
Analysis of Networked Systems

F. Coloniuss, U. Helmke, J. Jordan, C. Kawan, R. Sailer, and F. Wirth

This chapter deals with particular properties of dynamic systems that are important when controlling the system over digital networks. It is shown how the observability of the system depends upon the network properties. Then the minimal bit rate that is necessary to solve a control task is considered. Finally, a method for the dynamic quantization of the feedback information that is used for stabilizing an unstable plant is developed.

Chapter Contents

2.1 Overview	32
2.2 Observability of Networked Systems	34
2.2.1 Motivation and Earlier Results	34
2.2.2 Reachability and Observability of Networks	36
2.2.3 Sensitivity Analysis of Networks	47
2.3 Minimal Bit Rates and Entropy for Control Tasks ...	52
2.3.1 Motivation and Earlier Results	52
2.3.2 Invariance Entropy	53
2.3.3 Entropy for Exponential Stabilization	62
2.4 Dynamic Quantization for Feedback Stabilization with Delayed Data-Rate Limited Communication	69
2.4.1 Problem Statement	69
2.4.2 Dynamic Quantization	69
2.4.3 Markovian Communication Models	76

2.1 Overview

In this chapter fundamental properties of networked control systems are discussed that characterize obstructions to and requirements for the control of interconnections through digital channels. On the one hand, this chapter relates to general problems of the characterization of controllability and observability properties of interconnected systems, where information is distributed via communication channels. On the other hand, it is of interest to characterize the necessary bandwidth required to control a system or an interconnection of systems. The latter question is related to dynamic properties of the physical system as well as the communication channel under consideration.

Interconnected dynamic systems describe a rather broad class of networked control systems where finitely many subsystems or ‘agents’ are interconnected to construct a network of dynamic systems. The interconnection structure can be static or dynamic via physical couplings or through communication channels. In either case, the interconnection structure is described by a coupling graph that defines the information flow in the network. Controls are inserted into the network to be broadcasted to all nodes and may thus be non-decentralized. Structural properties of such networks refer to questions such as e.g. reachability and observability, the estimation of the amount of necessary information to be broadcasted in order to e.g. stabilize the network, or questions about quantization and stability. The control-theoretic analysis of such interconnected systems has already led to a number of fundamental insights, which are briefly described next.

The status of the theory concerning control of one or several dynamic systems using multiple sensors and actuators over a digital communication network is far from being complete. Here, in spite of the rapid growth of communication capacities, the minimal bit rate for communication may be a limit for control performance. This poses fundamental mathematical questions. Although these questions have already been realized in the 1990’s, they are not yet satisfactorily solved. It is still a challenging task to determine minimal bit rates for standard control problems like stabilizing a of single dynamic system or rendering a subset of the state space invariant. It turns out that relations to the mathematical theory of (uncontrolled) dynamic systems involving topological entropy appear to be a key to the understanding of these problems. Here our work has essentially been restricted to these fundamental questions for single control systems. It is our hope that the results will contribute to the mathematical tools for understanding minimal bit rates for interconnected systems, which presently are under investigation.

A complete characterization of the reachability and observability properties of standard parallel, series and feedback interconnections of linear systems is well-known since the early 1970’s. While parallel interconnections of identical linear systems can never be controllable or observable, for nonlinear systems this fact is no longer true. It is known that the parallel coupling of identical nonlinear systems can be locally accessible, at least for generic

initial states. The extension of such results for general interconnection structures is an open research problem. For arbitrary homogeneous networks of linear systems, i.e., for networks of identical SISO systems, a very concise characterization of reachability and observability is known [165]. However, the extension of such results to arbitrary heterogeneous networks of control systems is an open research problem.

This chapter presents several new results on fundamental limitations for the control and estimation of interconnected systems:

- **Observability of networked systems:** Section 2.2 explores the extensions of the established observability and reachability analysis of homogeneous networks to the general case of heterogeneous networks of linear systems. A basic characterization of generic structural observability of networks of first order driftless integrators is generalized to arbitrary homogeneous networks of SISO systems. Sufficient conditions for observability of the series connection of two nonlinear systems are presented. New results on the reachability of switching networks are derived, where the interconnection parameters are considered as independent control variables.
- **Minimal bit rates for control:** Section 2.3 surveys entropy estimates for the invariance of subsets of the state space and establishes relations between invariance entropy and minimal data rates. In particular, for control sets the minimal sum of positive Lyapunov exponents along periodic solutions in the control set provides an upper bound for the strict invariance entropy. As an example, the invariance entropy for a controlled linear mathematical pendulum near the unstable position is computed. Furthermore, for nonlinear control systems, entropy estimates for the stabilization at an equilibrium with fixed exponential rate are given, which are exact in the linear case.
- **Dynamic quantization for feedback stabilization:** Section 2.4 discusses the concept of dynamic quantization and study the problem of minimal data rates in the context of this particular approach. An encoder/decoder scheme is presented for channels characterized by quantized, delayed packet-based communication with the additional possibility of packet drops. The scheme guarantees stabilization provided that the required bandwidth is available as prescribed by easily identifiable system parameters. The novel idea introduced to the classic concept of dynamic quantization is that time stamps in data packets as well as a feedback loop in the communication channel using acknowledgments are sufficient to handle unreliable packet delivery. This approach can also be analyzed for communication channels of a stochastic nature in which delays and packet delivery are described by a Markov process.

The subsequent research marks the beginning of a new theory of networked control system that is in close analogy to Shannon's theory. The development

of mathematical tools for establishing fundamental limits of performance together with finding effective algorithms to achieve these limits present challenges for the future system-theoretic research.

