocol is very efficient at producing precise estimates at a remote distance;

2. The required communication rate is much lower than the upper bound on the theoretical rate, which is due both to the usage of an event-triggered communication protocol, as well as conservatism in the error bounds;

3. It is possible to exchange more precision for a higher communication rate and vice-versa, by tuning the parameter N;

4. Although a proportion of the transmitted bits is used to transmit $\hat{\theta}$, which is not included in the observation error, it is an essential part of the dynamics of the robot and hence plays a crucial role in increasing the time between two subsequent communications (by improving the quality of the remote estimate).

Further research on this topic includes configurations with piece-wise constant input, tracking of the full state instead of only the position, improving the covering procedure to improve the theoretical bounds, and studying the fundamental minimum requires capacity to observe unicycle-type robots.

Appendices

6.A Proof of Proposition 6.4

Proof: We start by defining the error $e(t) = [(x(t) - \hat{x}(t))^\intercal \ (\theta(t) - \hat{\theta}(t))]^\intercal$. The dynamics of the error are

$$\begin{bmatrix} \dot{e}_1(t) \\ \dot{e}_2(t) \\ \dot{e}_3(t) \end{bmatrix} = \begin{bmatrix} u_x(t) \left(\cos(\theta(t)) - \cos(\hat{\theta}(t))\right) + d_x(t)\cos(\theta(t)) \\ u_x(t) \left(\sin(\theta(t)) - \cos(\hat{\theta}(t))\right) + d_x(t)\sin(\theta(t)) \\ d_\theta(t) \end{bmatrix}, \tag{6.18}$$

which implies that

$$\begin{aligned} e_3(t+\hat{t}) &\leq e_3(t) + \delta_\theta \hat{t}, \\ e_3(t+\hat{t}) &\geq e_3(t) - \delta_\theta \hat{t}. \end{aligned} \tag{6.19}$$

Since $\cos(\theta), \sin(\theta) \in [-1, 1]$, $\forall \theta \in \mathbb{R}$, and due to (6.3) we clearly have that the solutions $e(t+\hat{t})$ of (6.18) satisfy

$$\begin{aligned} e_1(t+\hat{t}) &\leq e_1(t) + |\bar{u}_x|\hat{t} + \delta_x \hat{t}, \\ e_2(t+\hat{t}) &\leq e_2(t) + |\bar{u}_x|\hat{t} + \delta_x \hat{t}, \end{aligned} \tag{6.20}$$

and

$$\begin{aligned} e_1(t+\hat{t}) &\geq e_1(t) - |\bar{u}_x|\hat{t} - \delta_x \hat{t}, \\ e_2(t+\hat{t}) &\geq e_2(t) - |\bar{u}_x|\hat{t} - \delta_x \hat{t}. \end{aligned} \tag{6.21}$$

Since communications reset the observation error to ϵ_x, it remains to evaluate the observation error in between communication instants. We thus evaluate $\left\|x(t+\hat{t}) - \hat{x}(t+\hat{t})\right\|$ for $\hat{t} \geq 0$, assuming that a communication occurred at time t. There are two possible situations: either $N\bar{t} \geq \hat{t} \geq 0$, or $\hat{t} > N\bar{t}$.

Situation 1: $N\bar{t} \geq \hat{t} \geq 0$

This implies that the triggering condition (6.14) has not been checked and will only be checked at $\hat{t} = N\bar{t}$. For $\left\|x(t+\hat{t}) - \hat{x}(t+\hat{t})\right\|_2 = \left\|e_{1:2}(t+\hat{t})\right\|_2$, inequalities (6.20) and (6.21) imply that

$$\begin{aligned} \left\|e_{1:2}(t+\hat{t})\right\|_2 &\leq \left\|e_{1:2}(t)\right\|_2 + |\bar{u}_x|\hat{t} + \delta_x \hat{t} + |\bar{u}_x|\hat{t} + \delta_x \hat{t} \\ &\leq \epsilon_x + 2|\bar{u}_x|\hat{t} + 2\delta_x \hat{t}, \end{aligned}$$

which, from (6.11), implies that (6.16) holds for $\hat{t} : N\bar{t} \geq \hat{t} \geq 0$.

Situation 2: $\hat{t} > N\bar{t}$

This implies that (6.14) held at most $\bar{t}$ time instants ago (otherwise a communication would have occurred). We thus have

$$\left\|e_{1:2}(t+\hat{t}-\bar{t})\right\|_2 \leq \epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t} - 2|\bar{u}_x|\bar{t} - 2\delta_x \bar{t} \tag{6.22}$$

and hence, from (6.20) and (6.21), we have that

$$\left\|e_{1:2}(t+\hat{t})\right\|_2 \leq \epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t}. \tag{6.23}$$

□

6.B Proof of Theorem 6.6

Proof: The rate R depends on the number of bits b_j that are sent as each communication instant, which itself depends on the alphabet length. The alphabet

length is equal to the product of the number of elements in the covering of I_j, as well as the number elements in the covering of S^1. The latter is easily computed: with intervals of length $2\epsilon_\theta$, $\left\lceil \frac{\pi}{\epsilon_\theta} \right\rceil$ intervals are required.

The set I_j, on the other hand, is a ring of inner radius $\epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t} - 2|\bar{u}_x|\bar{t} - 2\delta_x\bar{t}$ and outer radius $\epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t}$. Recall that for any disk ϵ_x, there is a square of side $\sqrt{2}\epsilon$ inscribed. To cover a ring of outer radius r and inner radius $r - \epsilon_x$, it suffices to employ $\left\lceil \frac{2\pi r}{\sqrt{2}\epsilon_x} \right\rceil$ squares of side $\sqrt{2}\epsilon_x$ (see Fig. 6.8), and hence $\left\lceil \frac{2r\pi}{\sqrt{2}\epsilon_x} \right\rceil$ disks of radius ϵ_x. By splitting I_j in rings of thickness $\sqrt{2}\epsilon_x$, no more than $\left\lceil \frac{2(\epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t})\pi}{\sqrt{2}\epsilon_x} \right\rceil$ disks of radius ϵ_x are required to cover each ring. Since there are at most $\left\lceil \frac{2|\bar{u}_x|\bar{t} + 2\delta_x\bar{t}}{\sqrt{2}\epsilon_x} \right\rceil$ rings of thickness $\sqrt{2}\epsilon_x$ that form I_j, in total $\left\lceil \frac{2(\epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t})\pi}{\sqrt{2}\epsilon_x} \right\rceil \left\lceil \frac{2|\bar{u}_x|\bar{t} + 2\delta_x\bar{t}}{\sqrt{2}\epsilon_x} \right\rceil$ disks of radius ϵ_x are sufficient to cover I_j. We thus have that

$$l_j \leq \left\lceil \frac{\pi}{\epsilon_\theta} \right\rceil \left\lceil \frac{2(\epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t})\pi}{\sqrt{2}\epsilon_x} \right\rceil \left\lceil \frac{2|\bar{u}_x|\bar{t} + 2\delta_x\bar{t}}{\sqrt{2}\epsilon_x} \right\rceil.$$

This implies that

$$\begin{aligned} b_j &\coloneqq \lceil \log_2 l_j \rceil \\ &\leq \left\lceil \log_2 \left\lceil \frac{\pi}{\epsilon_\theta} \right\rceil \left\lceil \frac{2(\epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t})\pi}{\sqrt{2}\epsilon_x} \right\rceil \left\lceil \frac{2|\bar{u}_x|\bar{t} + 2\delta_x\bar{t}}{\sqrt{2}\epsilon_x} \right\rceil \right\rceil. \end{aligned}$$

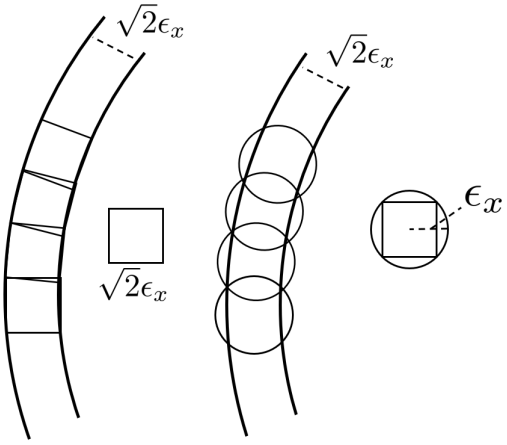

Figure 6.8: Figure depicting how balls of radius ϵ_x are used to cover a ring of thickness $\sqrt{2}\epsilon_x$ by the usage of inscribed squares of side $\sqrt{2}\epsilon_x$.