2.2 Observability of Networked Systems

2.2.1 Motivation and Earlier Results

Large-scale networks of interconnected dynamic systems pose a number of challenging tasks for the theoretical analysis of control-theoretic issues. For example, in analyzing sensor-/actuator networks or large-scale biological networks it becomes increasingly important to estimate the state variables, to identify relevant system parameters, or to characterize the most sensitive interconnection parameters. This leads to the important questions of reachability, observability, state estimation and observer design for networks of systems.

There is a recent resurgence of interest in the analysis of controllability properties for parallel interconnection structures. The paper [39] may serve as a good starting point. In [39], controllability properties of a parallel connection of finitely many nonlinear systems are established. Somewhat surprisingly, local accessibility can hold even if the systems are identical. Early work in this area focused on the analysis of standard structures for linear time-invariant systems, such as series [46], parallel [128] and feedback interconnections [46].

In this section, our main focus is on the observability analysis of heterogeneous networks, whose interconnection structure is fixed by a coupling graph. For homogeneous networks of identical linear SISO systems a characterization of reachability has been first obtained by [166]; see also [129] for a more concise proof. The more general scenario that we consider is that of a finite number of arbitrary linear or nonlinear control systems, called *node systems*, that are connected through static or dynamic coupling laws. Since communication between the systems is mainly assumed to occur through the interconnection channels, dynamic coupling laws can model the effects of delays or time-varying topologies in the network. The models for the node systems are defined by first order or higher order difference or differential equations. We show that using higher order polynomial matrix fraction decompositions, allows one to obtain very compact coprimeness conditions for reachability and observability. Our main tool is a permanence principle for strict system equivalence that implies very concise results on observability and reachability of interconnected systems. This leads to elegant new proofs for classical conditions for reachability and observability of series, parallel and feedback interconnections. We also derive new results for series or parallel connections of homogeneous networks.

Another direction of research deals with the situation of networks that are controlled by switching interconnections; see e.g. [227]. Thus the interconnections are considered as input variables that are tuned to achieve better

performance of the overall network. In [40], BROCKETT has characterized the system Lie algebras of feedback systems, where the feedback gains act as control variables. In [103] we have generalized this characterization to homogeneous networks of SISO systems.

For graph-theoretic characterizations for generic controllability and generic observability for structured systems we refer to [100, 232]. This early work applies to networks whose node dynamics are described by simple first-order integrators. Using recent results from [129], we obtain a generalization to SISO node systems with arbitrary linear dynamics. In [401], we analyze the zero properties of blocked linear systems resulting from blocking of linear time-invariant systems. This is an important step towards analyzing zeros of networks with periodic interconnection topologies. Global observability of Morse-Smale nonlinear systems has been analyzed in [1]. In [102], we extended these results to a global observability result on real analytic systems, using weak assumptions on the dynamics of the limit sets. Our work is also connected with results on synchronizing subspaces and geometric control introduced in Section 6.3. Entropy bounds for conditioned invariant subspaces of linear systems were established in [69].

To analyze the behavior of large-scale networks of interacting dynamic systems it is of vital interest to identify those components which are highly sensitive to errors. It is well known that round-off noise and quantization errors in the implementation of linear systems in digital processing devices might lead to a substantial different performance, cf. [268]. This gave rise to a systematic sensitivity analysis for single linear systems, see [133], [173]. These investigations show that the minimum sensitivity realizations are balanced realizations. Since minimum sensitivity realizations generalize the class of balanced realizations, there is a natural relation to balanced truncation and model reduction.

In this project, we examined the sensitivity analysis from the network perspective. Precisely, we consider networks of linear discrete-time systems and develop a systematic sensitivity theory. To this end, we define a sensitivity measure that assigns to each realization of the network structure a value quantifying its sensitivity with respect to network parameter variations. In addition, new sensitivity Gramians are introduced so that the network sensitivity can be measured by the sum of their traces. In a first step we consider homogeneous networks of single-input-single-output systems and concentrate on the network sensitivity with respect to input-to-state and state-to-output parameter variations. It turns out that the network has minimal sensitivity if and only if the associated controllability and observability sensitivity Gramians are equal. This can be interpreted in the way that symmetric networks are sensitivity optimal. Furthermore, we address the issue of existence and uniqueness of sensitivity optimal network realizations. Moreover, we show that sensitivity optimal realizations can be computed by a Newton algorithm which is locally quadratically convergent. Dependent on the dynamics of the node systems we derive estimates of the sensitivity Gramians in terms of the classical Gramians corresponding to the realization of the network topology

and the realization of the overall network. These estimates result in some interesting perspectives on model reduction based on sensitivity optimal design, which is briefly addressed, cf. [176]. The next step is to extend these results to heterogeneous networks of multi-input-multi-output linear systems. This is the content of current and future research, cf. [177].

2.2.2 Reachability and Observability of Networks

Models of Interconnected Systems. We consider heterogeneous networks described interconnecting N linear discrete-time node systems

$$\begin{aligned} \mathbf{x}_i(k+1) &= \boldsymbol{\alpha}_i \mathbf{x}_i(k) + \boldsymbol{\beta}_i \mathbf{v}_i(k) \\ \mathbf{w}_i(k) &= \boldsymbol{\gamma}_i \mathbf{x}_i(k), \quad i = 1, \dots, N \end{aligned} \quad (2.1)$$

with $\boldsymbol{\alpha}_i \in \mathbb{F}^{n_i \times n_i}$, $\boldsymbol{\beta}_i \in \mathbb{F}^{n_i \times m_i}$ and $\boldsymbol{\gamma}_i \in \mathbb{F}^{p_i \times n_i}$. Here \mathbb{F} is an arbitrary field. As a standard assumption, each system $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \boldsymbol{\gamma}_i)$, $(i = 1, \dots, N)$ is assumed to be observable and controllable. The systems are coupled via a static linear *coupling law*

$$\mathbf{v}_i(k) = \sum_{j=1}^N \mathbf{L}_{ij} \mathbf{w}_j(k) + \mathbf{B}_i \mathbf{u}(k) \in \mathbb{F}^{m_i} \quad (2.2)$$

with $\mathbf{L}_{ij} \in \mathbb{F}^{m_i \times p_j}$ and $\mathbf{B}_i \in \mathbb{F}^{m_i \times m}$. Here, the matrix $\mathbf{L} = (\mathbf{L}_{ij}) \in \mathbb{F}^{\overline{m} \times \overline{p}}$ with $\overline{m} := m_1 + \dots + m_N$ and $\overline{p} := p_1 + \dots + p_N$ models the *interconnection structure*. The *interconnection output* is given by

$$\mathbf{y}(k) = \sum_{i=1}^N \mathbf{C}_i \mathbf{w}_i(k) \quad \text{with } \mathbf{C}_i \in \mathbb{F}^{p \times p_i}, \quad i = 1, \dots, N.$$

The *input/output interconnection* is defined by the matrices

$$\mathbf{C} := (\mathbf{C}_1, \dots, \mathbf{C}_N) \in \mathbb{F}^{p \times \overline{p}} \quad \text{and} \quad \mathbf{B} := (\mathbf{B}_1^T, \dots, \mathbf{B}_N^T)^T \in \mathbb{F}^{\overline{p} \times m}$$

with $\overline{n} := n_1 + \dots + n_N$. Let $\mathbf{x}(k) = (\mathbf{x}_1^T(k), \dots, \mathbf{x}_N^T(k))^T \in \mathbb{F}^{\overline{n}}$ denote the global state of the network. Of course, more complicated, dynamic couplings laws are also of interest, in particular towards modeling communication constraints, packet loss and so on. Certainly, the analysis of any such general dynamic network structures rests on a solid understanding of the influence of static coupling laws. Therefore, we focus our subsequent analysis on static interconnections.

Define

$$\begin{aligned} \mathbf{A}_0 &:= \text{diag}(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N) \in \mathbb{F}^{\overline{n} \times \overline{n}}, \\ \mathbf{B}_0 &:= \text{diag}(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_N) \in \mathbb{F}^{\overline{n} \times \overline{m}}, \\ \mathbf{C}_0 &:= \text{diag}(\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_N) \in \mathbb{F}^{\overline{p} \times \overline{n}}. \end{aligned}$$

Then the interconnected system has the following state-space representation

$$\begin{aligned}\mathbf{x}(k+1) &= \bar{\mathbf{A}}\mathbf{x}(k) + \bar{\mathbf{B}}\mathbf{u}(k) \\ \mathbf{y}(k) &= \bar{\mathbf{C}}\mathbf{x}(k)\end{aligned}\quad (2.3)$$

with

$$\bar{\mathbf{A}} := \mathbf{A}_0 + \mathbf{B}_0\mathbf{L}\mathbf{C}_0 \in \mathbb{F}^{\bar{n} \times \bar{n}}, \quad \bar{\mathbf{B}} := \mathbf{B}_0\mathbf{B} \in \mathbb{F}^{\bar{n} \times m}, \quad \bar{\mathbf{C}} := \mathbf{C}\mathbf{C}_0 \in \mathbb{F}^{\bar{n} \times m}.$$

As a special case we also consider *homogeneous networks*, i.e., the case where $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1 = \dots = \boldsymbol{\alpha}_N$, $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \dots = \boldsymbol{\beta}_N$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 = \dots = \boldsymbol{\gamma}_N$ denote identical SISO systems. The state space representation of a homogeneous network has the tensor product form

$$\bar{\mathbf{A}} = I_N \otimes \boldsymbol{\alpha} + \mathbf{L} \otimes \boldsymbol{\beta}\boldsymbol{\gamma}, \quad \bar{\mathbf{B}} = \mathbf{B} \otimes \boldsymbol{\beta} \quad \text{and} \quad \bar{\mathbf{C}} = \mathbf{C} \otimes \boldsymbol{\gamma}. \quad (2.4)$$

In order to analyze the influence of the interconnection matrices on the reachability and observability properties it is convenient to describe the interconnected systems in terms of the transfer functions of the node systems. Define the *i-th node transfer function* as the strictly proper transfer function

$$\mathbf{G}_i(z) = \boldsymbol{\gamma}_i(z\mathbf{I} - \boldsymbol{\alpha}_i)^{-1}\boldsymbol{\beta}_i. \quad (2.5)$$

of McMillan degree n_i . We define the *node transfer function* as

$$\mathbf{G}(z) := \text{diag}(\mathbf{G}_1(z), \dots, \mathbf{G}_N(z)) = \mathbf{C}_0(z\mathbf{I} - \mathbf{A}_0)^{-1}\mathbf{B}_0.$$

In the case that \mathbf{L} is a square matrix we define the *interconnection transfer function* as

$$\boldsymbol{\phi}(z) = \mathbf{C}(z\mathbf{I} - \mathbf{L})^{-1}\mathbf{B}.$$

The global *network transfer function* then is defined as

$$\bar{\boldsymbol{\Phi}}(z) = \bar{\mathbf{C}}(z\mathbf{I} - \bar{\mathbf{A}})^{-1}\bar{\mathbf{B}}. \quad (2.6)$$