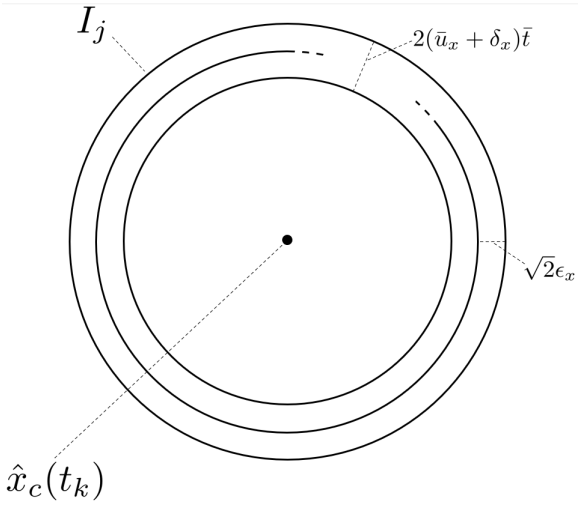

Figure 6.9: Figure depicting the region I_j and how it is split up into rings of thickness $\sqrt{2}\epsilon_x$.

The rate is defined in (6.10) as

$$R \coloneqq \lim_{j\to\infty} \frac{1}{s_j} \sum_{i=1}^{j} b_i.$$

It is dependent on the communication instants s_j, which depend on the particular realization of the system. At least N sampling instants elapse in between two subsequent communications. This implies that $s_j \geq jN\bar{t}$ and hence

$$R \leq \lim_{j\to\infty} \frac{1}{jN\bar{t}} \sum_{i=1}^{j} b_j,$$

which, since b_i is constant, implies that

$$R \leq \frac{b_j}{N\bar{t}},$$

and hence

$$R \leq \frac{\left\lceil \log_2 \left\lceil \frac{\pi}{\epsilon_\theta} \right\rceil \left\lceil \frac{2(\epsilon_x + 2|\bar{u}_x|N\bar{t} + 2\delta_x N\bar{t})\pi}{\sqrt{2}\epsilon_x} \right\rceil \left\lceil \frac{2|\bar{u}_x|\bar{t} + 2\delta_x \bar{t}}{\sqrt{2}\epsilon_x} \right\rceil \right\rceil}{N\bar{t}}.$$

□

Chapter 7

Conclusion and Recommendations

7.1 Concluding Remarks

As was discussed extensively in the introduction, the widespread of communication technologies means that many applications featuring dynamical systems now also include some form of wireless communication. These communication technologies have limitations that are inherent to their nature. These limitations are limited packet size, limited packet transmission rate, losses, noise/perturbations, and time-delays (see Section 1.1.2 for a more detailed explanation of these limitations). When these wireless communication technologies are used in conjunction with a controlled/observed dynamical system which has one of the following sources of uncertainty: parametric uncertainty, perturbations/noise, and sensitivity to initial conditions (see Section 1.1.3 for a more detailed explanation on each of these sources of uncertainty), it becomes necessary to design specific communication strategies to carry over information about the system via these communication channels whilst dealing with the limitations.

It is important to simultaneously handle the limitations due to communication technologies and solve the underlying control/observation problem, as opposed to dealing with each problem separately. Integrated approaches offer more possibilities than simply solving the communication problem separately from the control problem, as was highlighted in Section 1.1.1.

The objective of this thesis is to develop tools for interactions between dynamical systems and communication technologies. The following results are presented:

1. (**Chapter 2**) A data-rate constrained observation scheme for nonlinear systems (both continuous- and discrete-time) with sensitivity to initial conditions that is robust towards losses in the communication channel is developed.

2. (**Chapter 3**) For a network of agents described by identical discrete-time dynamical systems, several consensus protocols (consisting of several devices) are obtained (for different network topologies) which keep the agents in consensus whilst limiting the load on the communication.

3. (**Chapter 4**) An event-triggered communication protocol for the remote observation of continuous-time Lipschitz-nonlinear systems with bounded state perturbations and measurement noise is developed.

4. (**Chapter 5**) An event-triggered communication protocol for the remote observation of steered discrete-time Lipschitz-nonlinear systems with bounded state perturbations and measurement noise is obtained.

5. (**Chapter 6**) An event-triggered communication protocol for the remote observation of unicycle robots, with experimental validation is developed and validated.

Each of these results is written in the form of a paper which has either been published or submitted to a peer-reviewed journal/conference. Chapters 2 to 6 each correspond to one result and paper. What follows is a summary of each of these results.

7.1.1 Data-Rate Constrained Observers of Nonlinear Systems

The first result is a data-rate constrained observer for a nonlinear dynamical system with robustness towards losses. By means of a one-way communication channel, a system is connected to a remote location where online estimates of the state of the system should be reconstructed. The source of uncertainty is in the form of sensitivity to initial conditions. The communication channel can only transmit limited amounts of data per unit of time. A solution that is robust towards losses in the communication channel is provided, both for continuous-time and discrete-time systems. For this solution, bounds on the required communication rate are provided in terms of the upper box dimension of the state space of the dynamical system and an upper bound on the largest singular value of the system's Jacobian. Next, theorems that provide an analytical bound on the required minimum communication rate are presented. The bounds are obtained by using the Lyapunov dimension of the dynamical system instead of the upper box dimension in the communication rate.

The theoretical results are compared in simulations for two systems: the Lozi map and the Lorentz system. All simulations (for this result as for the next) are obtained by using Monte-Carlo methods. The simulations confirm that the proposed communication protocol is implementable on any channel with a capacity equal to or larger than the analytical upper bound provided by

the theorems. The novelty of this result is the robustness towards losses in the communication channel, which is a valuable property in light of the drawbacks of wireless communication technologies.

7.1.2 Data-Rate Constrained Consensus in Networks of Dynamical Systems

The second result is consensus for a network of agents, which communicate over data-rate constrained communication channels. A network of agents is considered whose dynamics are determined by a nonlinear discrete-time dynamical system. In this case, the source of uncertainty stems from a sensitivity to initial conditions. Each agent is equipped with a smart sensor (a device capable of measuring the state and performing some computations) and a controller. All agents are interconnected through channels that are limited in terms of communication capacity. The smart sensor and controller of each agent are placed at locations remote from one another such that the smart sensor and controller of each agent need to use the communication network as well to exchange information. The topology of the network of communication channels is represented through a communication adjacency matrix. By exchanging messages, the sensors and controllers should steer the agents so that they achieve a particular type of consensus.