It is useful to describe the network transfer function explicitly in terms of coprime factorizations of the node transfer function. Consider the case where the *i*-th node transfer function is strictly proper and is given in the Rosenbrock form as

$$\mathbf{G}_i(z) = \mathbf{V}_i(z)\mathbf{T}_i^{-1}(z)\mathbf{U}_i(z). \quad (2.7)$$

The pairs $(\mathbf{T}_i(z), \mathbf{U}_i(z))$ and $(\mathbf{T}_i(z), \mathbf{V}_i(z))$ are assumed to be left or right coprime polynomial matrices, respectively. Moreover, we assume that the factorizations are minimal in the sense that $\deg \det \mathbf{T}_i(z) = n_i$. Let

$$\begin{aligned}\mathbf{T}(z) &= \text{diag}(\mathbf{T}_1(z), \dots, \mathbf{T}_N(z)) \\ \mathbf{U}(z) &= \text{diag}(\mathbf{U}_1(z), \dots, \mathbf{U}_N(z)) \\ \mathbf{V}(z) &= \text{diag}(\mathbf{V}_1(z), \dots, \mathbf{V}_N(z)).\end{aligned}$$

Then we have the left- and right coprime factorization of a strictly proper transfer function

$$\mathbf{G}(z) = \mathbf{V}(z)\mathbf{T}(z)^{-1}\mathbf{U}(z).$$

Note that $\mathbf{T}(z) - \mathbf{U}(z)\mathbf{L}\mathbf{V}(z)$ is a nonsingular polynomial matrix for any constant interconnection matrix \mathbf{L} . Thus the transfer network function from \mathbf{u} to \mathbf{y} is well defined and given by

$$\bar{\Phi}(z) = \mathbf{C}\mathbf{V}(z)(\mathbf{T}(z) - \mathbf{U}(z)\mathbf{L}\mathbf{V}(z))^{-1}\mathbf{U}(z)\mathbf{B}.$$

This explicit representation of the network transfer function turns out to be very useful for analyzing the system-theoretic properties of the network.

Conditions for Reachability and Observability. In [129] we extended the classical notion of strict system equivalence to networks of linear systems. This was the key to proving the following extension of the classical Hautus criterion.

Theorem 2.1.

a) $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ is reachable if and only if

$$\text{rank}(\mathbf{T}(z) - \mathbf{U}(z)\mathbf{L}\mathbf{V}(z), \mathbf{U}(z)\mathbf{B}) = \bar{n} \quad \forall z \in \mathbb{C}.$$

b) $(\bar{\mathbf{C}}, \bar{\mathbf{A}})$ is observable if and only if

$$\text{rank} \begin{pmatrix} \mathbf{T}(z) - \mathbf{U}(z)\mathbf{L}\mathbf{V}(z) \\ \mathbf{C}\mathbf{V}(z) \end{pmatrix} = \bar{n} \quad \forall z \in \mathbb{C}.$$

In [129] this result is extended to dynamic coupling schemes. In particular, this result yields new and efficient algebraic tests for controllability and observability of networked control systems. It is a generalization of a preliminary result by Hara et. al. for homogeneous networks, i.e., from systems of type (2.4), to arbitrary heterogeneous networks. More precisely, Theorem 2.1 implies the following result.

Theorem 2.2. [129, 166] *A homogeneous network is reachable (observable) if and only if the realization $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ of the interconnection transfer function $\phi(z)$ is reachable (observable). In particular, reachability of $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$ is independent of the choice of the node transfer function $g(z)$, as long as $g(z)$ is scalar rational and strictly proper.*

Another application of Theorem 2.1 is to derive graph-theoretic conditions for *generic observability* of homogeneous networks, for $\mathbb{F} = \mathbb{C}$. We assume that for a given set of edges $E \subset \{1, \dots, N\} \times \{1, \dots, N\}$ we have $L_{ij} = 0$ if $(i, j) \notin E$. The directed graph $\mathcal{G}(E, \{1, \dots, N\})$ is called the *coupling graph*.

The set of structured matrices $\mathbf{L} \in \mathbb{C}^{N \times N}$ which corresponds to a given coupling graph $\mathcal{G}(E, \{1, \dots, N\})$ is denoted by $D_{\mathcal{G}}$. Note that $D_{\mathcal{G}}$ is a vector space of dimension $|E|$. In the following we fix the system parameters $(\alpha, \beta, \gamma, \mathbf{B}, \mathbf{C})$ and the coupling graph \mathcal{G} and consider the coupling parameters $L_{ij} \in \mathbb{C}$ with $(i, j) \in E$ as free parameters. Thus the set of admissible couplings is equal to $\mathbb{C}^{|E|}$. System (2.3) is *generically controllable (generically observable)* if it is controllable (observable) for a generic set of parameters \mathbf{L} with. Here, we say a nonempty subset $M \subset D_{\mathcal{G}}$ is generic if M is not contained in the zero set of a nontrivial polynomial in $D_{\mathcal{G}}$.

In the special case $n = 1$ and $\alpha = 0$ and $\beta = \gamma = 1$ this concept reduces to the well-known concept of generic observability for structured matrices (see for example [100]), i.e., the system

$$\mathbf{x}(k+1) = \mathbf{L}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k), \quad \mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) \quad (2.8)$$

with a structured matrix \mathbf{L} . Theorem 2.1 yields the following connection between linear structured systems and homogeneous SISO networks.

Corollary 2.1. *The homogeneous network (2.4) for $\mathbb{F} = \mathbb{C}$ is generically controllable (generically observable) if and only if the structured system (2.8) is generically controllable (generically observable).*

Note that there exist graph-theoretic characterizations for generic controllability and generic observability for structured systems, see [100, 232]. More precisely, a structured system is generically controllable if the corresponding structure graph fulfills the cactus-condition, i.e., in \mathcal{G} there exists a vertex disjoint union of cacti that covers all state vertices.

Observability under Transmission Delays and Switching. As a special case of the general dynamic coupling laws we characterize reachability and observability for systems with delayed interconnections

$$\mathbf{v}_i(t) = \sum_{j=1}^N \mathbf{L}_{ij} \mathbf{w}_j(t - \lambda_{ij}) + \mathbf{B}_i \mathbf{u}(t),$$

where λ_{ij} are assumed to be arbitrary nonnegative integers. This model can be realized as follows using a dynamic interconnection law $\mathbf{E}(\sigma)\mathbf{v} = \mathbf{A}(\sigma)\mathbf{w} + \mathbf{B}(\sigma)\mathbf{u}$, where $\mathbf{E}(z)$ is a square nonsingular polynomial matrix and $\mathbf{A}(z), \mathbf{B}(z)$ appropriately sized polynomial matrices. We assume that the network transfer function is given in terms of coprime factorizations of the node transfer functions

$$\mathbf{G}_i(z) = \mathbf{Q}_i(z)^{-1} \mathbf{P}_i(z) = \overline{\mathbf{P}}_i(z) \overline{\mathbf{Q}}_i(z)^{-1} \quad (2.9)$$

where we assume that the factorizations are minimal in the sense that $\delta(\mathbf{G}_i) = \deg \det \mathbf{Q}_i = \deg \det \overline{\mathbf{Q}}_i$. We denote $\lambda_i := \max_{j=1, \dots, N} \lambda_{ij}$, $i = 1, \dots, N$ and define $\bar{\lambda}_{ij} = \lambda_i - \lambda_{ij} \geq 0$. Define polynomial matrices $\mathbf{E}(z) = \text{diag}(z^{\lambda_1}, \dots, z^{\lambda_N})$, $\mathbf{A}(z) = (\mathbf{L}_{ij} z^{\bar{\lambda}_{ij}})_{i,j=1, \dots, N} \in \mathbb{F}^{\overline{p} \times \overline{m}}[z]$, $\mathbf{B}(z) := \mathbf{E}(z) \mathbf{B}$ and $\mathbf{C}(z) = \mathbf{C}$.

Theorem 2.3. *The shift realization of the delayed network is reachable if and only if*

$$\text{rank} \left(\mathbf{L}_{ij} z^{\bar{\lambda}_{ij}} \bar{\mathbf{P}}_j(z) - z^{\lambda_i} \bar{\mathbf{Q}}_j(z) \delta_{ij}, z^{\lambda_i} \mathbf{B}_i \right) = \bar{n}, \quad \forall z \in \bar{\mathbb{F}}.$$

The shift realization of the delayed network is observable if and only if

$$\text{rank} \left(\left(\mathbf{L}_{ij} z^{\bar{\lambda}_{ij}} \bar{\mathbf{P}}_j(z) - z^{\lambda_i} \bar{\mathbf{Q}}_j(z) \delta_{ij} \right) \right. \\ \left. \mathbf{C} \bar{\mathbf{P}}(z) \right) = \bar{n}, \quad \forall z \in \bar{\mathbb{F}}.$$

As a special case we consider homogeneous networks of identical SISO systems with node transfer functions $G_i(z) = \frac{p(z)}{q(z)}$ satisfying $p(0) \neq 0$. In this situation the above result implies

Corollary 2.2. *Consider a network of identical node transfer functions*

$G_i(z) = \frac{p(z)}{q(z)}$ *satisfying* $p(0) \neq 0$. *Assume that* \mathbf{L} *is invertible and* $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ *is observable. Assume further that all delays* λ_{ij} *are equal to* $\lambda \geq 1$. *Then the delayed network is observable, independently of the value of* λ .

An interesting class of observability problems for networks of linear systems arises when the network parameters are allowed to vary with time. A special class of interest here are switching networks that are defined by time-varying state interconnection matrices $\mathbf{L}(t)$. Here the interconnection matrices $\mathbf{L}(t)$ are considered as matrix-valued control variables and the task becomes to investigate the controllability and observability properties of the resulting network. This leads to bilinear control and observation problems and therefore requires the use of techniques from geometric nonlinear control.

We begin with a brief summary of some early results on controllability of bilinear systems that are of interest here. In [40] BROCKETT characterized completely the system Lie algebras of the bilinear output-feedback control system

$$\dot{\mathbf{x}}(t) = (\boldsymbol{\alpha} + u(t)\boldsymbol{\beta}\boldsymbol{\gamma}) \mathbf{x}(t) \tag{2.10}$$

for a SISO system $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$.

Theorem 2.4. [40] *Let* $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ *be controllable and observable,* $g(z) = \boldsymbol{\gamma}(z\mathbf{I} - \boldsymbol{\alpha})^{-1}\boldsymbol{\beta}$. *The system Lie algebra of the output-feedback control system (2.10) is:*

- a) $sp(n, \mathbb{R}) \iff g(z) = g(-z)$
- b) $sp(n, \mathbb{R}) \oplus \mathbb{R}I \iff g(r + \alpha) = g(-r + \alpha)$ *for some* $r \neq 0$.
- c) $sl_n(\mathbb{R}) \iff \forall r : g(z + r) \neq g(-z + r), \boldsymbol{\gamma}\boldsymbol{\beta} = \text{trace } \boldsymbol{\alpha} = 0$.
- d) $gl_n(\mathbb{R}) \iff \textit{else}$.