Three different designs of smart sensors, controllers, and communication protocols that achieve this feature are presented, each with an increasing degree of interaction between the agents. For each protocol, a theorem is presented, providing conditions on the sufficient minimal data rates to implement them, as well as the requirements in terms of the communication adjacency matrix. It is shown that involving more agents in the decision on the common trajectory, which implies a higher degree of interaction, requires a higher number of communications to keep all systems in consensus. The protocols leading to consensus are tested via simulations on a network of Logistic maps and Hénon maps. For each of these systems, the theoretical bounds on the rate are compared to the rates observed in simulations, confirming the applicability and effectiveness of the consensus protocols. The novelty of the result is the data-rate constrained communication protocols for various network topologies.

7.1.3 An Event-Triggered Observation Scheme for Systems with Perturbations and Data-Rate Constraints

The third result is an event-triggered data-rate constrained observation scheme for a perturbed continuous-time Lipschitz-nonlinear dynamical system. Bounded perturbations form the source of uncertainty for this problem. The system is connected to a remote location by means of a communication channel that can

only send a limited amount of bits per unit of time. The goal is to provide estimates of the state of the system at the remote location whilst respecting a channel capacity constraint. A solution, in the form of an event-triggered communication protocol, is developed. The event-triggering mechanism compares the remote estimate with the current state and only sends a new estimate when the distance between both exceeds a certain threshold. This mechanism greatly reduces the average number of communications. For this communication protocol, a bound on the minimum channel capacity is provided, in terms of the underlying system's dynamics, and some tunable constants. The observation scheme's efficiency is then tested through simulations on unicycle-type robot systems. The simulations show that by using an event-triggered mechanism, it is possible to greatly reduce the average number of communications, and hence reduce the load on the communication channel. The novelty of this result is considering an event-triggered data-rate constrained observer for Lipschitz-nonlinear systems and providing analytical upper bounds on the necessary channel rate to implement the communication protocol.

7.1.4 Remote State Estimation of Steered Systems with Limited Communications: an Event-Triggered Approach

The tracking of the state of a perturbed system that is steered by a measured reference signal is the fourth result. A single discrete-time Lipschitz-nonlinear dynamical system is steered by an external signal which is measured and subjected to bounded state perturbations and measurement error. The steering signal is modeled as a parameter, which is a priori unknown but measured in real-time. The steering signal (parameter), perturbations, and measurement noise act as sources of uncertainty. This system is connected to a remote location by means of a communication channel. The objective is to provide estimates of the state at the remote location by sending messages via the communication channel and to limit the bandwidth usage of the communication. A solution in the form of several interacting agents is proposed. This solution makes use of an event-triggering mechanism to reduce bandwidth usage. The theoretical maximum communication rate resulting from the communication protocol is computed. This theoretical rate is then compared to the actual communication rate by means of simulations on several dynamical systems. The effectiveness of the event-triggered scheme is demonstrated by these simulations. The novelty of the result is the event-triggered protocol for discrete-time steered Lipschitz-nonlinear systems.

7.1.5 Observing a Unicycle Robot with Data-Rate Constraints: a Case Study

The final result of the thesis is a data-rate constrained observer specifically designed for unicycle-type robots, which is experimentally validated on a Turtlebot2 robot. A unicycle robot is steered by an external and known in advance steering signal. It is also subjected to random bounded perturbations. The robots send messages via WiFi to a remote location where estimates of the position of the robot should be maintained. An event-triggered communication protocol that achieved this whilst using little data is developed. For this communication protocol, an upper bound on the maximum number of communications per unit of time is obtained. This theoretical upper bound is compared to the actual rate resulting from the implementation of the protocol through various experiments. The experiments proved that the event-triggeredness of the protocol was very efficient at reducing the average number of communications. The novelty of the result is that the observer is specifically designed for unicycle-type robots. The importance of this result is that it confirms, through experiments, the effectiveness and applicability of the third and fourth results to real-life problems.

7.2 Future Work and Recommendations

7.2.1 Future Work

Extensions

For each of the aforementioned results, further works could be carried out, improving on the existing results.

- (**Chapter 2**): Since the communication protocol is robust towards losses, an appropriate continuation would be to determine a mathematical model for the losses, as well as results providing bounds on the error, when losses occur. The best approach to achieve this is by including a stochastic loss model.
- (**Chapter 3**): Each of the designed consensus protocols requires a specific communication network topology. One interesting extension would consist in adapting the consensus protocol so that it can accept any network topology, e.g., by having nodes carry over messages to third nodes (that is, to have chains of communication in the network). Another extension would be to consider configurations with systems whose dynamics are not necessarily affine with respect to the control input.
- (**Chapters 4 and 5**): Although the simulations prove that the event-triggered communication protocol is efficient at reducing the required communication rate, they also prove that the error bounds are conservative.

The reasons for this conservatism lie partially in the Lyapunov-liken approach that is used in the LMI formulation. One possible extension would be to improve on the error bounds.

- (**Chapter 6**): The protocol that was designed implied that a constant control input was being applied to the unicycle robots. The communication protocol could easily be extended to the case of piece-wise constant control inputs by adding additional messages containing information necessary to reconstruct this piece-wise constant control input.

Combining Several Sources of Uncertainty

Many solutions that have been developed for problems involving dynamical systems and communication technologies focus on one of the three sources of uncertainty (sensitivity to initial conditions, perturbations/noise, and parametric uncertainty), and provide solutions specifically for that source of uncertainty. What functions for one source of uncertainty generally doesn't apply to others. An example of this is the notions of entropy that have been used for systems with sensitivity to initial conditions but that becomes difficult to use in the case of perturbations. The reasons for this being that most notions of entropy are defined as asymptotic quantities which, in the case of perturbed systems are simply infinite. In this thesis, an effort to combine several sources of uncertainty was made but no general result for problems combining sensitivity to initial conditions, perturbation/noise, and parametric uncertainty have been obtained.

Each of these sources models a property of real-life systems and it is generally not possible to model e.g., perturbations as parametric uncertainty and vice-versa. Designing control tools that combine all three sources of uncertainty is thus an important direction for research and should be pursued in the future.

Dealing With More Drawbacks

Packet-based communication technologies suffer from five different types of drawbacks: limited packet size, limited transmission rate, corrupted packets, packet losses, and delays. In this thesis, three out of the five drawbacks have been addressed directly but solutions were provided to deal with packet corruption and time delays. Up to now, there have been few works in the literature dealing with all drawbacks together. Providing solutions that are robust towards losses and time-delays in the communication channel whilst simultaneously taking into account the limited packet size and transmission rate would be an important achievement, considering that most modern wireless technologies suffer from all four drawbacks (Wi-Fi, 4G, 5G).

7.2.2 Final Recommendations

In this thesis, the focus has been on providing general solutions for problems involving several sources of uncertainty and dealing with several drawbacks of communication technologies. The main reason for this approach is that the aforementioned limitations/problems all correspond to difficulties encountered when dealing with real-life applications. The focus in research on interactions between dynamical systems and communication technologies, and more generally engineering, should remain on solving problems stemming from real-life applications and designing solutions that are applicable in practice.

Bibliography

K.-E. Åarzén. A simple event-based PID controller. *IFAC Proceedings Volumes*, 32(2):8687–8692, 1999.

M. Abdelrahim, V. S. Dolk, S. Member, and W. P. M. H. Heemels. Event-Triggered Quantized Control for Input-to-State Stabilization of Linear Systems With Distributed Output Sensors. *IEEE Transactions on Automatic Control*, 64(12):4952–4967, 2019.

R. L. Adler, A. G. Konheim, and M. H. McAndrew. Topological entropy. *Transactions of the American Mathematical Society*, 114(2):309–319, 1965.

A. D. Ames, A. Abate, and S. Sastry. Sufficient Conditions for the Existence of Zeno Behavior. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 686–701, Seville, 2005.

B. R. Andrievsky, A. S. Matveev, and A. L. Fradkov. Control and Estimation under Information Constraints: Toward a Unified Theory of Control, Computation and Communications. *Automation and Remote Control*, 71(4):572–633, 2010.