These result shows that any homogeneous network of identical first order integrators, controlled by independent all-to-all interconnection functions is accessible. More generally, consider arbitrary SISO node dynamics and more general interconnection schemes. We have obtained the more general result.

Theorem 2.5. [103] *Let (α, β, γ) be controllable and observable. Let*

$$\mathcal{L}_{\mathcal{G}}(u(t)) := \sum_{(i,j) \in E} u_{ij}(t) L_{ij}$$

be the adjacency matrix of the graph $\mathcal{G} = (E, V)$ with $N > 1$ nodes with independent input function $u_{ij}(t)$. Then the homogeneous network

$$\dot{\mathbf{x}}(t) = (I_N \otimes \alpha + \mathcal{L}_{\mathcal{G}}(u(t)) \otimes \beta \gamma) \mathbf{x}(t) \quad (2.11)$$

is accessible if and only if \mathcal{G} is strongly connected. The system Lie algebra of (2.11) is either $sl_{nN}(\mathbb{R})$ or $gl_{nN}(\mathbb{R})$.

Controllability and Observability of Ensembles of Systems. Spatially-invariant systems, such as the heat equation, provide interesting examples of distributed parameter systems, where control actions and measurements can take place in a spatially distributed way. Using Fourier transform techniques, spatially-invariant control systems can be identified with parameter-dependent families of linear systems; see e.g. [15] for a systematic outline of this approach. In many applications such as quantum control or the controllability of swarms, it is of interest to solve these problems using open-loop controls. Thus, given a family of desired terminal states, we attempt to construct a parameter-independent input function that steers the zero-state to these states, simultaneously for all parameter values. For simplicity we work in a Hilbert space context and focus on the notion of L^2 -ensemble observability and controllability and consider the continuous-time case; the discrete-time case is treated similarly.

We begin with the controllability analysis of parameter-dependent linear systems of the form

$$\begin{aligned} \frac{\partial}{\partial t} \mathbf{x}(t, \theta) &= \mathbf{A}(\theta) \mathbf{x}(t, \theta) + \mathbf{B}(\theta) \mathbf{u}(t) \\ \mathbf{x}(0, \theta) &= \mathbf{0}. \end{aligned} \quad (2.12)$$

The system matrices $\mathbf{A}(\theta) \in \mathbb{R}^{n \times n}$, $\mathbf{B}(\theta) \in \mathbb{R}^{n \times m}$ are assumed to vary continuously in a compact interval $\mathbf{P} := [\theta_{\min} \theta_{\max}] \subset \mathbb{R}$.

Definition 2.1. *The system (2.12) is called L^2 -ensemble controllable, if there exists a finite time $T > 0$ and an input function $\mathbf{u} : [0, T] \rightarrow \mathbb{R}^m$ that steers the initial state $\mathbf{x}(0, \theta) = \mathbf{0}$ in time T into an ε -neighborhood of the desired state $\mathbf{x}_*(\theta)$, simultaneously for all parameters $\theta \in \mathbf{P}$; i.e. if*

$$\left(\int_{\mathbf{P}} \|\mathbf{x}(T, \theta) - \mathbf{x}_*(\theta)\|^2 d\theta \right)^{\frac{1}{2}} < \varepsilon. \quad (2.13)$$

The following statement contains necessary and sufficient conditions that systems (2.12) is L^2 -ensemble controllable.

Theorem 2.6. [174] *Let $\mathbf{P} := [\theta_{\min}, \theta_{\max}] \subset \mathbb{R}$ be a compact interval. A continuous family $(\mathbf{A}(\theta), \mathbf{B}(\theta))$ of linear systems is L^2 -ensemble controllable provided the following conditions are satisfied:*

- (i) $(\mathbf{A}(\theta), \mathbf{B}(\theta))$ is reachable for all $\theta \in \mathbf{P}$.
- (ii) The input Hermite indices $K_1(\theta), \dots, K_m(\theta)$ of $(\mathbf{A}(\theta), \mathbf{B}(\theta))$ are independent of $\theta \in \mathbf{P}$.
- (iii) For any pair of distinct parameters $\theta, \theta' \in \mathbf{P}, \theta \neq \theta'$, the spectra of $\mathbf{A}(\theta)$ and $\mathbf{A}(\theta')$ are disjoint:

$$\sigma(\mathbf{A}(\theta)) \cap \sigma(\mathbf{A}(\theta')) = \emptyset. \quad (2.14)$$

- (iv) For each $\theta \in \mathbf{P}$, the eigenvalues of $\mathbf{A}(\theta)$ have algebraic multiplicity one.

This result is of significance for robust output feedback control of a given linear SISO system $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ which is supposed to be controllable and observable. Consider a compact interval $\mathbf{P} = [\theta_{\min}, \theta_{\max}]$ of gain parameters. Then, for any $\theta \in \mathbf{P}$, the characteristic polynomial of the closed loop system is of the form

$$\det(z\mathbf{I} - (\mathbf{A} + \theta\mathbf{b}\mathbf{c})) = q(z) + \theta p(z)$$

for some appropriate coprime polynomials p, q . This implies that for $\theta \neq \theta' \in \mathbf{P}$, there exists no complex number z so that

$$q(z) + \theta p(z) = 0 = q(z) + \theta' p(z).$$

Corollary 2.3. [174] *Let $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ be a controllable and observable SISO system. Then output feedback system $(\mathbf{A} + \theta\mathbf{b}\mathbf{c}, \mathbf{b})$ is L^2 -ensemble controllable if $\mathbf{A} + \theta\mathbf{b}\mathbf{c}$ has distinct eigenvalues for all $\theta \in [\theta_{\min}, \theta_{\max}]$.*

By duality, Theorem 2.6 can also be used to derive conditions on observability of parameter dependent linear systems. To this end, we consider the parameter-dependent system

$$\begin{aligned} \frac{\partial}{\partial t} \mathbf{x}(t, \theta) &= \mathbf{A}(\theta) \mathbf{x}(t, \theta), \quad \mathbf{x}(0, \cdot) \in L^2(\mathbf{P}, \mathbb{R}^n), \\ \mathbf{y}(t) &= \int_{\mathbf{P}} \mathbf{C}(\theta) \mathbf{x}(t, \theta) d\theta. \end{aligned} \quad (2.15)$$

As before, we assume that $\mathbf{A}(\theta) \in \mathbb{R}^{n \times n}$ and $\mathbf{C}(\theta) \in \mathbb{R}^{p \times n}$ vary continuously in a compact parameter domain $\mathbf{P} \subset \mathbb{R}$.

Definition 2.2. *The system (2.15) is called L^2 -ensemble-observable, if there exists $T > 0$ such that $\mathbf{y}(t) = \mathbf{0}$ on $[0, T]$ implies $\mathbf{x}(0, \theta) = \mathbf{0}$ for all $\theta \in \mathbf{P}$.*

To analyze the observability properties of system (2.15), we note that it is equivalent to the system

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathcal{A}\mathbf{x}(t), & \mathbf{x}(0) &\in L^2(\mathbf{P}, \mathbb{R}^n), \\ \mathbf{y}(t) &= \mathcal{C}\mathbf{x}(t) \end{aligned} \tag{2.16}$$

on the Hilbert space $X = L^2(\mathbf{P}, \mathbb{R}^n)$, where the bounded linear operators \mathcal{A} and \mathcal{C} are defined by

$$\mathcal{A} : X \longrightarrow X, \quad (\mathcal{A}\mathbf{x})(\theta) = \mathbf{A}(\theta)\mathbf{x}(\theta)$$

and

$$\mathcal{C} : X \longrightarrow \mathbb{R}^p, \quad (\mathcal{C}\mathbf{x})(\theta) = \int_{\mathbf{P}} \mathbf{C}(\theta)\mathbf{x}(\theta) \, d\theta,$$

respectively. Note that the notion of L^2 -ensemble observability is equivalent to the notion of approximate observability for linear systems on Hilbert spaces. In particular, (2.16) is approximate observable if and only if the dual system

$$\frac{\partial}{\partial t} \mathbf{x}(t, \theta) = \mathbf{A}(\theta)^T \mathbf{x}(t, \theta) + \mathbf{C}(\theta)^T \mathbf{u}(t), \quad \mathbf{x}(0, \theta) = \mathbf{0} \tag{2.17}$$

is L^2 -ensemble controllable. Thus we can apply Theorem 2.6 to (2.17) and obtain.

Theorem 2.7. [174] *Let $\mathbf{P} = [\theta_{\min} \ \theta_{\max}] \subset \mathbb{R}$ be a compact interval. A continuous family $(\mathbf{A}(\theta), \mathbf{C}(\theta))$ of linear systems is L^2 -ensemble observable, provided the following conditions are satisfied:*

- (i) $(\mathbf{A}(\theta), \mathbf{C}(\theta))$ is observable for all $\theta \in \mathbf{P}$.
- (ii) The output Hermite indices $K_1(\theta), \dots, K_m(\theta)$ of $(\mathbf{A}(\theta), \mathbf{C}(\theta))$ are independent of $\theta \in \mathbf{P}$.
- (iii) For any pair of distinct parameters $\theta, \theta' \in \mathbf{P}, \theta \neq \theta'$, the spectra of $\mathbf{A}(\theta)$ and $\mathbf{A}(\theta')$ are disjoint:

$$\sigma(\mathbf{A}(\theta)) \cap \sigma(\mathbf{A}(\theta')) = \emptyset. \tag{2.18}$$

- (iv) For each $\theta \in \mathbf{P}$, the eigenvalues of $\mathbf{A}(\theta)$ have algebraic multiplicity one.

Global Observability of Nonlinear Systems. As a preliminary step to investigate interconnections of nonlinear systems we discuss observability of nonlinear systems, with special emphasis on state constraints. Using a generalization of an approach by [1], we derive observability conditions for real analytic vector fields, i.e.,

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)), \quad \mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t)) \quad (2.19)$$

on a manifold M . This work is done as a preliminary step to a subsequent observability analysis of interconnected nonlinear systems. Necessary and sufficient condition for observability of analytic systems are due to [74], who showed that observability holds if and only if the observation space \mathcal{O} , spanned by iterated Lie derivatives $L_f^j \mathbf{h}$, separates points on M . However, this condition is hard to test, since the observation space \mathcal{O} is infinite dimensional.

If the limiting dynamics on an attractor of \mathbf{f} is known a-priori, then a sensible observability condition might consist of combining local observability criteria together with some global information on the phase portrait on the attractor. Aeyels[1] has been the first who proposed such an asymptotic dynamics based observability condition for the special class of nonlinear Morse-Smale systems. A lot of examples of particular interest, such as Riccati equations do not satisfy this condition. The following result extends that of [1] to a wide class of systems.

Theorem 2.8. [102] *Let $\mathbf{f} : M \rightarrow TM$ be a real analytic complete vector field on a manifold M and let $\mathbf{h} : M \rightarrow \mathbb{R}^p$ be real analytic. Assume:*

A1: The ω -limit set $\omega(\mathbf{x})$ of any $\mathbf{x} \in M$ is a nonempty compact subset of M .