T. Arai and J. Ota. Dwarf intelligence - A large object carried by seven dwarves. *Robotics and Autonomous Systems*, 18:149–155, 1996.

B. Asadi Khashooei, D. J. Antunes, and W. P. Heemels. A Consistent Threshold-Based Policy for Event-Triggered Control. *IEEE Control Systems Letters*, 2 (3):447–452, 2018.

K. J. Åström and B. Bernhardsson. Comparison of periodic and event based sampling for first-order stochastic systems. In *Proceedings of the 14th IFAC World Congress*, pages 5006–5011, Beijing, 1999.

K. M. Awan, P. A. Shah, K. Iqbal, S. Gillani, W. Ahmad, and Y. Nam. Underwater Wireless Sensor Networks: A Review of Recent Issues and Challenges. *Wireless Communications and Mobile Computing*, 2019, 2019.

M. A. Aziz-Alaoui, C. Robert, and C. Grebogi. Dynamics of a Henon-Lozi-type map. *Chaos Solitons and Fractals*, 12:2323–2341, 2001.

J. Baillieul. Data-rate requirements for nonlinear feedback control. *Proc. 6th IFAC Symp. Nonlinear Control Syst.*, pages 1277–1282, 2004.

J. Baillieul and P. J. Antsaklis. Control and Communication Challenges in Networked Real-Time Systems. *Proceedings of the IEEE*, 95(1):9–28, 2007.

A. Bemporad, W. P. Heemels, and M. Johansson. *Networked Control Systems*, volume 406. Heidelberg, Springer-Verlag Berlin, 2010.

V. A. Boichenko, G. A. Leonov, and V. Reitman. *Dimension Theory for Ordinary Differential Equations*. Springer Vieweg Teubner Verlag, 2005.

D. P. Borgers and W. P. H. Heemels. Event-separation properties of event-triggered control systems. *IEEE Transactions on Automatic Control*, 59(10): 2644–2656, 2014.

R. W. Brockett and D. Liberzon. Quantized Feedback Stabilization of Linear Systems. *IEEE Transactions on Automatic Control*, 45(7):1279–1289, 2000.

R. A. Brooks, P. Maes, M. J. Mataric, and G. More. Lunar Base Construction Robots. In *Proceedings of the IEEE International Workshop on Intelligent Robots and Systems*, pages 389–392, Ibaraki, 1990.

F. Colonius. Minimal Bit Rates and Entropy for Exponential Stabilization. *SIAM Journal on Control and Optimization*, 50(5):2988–3010, 2012.

F. Colonius, C. Kawan, and G. Nair. A note on topological feedback entropy and invariance entropy. *Systems and Control Letters*, 62(5):377–381, 2013.

D. W. Davies. An historical study of the beginnings of packet switching. *Computer Journal*, 44(3):152–162, 2001.

C. De Persis. A Note on Stabilization via Communication Channel in the presence of Input Constraints. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 187–192, Maui, 2003.

D. F. Delchamps. Stabilizing a Linear System with Quantized State Feedback. *IEEE Transactions on Automatic Control*, 35(8):916–924, 1990.

V. S. Dolk, D. P. Borgers, and W. P. Heemels. Output-based and decentralized dynamic event-Triggered control with guaranteed Lp-Gain performance and zeno-freeness. *IEEE Transactions on Automatic Control*, 62(1):34–49, 2017.

W. Dong. Consensus of High-Order Nonlinear Continuous-Time Systems With Uncertainty and Limited Communication Data Rate. *IEEE Transactions on Automatic Control*, 64(5):2100–2107, 2019.

M. C. Donkers and W. P. Heemels. Output-based event-triggered control with guaranteed L infinity- gain and improved and decentralized event-triggering. *IEEE Transactions on Automatic Control*, 57(6):1362–1376, 2012.

A. Douady and J. Oesterle. Dimension de Hausdorff des attracteurs. *Comptes Rendus Acad. Sci. Ser. A*, 290:1135–1138, 1980.

T. Downarowicz. *Entropy in Dynamical Systems*. Cambridge University Press, 2011.

Z. Elhadj. *Lozi Mappings: Theory and Applications*. CRC Press, 2013.

N. Elia and S. K. Mitter. Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 46(9):1384–1400, 2001.

J. Ellenberg. *How Not to Be Wrong, The Power of Mathematical Thinking*. The Penguin Press, New York, 2014.

T. Eren, P. N. Belhumeur, and A. S. Morse. Closing ranks in vehicle formations based on rigidity. In *Proceedings of the 41st IEEE Conference on Decision and Control*, volume 3, pages 2959–2964, Las Vegas, 2002.

D. Eustace, D. P. Barnes, and J. O. Gray. Co-operant mobile robots for industrial applications. *Proceedings of the Industrial Electronics Conference*, 1:39–44, 1993.

W. J. Evers and H. Nijmeijer. Practical stabilization of a mobile robot using saturated control. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 2394–2399, San Diego, 2006.

K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley & Sons: Hoboken, 1997.

S. Fang, J. Chen, and H. Ishii. *Towards Integrating Control and Information Theories*, volume 465. Springer International Publishing Switzerland, 2017. URL http://link.springer.com/10.1007/978-3-319-49289-6.

J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, 49(9):1465–1476, 2004.

C. Fiter, L. Hetel, W. Perruquetti, and J.-P. Richard. A State Dependent Sampling for Linear State Feedback. *Automatica*, 48(8):1860–1867, 2012.

C. Fiter, L. Hetel, W. Perruquetti, and J.-P. Richard. A robust stability framework for LTI systems with time-varying sampling. *Automatica*, 54:56–64, 2015.

A. L. Fradkov, B. Andrievsky, and R. J. Evans. Chaotic observer-based synchronization under information constraints. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 73(6):066209, 2006.

A. L. Fradkov, B. Andrievsky, and R. J. Evans. Synchronization of nonlinear systems under information constraints. *Chaos*, 18(3):037109–8, 2008a.

A. L. Fradkov, B. Andrievsky, and R. J. Evans. Controlled synchronization under information constraints. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 78(3):036210, 2008b.

E. Fridman, A. Seuret, and J. P. Richard. Robust sampled-data stabilization of linear systems: An input delay approach. *Automatica*, 40(8):1441–1446, 2004.

M. Fu and L. Xie. The Sector Bound Approach to Quantized. *IEEE Transactions on Automatic Control*, 50(11):1698–1711, 2005.

E. Garcia, P. J. Antsaklis, and L. A. Montestruque. *Model-Based Control of Networked Systems*. Springer International Publishing Switzerland, 2010.

B. Grandvallet, A. Zemouche, M. Souley-Ali, and M. Boutayeb. New LMI Condition for Observer-Based H-infinity Stabilization of a Class of Nonlinear Discrete-Time Systems. *Siam Journal on Control and Optimization*, 51 (1):784–800, 2013.

F. Granelli, D. Kliazovich, and N. da Fonseca. Performance limitations of IEEE 802.11 networks and potential enhancements. *International Journal of Computer Research*, 14(1/2):85–108, 2007.

M. Guerra, D. Efimov, G. Zheng, and W. Perruquetti. Finite-time supervisory stabilization for a class of nonholonomic mobile robots under input disturbances. In *Proceedings of the 19th IFAC World Congress*, pages 4867–4872, Cape Town, 2014.

M. Guerra, D. Efimov, G. Zheng, and W. Perruquetti. Finite-time obstacle avoidance for unicycle-like robot subject to additive input disturbances. *Autonomous Robots*, 41(1):19–30, 2017.

A. Guillemin. *El mundo fisico: gravedad, gravitacion, luz, calor, electricidad, magnetismo, etc. Ilustracion compuesta de numerosas vinetas intercaladas en el texto*. Montaner Y Simon, Editores, 1885.

A. Hamilton, S. Holdcroft, D. Fenucci, P. Mitchell, N. Morozs, A. Munafò, and J. Sitbon. Adaptable underwater networks: The relation between autonomy and communications. *Remote Sensing*, 12(20):1–22, 2020.

D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi. Stochastic event-triggered sensor schedule for remote state estimation. *IEEE Transactions on Automatic Control*, 60(10):2661–2675, 2015.

W. P. Heemels, R. J. Gorter, A. Van Zijl, P. P. Van Den Bosch, S. Weiland, W. H. Hendrix, and M. R. Vonder. Asynchronous measurement and control: A case study on motor synchronization. *Control Engineering Practice*, 7(12): 1467–1482, 1999.

W. P. Heemels, A. R. Teel, N. V. D. Wouw, and D. Nesic. Constraints : Tradeoffs Between Transmission Intervals , Delays and Performance. *IEEE Transactions on Automatic Control*, 55(8):1781–1796, 2010.

W. P. Heemels, K. H. Johansson, and P. Tabuada. An Introduction to Event-triggered and Self-triggered Control. In *Proceedings of the 51st IEEE Conference on Decision and Control*, pages 3270–3285, Maui, 2012.

T. Henningsson, E. Johannesson, and A. Cervin. Sporadic event-based control of first-order linear stochastic systems. *Automatica*, 44(11):2890–2895, 2008.

M. Henon. A Two-dimensional Mapping with a Strange Attractor. *Communications in Mathematical Physics*, 50:69–77, 1976.

J. Hespanha, P. Naghshtabrizi, and Y. Xu. A Survey of Recent Results in Networked Control Systems. *Proceedings of the IEEE*, 95(1):138–162, 2007.

L. Hetel, C. Fiter, H. Omran, A. Seuret, J.-p. Richard, and S. I. Niculescu. Recent developments on the stability of systems with aperiodic sampling : An overview. *Automatica*, 76:309–335, 2017.

G. W. Hill. On the part of the motion of the lunar perigee which is a function of the mean motions of the sun and moon. *Acta mathematica*, 8:1–36, 1886.

R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 2013.

J. Huang, D. Shi, and T. Chen. Event-triggered state estimation with an energy harvesting sensor. *IEEE Transactions on Automatic Control*, 62(9):4768–4775, 2017.

B. R. Hunt. Maximum local Lyapunov dimension bounds the box dimension of chaotic attractors. *Nonlinearity*, 9:845–852, 1996.

D. Hutt, K. Snell, and P. Belanger. Alexander Graham Bell's Photophone. *Optics and Photonics*, 4(6):20–25, 1993.

R. Isermann. *Digital Control Systems*. Springer-Verlag Berlin Heidelberg, 1981.

S. Ito. An Estimate from above for the Entropy and the Topological Entropy of a C1-diffeomorphism. *Proceedings of the Japanese Academy*, 46(1993):226–230, 1970.

Z. P. Jiang and H. Nijmeijer. Tracking Control of Mobile Robots: A Case Study in Backstepping. *Automatica*, 33(7):1393–1399, 1997.

P. J. Johnson and J. S. Bay. Distributed control of simulated autonomous mobile robot collectives in payload transportation. *Autonomous Robots*, 2(1):43–63, 1995.

J. L. Kaplan and J. A. Yorke. *Functional Differential Equations and Approximation of Fixed Points, Chaotic behavior of multidimensional difference equations.* Springer, Berlin, Heidelberg, 1978.

C. Kawan. Upper and lower estimates for invariance entropy. *Discrete & Continuous Dynamical Systems*, 290(1):169–186, 2011.

C. Kawan. *Invariance Entropy for Deterministic Control Systems An Introduction.* Springer Cham Heidelberg New York Dordrecht London, 2013.

C. Kawan. Exponential state estimation, entropy and Lyapunov exponents. *Systems and Control Letters*, 113:78–85, 2018.

C. Kawan and S. Yüksel. On optimal coding of non-linear dynamical systems. *IEEE Transactions on Information Theory*, 64(10):6816–6829, 2018.

C. Kawan, A. S. Matveev, and A. Y. Pogromsky. Data rate limits for the remote state estimation problem. In *Proceedings of the 21st IFAC World Congress*, Berlin, 2020.

C. Kawan, A. S. Matveev, and A. Y. Pogromsky. Remote state estimation problem: Towards the data-rate limit along the avenue of the second lyapunov method. *Automatica*, 125:109467, 2021.

K. Khalil. *Nonlinear Systems Third Edition.* Prentice Hall, 2002.

D. Kostic, S. Adinandra, J. Caarls, N. Van De Wouw, and H. Nijmeijer. Collision-free tracking control of unicycle mobile robots. In *Proceedings of the 48th IEEE Conference on Decision and Control held jointly with the 2009 28th Chinese Control Conference*, pages 5667–5672, Shanghai, 2009.

N. V. Kuznetsov. The Lyapunov dimension and its estimation via the Leonov method. *Physics Letters, Section A: General, Atomic and Solid State Physics*, 380(25-26):2142–2149, 2016.

F. Lamnabhi-Lagarrigue, A. Annaswamy, S. Engell, A. Isaksson, P. Khargonekar, R. M. Murray, H. Nijmeijer, T. Samad, D. Tilbury, and P. Van den Hof. Systems and Control for the future of humanity, research agenda: Current and future roles, impact and grand challenges. *Annual Reviews in Control*, 43:1–64, 2017.

B. Lathi and Z. Ding. *Modern Digital and Analog Communication Systems*. Oxford University Press, 2018.

F. Ledrappier and L.-S. Young. The Metric Entropy of Diffeomorphisms: Part II: Relations between Entropy, Exponents and Dimension. *The Annals of Mathematics*, 122(3):540–574, 1985.

G. A. Leonov. *Strange Attractors and Classical Stability Theory*. St. Petersburg University Press, 2007.

G. A. Leonov, N. V. Kuznetsov, N. A. Korzhemanova, and D. V. Kusakin. Lyapunov dimension formula for the global attractor of the Lorenz system. *Communications in Nonlinear Science and Numerical Simulation*, 41:84–103, 2016.

F. Lewis. *Applied Optimal Control & Estimation: Digital Design & Implementation*. Prentice Hall, 1992.

H. Li, G. Chen, T. Huang, Z. Dong, W. Zhu, and L. Gao. Event-Triggered Distributed Average Consensus over Directed Digital Networks with Limited Communication Bandwidth. *IEEE Transactions on Cybernetics*, 46(12):3098–3110, 2016.

L. Li, X. Wang, and M. Lemmon. Stabilizing bit-rate of perturbed event triggered control systems. In *Proceedings of the 4th IFAC Conf. Anal. Design Hybrid Systems*, pages 70–75, 2012.

L. Li, X. Wang, and M. D. Lemmon. Efficiently Attentive Event-Triggered Systems with Limited Bandwidth. *IEEE Transactions on Automatic Control*, 62(3):1491–1497, 2017.

T. Li and L. Xie. Distributed consensus over digital networks with limited bandwidth and time-varying topologies. *Automatica*, 47(9):2006–2015, 2011.

T. Li, M. Fu, L. Xie, and J.-f. Zhang. Distributed Consensus With Limited Communication Data Rate. *IEEE Transactions on Automatic Control*, 56(2): 279–292, 2011.

D. Liberzon. Hybrid feedback stabilization of systems with quantized signals. *Automatica*, 39(9):1543–1554, 2003a.

D. Liberzon. On Stabilization of Linear Systems with Limited Information. *IEEE Transactions on Automatic Control*, 48(2):304–307, 2003b.

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

D. Liberzon and S. Mitra. Entropy and minimal data rates for state estimation and model detection. In *Proceedings of the 16' International Conference on Hybrid Systems: Computation and Control*, pages 247–256, Vienna, 2016.

D. Liberzon and S. Mitra. Entropy and Minimal Bit Rates for State Estimation and Model Detection. *IEEE Transactions on Automatic Control*, 63(10):3330–3340, 2018.

D. Liberzon and D. Nešić. Input-to-state stabilization of linear systems with quantized state measurements. *IEEE Transactions on Automatic Control*, 52 (5):767–781, 2007.

Q. Ling and M. D. Lemmon. Stability of quantized control systems under dynamic bit assignment. *IEEE Transactions on Automatic Control*, 50(5):734–740, 2005.

S. Liu, T. Li, and L. Xie. Distributed Consensus for Multiagent Systems with Communication Delays and Limited Data Rate. *Siam Journal on Control and Optimization*, 49(6):2239–2262, 2011.

T. Liu and Z.-p. Jiang. Quantized Event-based Control of Nonlinear Systems. In *Proceedings of the 54th IEEE Conference on Decision and Control*, pages 4806–4811, Osaka, 2015. IEEE.

T. Liu and Z. P. Jiang. Event-Triggered Control of Nonlinear Systems with State Quantization. *IEEE Transactions on Automatic Control*, 64(2):797–803, 2019.

J. Lofberg. YALMIP : A toolbox for modeling and optimization in MATLAB. In *IEEE Internation Symposium on Computer Aided Control Systems Design*, pages 284–289, Taipei, 2004.

E. N. Lorenz. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

R. Lozi. Un Attracteur Étrange du Type de Hénon. *Journal de Physique Colloques*, 39(C5):9–10, 1978.

N. C. Martins, M. A. Dahleh, and N. Elia. Feedback Stabilization of Uncertain Systems in the Presence of a Direct link. *IEEE Transactions on Automatic Control*, 51(3):438–447, 2006.

A. S. Matveev. State estimation via limited capacity noisy communication channels. *Mathematics of Control, Signals, and Systems*, 20(2):1–357, 2008.

A. S. Matveev and A. Y. Pogromsky. Observation of nonlinear systems via finite capacity channels: Constructive data rate limits. *Automatica*, 70:217–229, 2016.

A. S. Matveev and A. Y. Pogromsky. Two Lyapunov methods in nonlinear state estimation via finite capacity communication channels. *IFAC-PapersOnLine*, 50(1):4132–4137, 2017.

A. S. Matveev and A. Y. Pogromsky. Observation of nonlinear systems via finite capacity channels, Part II: Restoration entropy and its estimates. *Automatica*, 103:189–199, 2019.

A. S. Matveev and A. V. Savkin. An Analogue of Shannon Information Theory for Detection and Stabilization via Noisy Discrete Communication Channels. *Siam Journal on Control and Optimization*, 46(4):1323–1367, 2007.

A. S. Matveev and A. V. Savkin. *Estimation and Control over Communication Networks*. Birkhäuser Boston Basel Berlin, 2009.

R. M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261(10), 1976.

L. Meier, J. Peschon, and R. M. Dressler. Optimal Control of Measurement Subsystems. *IEEE Transactions on Automatic Control*, AC-12(5):528–536, 1967.

Y. Meng and T. Li. Quantized observer-based coordination of linear multi-agent systems. In *Proceedings of the 19th IFAC World Congress*, pages 4693–4698. IFAC, 2014. URL http://dx.doi.org/10.3182/20140824-6-ZA-1003.00033.

Y. Meng, T. Li, and J. F. Zhang. Coordination over Multi-Agent Networks with Unmeasurable States and Finite-Level Quantization. *IEEE Transactions on Automatic Control*, 62(9):4647–4653, 2017.

M. Misiurewicz. Strange attractors for the Lozi mappings. *Annals of the New York Academy of Sciences*, pages 348–358, 1980.

M. Miskowicz. *Event-Based Control and Signal Processing*. CRC Press, 2016.

L. A. Montestruque and P. J. Antsaklis. Static and dynamic quantization in model-based networked control systems. *International Journal of Control*, 80 (1):87–101, 2007.

L. G. Moreira, S. Tarbouriech, A. Seuret, and J. M. Gomes da Silva. Observer-based event-triggered control in the presence of cone-bounded nonlinear inputs. *Nonlinear Analysis: Hybrid Systems*, 33:17–32, 2019.

M. Morse and G. A. . Hedlund. Symbolic Dynamics. *American Journal of Mathematics*, 60(4):815–866, 1938.

MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019. URL http://docs.mosek.com/9.0/toolbox/index.html.

M. Muehlebach and S. Trimpe. Distributed event-based state estimation for networked systems: An LMI approach. *IEEE Transactions on Automatic Control*, 63(1):269–276, 2018.

J. D. Murray. *Discrete Population Models for a Single Species*. Springer New York, 2002.

R. M. Murray, K. J. Åström, S. P. Boyd, R. W. Brockett, and G. Stein. Future Directions in Control in an Information-Rich World. *IEEE Control Systems Magazine*, 23(April):20–33, 2003.

G. N. Nair and R. J. Evans. Exponential stabilisability of multidimensional linear systems with finite data rates. *Automatica*, 39:585–593, 2003.

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

G. N. Nair, F. Fagnani, S. Zampieri, and R. Evans. Feedback Control Under Data Rate Constraints: An Overview. *Proceedings of the IEEE*, 95(1):108–137, 2007.

D. Nesic and D. Liberzon. A Unified Framework for Design and Analysis of Networked and Quantized Control Systems. *IEEE Transactions on Automatic Control*, 54(4):732–747, 2009.

H. Nijmeijer and A. van der Schaft. *Nonlinear Dynamical Control Systems*. Springer Science+Business Media, 2013.

R. Olfati-Saber and R. M. Murray. Graph rigidity and distributed formation stabilization of multi-vehicle systems. In *Proceedings of the 41st IEEE Conference on Decision and Control*, volume 3, pages 2965–2971, Las Vegas, 2002.

R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, 2004.

H. Omran, L. Hetel, J.-P. Richard, and F. Lamnabhi-Lagarrigue. Stability analysis of bilinear systems under aperiodic sampled-data control. *Automatica*, 50 (4):1288–1295, 2014.

H. Omran, L. Hetel, M. Petreczky, J.-P. Richard, and F. Lamnabhi-Lagarrigue. Stability analysis of some classes of input-affine nonlinear systems with aperiodic sampled-data control. *Automatica*, 70:266–274, 2016.

A. Y. Pogromsky and A. S. Matveev. Estimation of the topological entropy via the direct Lyapunov method. *Nonlinearity*, 24:1937–1959, 2011.

A. Y. Pogromsky and A. S. Matveev. Data rate limitations for observability of nonlinear systems. *IFAC-PapersOnLine*, 49(14):119–124, 2016a.

A. Y. Pogromsky and A. S. Matveev. Stability Analysis via Averaging Functions. *IEEE Transactions on Automatic Control*, 61(4):1081–1086, 2016b.

R. Postoyan and D. Nesic. A framework for the observer design for networked control systems. *IEEE Transactions on Automatic Control*, 57(5):1309–1314, 2012.

S. Raghavan and J. K. Hedrick. Hyperbolic observer design for a class of nonlinear systems. *International Journal of Control*, 59(2):515–528, 1994.

R. Rajamani. Observers for Lipschitz Nonlinear Systems. *IEEE Transactions on Automatic Control*, 43(3):397–401, 1998.

W. Ren and R. W. Beard. *Distributed Consensus in Multi-vehicle Cooperative Control, Theory and Applications*. Springer-Verlag London Limited, 2008.

C. W. Reynolds. Flocks, Herds, and Schools: A Distributed Behavioral Model. *Computer Graphics*, 21(4):25–34, 1987.

J.-P. Richard and T. Divoux. *Systèmes commandés en réseau*. Hermes Science Publications, 2007.

E. Rosenberg. *Fractal Dimensions of Networks*. Springer Nature Switzerland AG, 2020.

W. Rudin. *Principles of Mathematical Analysis*. MaGraw Hill: New York, 1976.

A. Sahai and S. Mitter. The Necessity and Sufficiency of Anytime Capacity for Stabilization of a Linear System Over a Noisy Communication Link — Part I : Scalar Systems. *IEEE Transactions on Information Theory*, 52(8):3369–3395, 2006.

A. V. Savkin. Analysis and synthesis of networked control systems: Topological entropy, observability, robustness and optimal control. *Automatica*, 42:51–62, 2006.

R. E. Seifullaev and A. L. Fradkov. Event-Triggered Control of Sampled-Data Nonlinear Systems. *IFAC-PapersOnLine*, 49(14):12–17, 2016.

C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

D. Shi, T. Chen, and M. Darouach. Event-based state estimation of linear dynamic systems with unknown exogenous inputs. *Automatica*, 69:275–288, 2016.

H. Sibai and S. Mitra. Optimal Data Rate for State Estimation of Switched Nonlinear Systems. In *Proceedings of the 17' International Conference on Hybrid Systems: Computation and Control*, pages 71–80, Pittsburgh, 2017.

H. Sibai and S. Mitra. State Estimation of Dynamical Systems with Unknown Inputs: Entropy and Bit Rates. In *Proceedings of the 18' International Conference on Hybrid Systems: Computation and Control*, pages 217–226, Porto, 2018.

S. Siegmund and P. Taraba. Approximation of box dimension of attractors using the subdivision algorithm. *Dynamical Systems*, 21(1):1–24, 2006.

T. Simsek, R. Jain, and P. Varaiya. Scalar Estimation and Control With Noisy Binary Observations. *IEEE Transactions on Automatic Control*, 49(9):1598–1603, 2004.

B. Sinopoli, C. Sharp, L. Schenato, S. Schaffert, and S. S. Sastry. Distributed control applications within sensor networks. *Proceedings of the IEEE*, 91(8): 1235–1245, 2003.

C. Sparrow. *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors.* Springer-Verlag: New York, 1982.

T. Stojanovski, L. Kocarev, and R. Harris. Applications of Symbolic Dynamics in Chaos Synchronization. *IEEE Transactions on Circuits and Systems*, 44 (10):1014–1018, 1997.

S. H. Strogatz. *Nonlinear Dynamics and Chaos.* Perseus Books Publishing, 1994.

P. Tabuada. Event-Triggered Real-Time Scheduling of Stabilizing Control Tasts. *IEEE Transactions on Automatic Control*, 52(9):1680–1685, 2007.

F. Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick*, pages 366–381. Springer, Berlin, Heidelberg, 1980.

P. Tallapragada and J. Cortes. Event-Triggered Stabilization of Linear Systems Under Bounded Bit Rates. *IEEE Transactions on Automatic Control*, 61(6): 1575–1589, 2016.

A. Tanwani, C. Prieur, and M. Fiacchini. Observer-based feedback stabilization of linear systems with event-triggered sampling and dynamic quantization. *Systems and Control Letters*, 94:46–56, 2016.

S. Tarbouriech, A. Seuret, L. G. Moreira, and J. M. da Silva. Observer-based event-triggered control for linear systems subject to cone-bounded nonlinearities. *IFAC-PapersOnLine*, 50(1):7893–7898, 2017.

G. Teschl. Ordinary differential equations and Dynamical Systems. *Lecture Notes from http://www. mat. univie. ac. at/gerald*, 2004.

J. Toner and Y. Tu. Flocks, herds, and schools: A quantitative theory of flocking. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 58(4):4828–4858, 1998.

S. Trimpe. Event-based state estimation: An emulationbased approach. *IET Control Theory and Applications*, 11(11):1684–1693, 2017.

N. Tufillaro, T. Abbott, and J. P. Reilly. *An Experimental Approach to Nonlinear Dynamics and Chaos*. Addison-Wesley, 1992.

A. J. van der Schaft and H. Schumacher. *An Introduction to Hybrid Dynamical Systems*. Springer-Verlag London, 2000.

T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel Type of Phase Transition in a System of Self-Driven Particles. *Physical Review Letters*, 75(6):1226–1229, 1995.

R. Vidal, O. Shakemia, and S. Sastry. Formation Control of Nonholonomic Mobile Robots with Omnidirectional Visual Servoing and Motion Segmentation. In *Proceedings of the 2003 IEEE International Conference on Robotics and Automation*, pages 584–589, Taipei, 2003.

Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. Continuous Time Observers of Nonlinear Systems with Data-Rate Constraints. In *Proceedings of the 5th IFAC Conference on Analysis and Control of Chaotic Systems*, Eindhoven, 2018a.

Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. A Data Rate Constrained Observer for Discrete Nonlinear Systems. In *Proceedings of the 57th IEEE Conference on Decision and Control*, pages 3355–3360, Miami Beach, 2018b.

Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. Data-Rate Constrained Observers of Nonlinear Systems. *Entropy*, 21(282):1–29, 2019.

Q. Voortman, D. Efimov, A. Y. Pogromsky, and J.-P. Richard. An Event-Triggered Observation Scheme for Systems with Perturbations and Data-Rate Constraints. *Submitted to Automatica*, 2020a.

Q. Voortman, D. Efimov, A. Y. Pogromsky, J. P. Richard, and H. Nijmeijer. Event-triggered Data-efficient Observers of Perturbed Systems. In *Proceedings of the 21st IFAC World Congress*, Berlin, 2020b.

Q. Voortman, D. Efimov, A. Y. Pogromsky, J.-p. Richard, and H. Nijmeijer. Synchronization of Perturbed Linear Systems with Data-Rate Constraints. In *Proceedings of the 59th IEEE Conference on Decision and Control*, Jeju Island, 2020c.

Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. Consensus of nonlinear systems with data-rate constraints. In *Proceedings of the 21st IFAC World Congress*, volume 53, pages 11050–11055, 2020d.

Q. Voortman, A. Y. Pogromsky, A. S. Matveev, and H. Nijmeijer. Data-Rate Constrained Consensus in Networks of Dynamical Systems. *Submitted to Automatica*, 2020e.

Q. Voortman, D. Efimov, A. Y. Pogromsky, J.-p. Richard, and H. Nijmeijer. Tracking State with Limited Communications : an Event-Triggered Approach. *Submitted to IEEE Transactions on Automatic Control*, 2021a.

Q. Voortman, D. Efimov, A. Y. Pogromsky, H. Silm, J.-p. Richard, and H. Nijmeijer. Observing Mobile Robots with Data-Rate Constraints : a Case Study. In *Proceedings of the 60th IEEE Conference on Decision and Control*, Austin, 2021b.

W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints - Part I: State estimation problems. *IEEE Transactions on Automatic Control*, 42(9):1294–1299, 1997.

M. Xia, V. Gupta, and P. J. Antsaklis. Networked State Estimation Over a Shared Communication Medium. *IEEE Transactions on Automatic Control*, 62(4):1729–1741, 2017.

H. Yamaguchi, T. Arai, and G. Beni. A distributed control scheme for multiple robotic vehicles to make group formations. *Robotics and Autonomous Systems*, 36(4):125–147, 2001.

J. K. Yook, D. M. Tilbury, and N. R. Soparkar. Trading computation for bandwidth: Reducing communication in distributed control systems using state estimators. *IEEE Transactions on Control Systems Technology*, 10(4):503–518, 2002.

K. You and L. Xie. Network topology and communication data rate for consensusability of discrete-time multi-agent systems. *IEEE Transactions on Automatic Control*, 56(10):2262–2275, 2011.

L.-S. Young. Entropy, Lyapunov Exponents, and Hausdorff Dimension in Differentiable Dynamical Systems. *IEEE Transactions on Circuits and Systems*, CAS-30(8):599–607, 1983.

S. Yüksel and T. Başar. *Stochastic Networked Control Systems*. Springer New York Heidelberg Dordrecht London, 2013.

L. Zaccarian. DC motors: dynamic model and control techniques. *Seminar handouts*, pages 1–23, 2005.

F. Zhang, M. Mazo, and N. van de Wouw. Absolute Stabilization of Lur'e Systems Under Event-Triggered Feedback. *IFAC-PapersOnLine*, 50(1):15301–15306, 2017.

Acknowledgements

Many different people participated in the journey that leads me to write these words today. For years, I dreamt of becoming a doctor. This dream will soon be a reality but it wouldn't be if it weren't for the amazing people that I met along the way. What follows is an attempt at thanking them for all they gave to me.

Allereerst zou ik Henk willen bedanken, bedankt voor u begeleiding tijdens mijn promotie. Dankzij u heb ik gedurende de laatste vier jaren ontzettend veel geleerd. Talloze keren heeft u nieuwe onderzoeksideën kunnen brengen aan dit project. Onze maandelijkse bijeenkomsten dienden als leidraad voor het proefschrift. Bedankt voor het vertrouwen dat u in mij heeft gesteld.

Jean-Pierre, je te remercie pour ta supervision pendant ces années. Tu as toujours su apporter le soutien nécessaire à l'avancement du projet tant au niveau de la recherche que des fastidieuses démarches administratives. Merci pour la justesse de tous les commentaires que tu as pu apporter à mes écrits, et pour la porte que tu m'as ouverte vers un monde de recherche que je ne connaissais pas.

I also want to thank Sasha, for the many fruitful discussions we have had. You introduced me to one of your favourite domains of research: control with data-rate constraints. You were always available to help me understand it better. Thank you for providing support during the more difficult periods of the thesis.

Denis, you greatly supported me during these years. Thank you for always being there to discuss my many questions and providing accurate solutions every time. I always appreciated the speed and efficiency at which you could provide the right comment to unblock when I was stuck on problems.

A sincere thank you to professor Maurice Heemels, professor Serdar Yüksel, and doctor Sophie Tarbouriech for accepting to be part of the defence committee and taking the time to review this manuscript.

I would like to thank professor Alexey Matveev for collaborating with me on several papers. Thank you for your patience while working through my mistakes. I was always impressed by the precision of your remarks and your writings.

Graag zou ik ook Wim willen bedanken. Dankzij u is het UCoCoS Project een groot succes geworden, en dankzij u heb ik een plaats kunnen vinden in het project. Laurentiu, merci de m'avoir soutenu dans l'organisation des séminaires

doctorants et inclus dans l'organisation des évènements. Rosane, merci pour l'organisation des séminaires et apéros du vendredi qui ont su enrichir et égayer les deux dernières années de ma thèse. Geertje, bedankt voor de praktische steun en je altijd blije humeur.

Comrade Rogov, colleagues at first, friends short thereafter. Sharing an office with you was a pleasure and so was spending three months in Lille with you. There were many great memories and many more to come. Thank you, Haik, for the collaboration on the mobile robots projects which ended up being a chapter of this thesis. Jijju, Libo, Deesh aka the rest of the UCoCoS crew: thank you for all the great moments (both work and private life related) we shared together. Alex, Cyrille, Nelson, Artem, Anatolii, Fiodar, Francois, Wenjie, Kostia, and Nicolas aka the Lille crew: you really embellished the final years of my PhD. I was happy to share an office (and a few (hobo) beers) with all of you guys. Brandon, Timo, Daniël, Arviandy, Robbin, Wouter, Isaac, with whom I spent the first two years in Eindhoven: thanks for collaborating on the DISC assignments and thank you for the enjoyable discussions.

Ces remerciements ne seraient pas complets sans un mot pour Edward, Muriel et Marie. Edward, parce-que grace à tes conseils avisés, j'ai réalisé tellement de choses. Jamais je n'aurai cru tout ceci possible il y a 5 ans ! Muriel parceque sans votre impulsion, sans la confiance que vous m'avez accordée, et sans l'Arboretum College, je n'aurais très probablement jamais entamé cette thèse. Je vous en serai toujours redevable. Finalement, Marie Martin, merci de m'avoir orienté sur la voie du doctorat, en me faisant comprendre qu'il y avait là une vocation à nourrir.

Een aantal leerkrachten van het Sint-Jan Berchmanscollege te Brussel hebben mijn verlangen om (wiskunde) te leren opgewekt. Hiervoor dank ik, onder andere, mijnheer Uyttersprot, mevrouw Du Ville, mijnheer Coppens en mevrouw Van Iseghem.

Merci à mes amis, la compagnie desquels a servi d'excellente motivation pour continuer à bosser pendant toutes ces années: Aurian, Max, Martin, Thomas, No, Oli, Jaaf et tous ceux que j'aurais pu oublier. My thanks of course also go to the Russian crew in Eindhoven: Dania, Ksu and Nikka. I will always cherish those precious moments we had in Eindhoven together. You made life in this foreign country so much more enjoyable.

Le remerciement final va à ma famille. D'abord au premier ingénieur Voortman: Bon-Papa. Avec Bonne-Maman vous avez ouvert la voie à trois générations d'universitaires accomplis et maintenant au premier docteur. Ensuite Papy et Mamy, qui m'ont toujours soutenu et encouragé dans mes choix d'études. Merci à Raph, Clara, Alex, Poutou, Cha et Gus, qui m'apportent tellement de bonheur dans la vie. Merci à Nicolas, pour tous ces moments inoubliables passés ensemble. Je n'oublie pas non plus le reste de la famille qui m'a toujours encouragé et soutenu. Les derniers mots sont pour vous, Papa et Maman. Merci pour votre soutien indéfectible qui a servi de phare pour guider ma vie. Sans vous je

n'aurais jamais fini ingénieur, et encore moins docteur. Merci pour tout ce que vous m'avez donné toutes ces années !

La dernière pensée quant à elle est pour l'élue de mon cœur, Marina. Merci infiniment pour ton soutien et tes encouragements.

Oppède, the 5th of August 2021.

Curriculum Vitae

Quentin Voortman was born in Etterbeek, Belgium, on the 28th of December 1991. He received his bachelor's degree in engineering and his master's degree in engineering in applied mathematics (cum laude) from the Université Catholique de Louvain in 2013 and 2015 respectively. His master's thesis focused on the aggregation of flexible customers on the energy market into virtual power plants.

After working as a self-employed private teacher for 1.5 years, he started a joint PhD between Eindhoven University of Technology and École Centrale de Lille in 2017, in a Marie-Curie project entitled: Understanding and Controlling Complex Systems (UCoCoS). His PhD was supervised by Henk Nijmeijer and Jean-Pierre Richard. The research focused on interactions between dynamical systems and communication technologies. The result of this research is contained in the present PhD dissertation. For his paper entitled Continuous Time Observers of Nonlinear Systems with Data-Rate Constraints he obtained an IFAC Young Author Award at the 5th IFAC Conference on Analysis and Control of Chaotic Systems. Since May 2021, he is employed at Flanders Make.