A2: The system is observable on each ω -limit set $\omega(\mathbf{x})$.

A3: For any points $\mathbf{x}_1, \mathbf{x}_2 \in M$ with $\omega(\mathbf{x}_1) \neq \omega(\mathbf{x}_2)$ one has $\mathbf{h}(\omega(\mathbf{x}_1)) \neq \mathbf{h}(\omega(\mathbf{x}_2))$.

A4: For any $\mathbf{x} \in M$ there exists some $p \in \omega(\mathbf{x})$ such that Σ is locally observable at p .

Then system (2.19) is globally observable on an arbitrary time interval $[0, T]$.

As an application towards interconnected systems we consider the **series connection** scheme

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{z}(t)), \quad \dot{\mathbf{z}}(t) = \mathbf{g}(\mathbf{x}(t)), \quad \mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t)) + \mathbf{k}(\mathbf{z}(t)) \quad (2.20)$$

where M is a compact manifold and $\mathbf{f} : M \times \mathbb{R}^m \rightarrow TM$, $\mathbf{g} : \mathbb{R}^m \rightarrow \mathbb{R}^m$, $\mathbf{h} : M \rightarrow \mathbb{R}^p$, $\mathbf{k} : \mathbb{R}^m \rightarrow \mathbb{R}^p$ are real analytic. The zero dynamics and the driving system are then defined as

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{0}), \quad \mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t)) \quad (2.21)$$

$$\dot{\mathbf{z}}(t) = \mathbf{g}(\mathbf{z}(t)), \quad \mathbf{w}(t) = \mathbf{k}(\mathbf{z}(t)), \quad (2.22)$$

respectively.

Theorem 2.9. [102] *Assume that $\mathbf{0} \in \mathbb{R}^m$ is a globally asymptotic stable equilibrium point of $\dot{\mathbf{z}}(t) = \mathbf{g}(\mathbf{z}(t))$ and (2.22) is observable. Assume further that the zero dynamics (2.21) satisfy*

- (i) *Every solution of (2.21) converges to a compact invariant subset $A \subset M$ and there exist finitely many ω -limit sets $\omega(x_i)$, $x_1, \dots, x_k \in M$ with $A := \bigcup_{i=1}^k \omega(x_i)$.*
- (ii) *Each $\omega(x_i)$ is minimal, nonempty, compact and distal.*
- (iii) *$A = \bigcup_{i=1}^k \omega(x_i)$ satisfies the non cycle condition, i.e. for any indices $1 \leq j_1 < \dots < j_r \leq k$ and $\Lambda_j := \omega(x_j)$ there exist no $q_1, \dots, q_r \notin A$, such that*

$$\alpha(q_i) \subset \Lambda_{j_i} \quad , \quad \omega(q_i) \subset \Lambda_{j_{i+1}}, \quad i = 1, \dots, r,$$

where $\Lambda_{j_{r+1}} := \Lambda_{j_1}$.

- (iv) *For each $i = 1, \dots, k$ there exists $a_i \in \omega(x_i)$ such that $h|_{\omega(x_i)}^{-1}(h(a_i)) = \{a_i\}$.*
- (v) *$h(\omega(x_i)) \neq h(\omega(x_j))$ holds whenever $i \neq j$.*
- (vi) *For any $i = 1, \dots, k$ there exists $p_i \in \omega(x_i)$ such that (2.21) is locally observable at p_i .*

Then system (2.20) is globally observable on an arbitrary time interval $[0, T]$.

As another application of Theorem 2.8 we consider linear perspective system on the projective space $\mathbb{P}(\mathbb{F}^n)$ of all real, and respectively, complex lines through the origin. Let $\langle \mathbf{X} \rangle$ denote the image space of a matrix \mathbf{X} . The system is given by

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X}, \quad \mathbf{Y} = \mathbf{C}\langle \mathbf{X} \rangle, \quad (2.23)$$

whose output is the image space $\langle \mathbf{X} \rangle$ under $\mathbf{C} \in \mathbb{F}^{p \times n}$. Our basic assumption is $p := \text{rank} \mathbf{C} \geq 2$. Equation (2.23) can be regarded as a real analytic system that acts on $\mathbb{P}(\mathbb{F}^n)$ via

$$\langle \mathbf{X} \rangle \mapsto e^{\mathbf{A}t} \langle \mathbf{X} \rangle \in \mathbb{P}(\mathbb{F}^n). \quad (2.24)$$

The corresponding output map is given by

$$\langle \mathbf{X} \rangle \mapsto \mathbf{Y} = \mathbf{C}\langle \mathbf{X} \rangle \in \mathbb{P}(\mathbb{F}^p). \quad (2.25)$$

Note that the image of (2.25) is not entirely contained in $\mathbb{P}(\mathbb{F}^p)$. However, $\mathbf{C}\langle \mathbf{X} \rangle$ belongs to $\mathbb{P}(\mathbb{F}^p)$ for almost all $\langle \mathbf{X} \rangle \in \mathbb{P}(\mathbb{F}^n)$. Then, the pair

(\mathbf{C}, \mathbf{A}) is called *perspectively observable* on $\mathbb{P}(\mathbb{F}^n)$ if for any two subspaces $\mathcal{V}_1, \mathcal{V}_2 \in \mathbb{P}(\mathbb{F}^n)$ the implication

$$\mathbf{C}e^{\mathbf{A}t}\mathcal{V}_1 = \mathbf{C}e^{\mathbf{A}t}\mathcal{V}_2 \text{ for almost all } t \in \mathbb{R} \implies \mathcal{V}_1 = \mathcal{V}_2 \tag{2.26}$$

holds. Note again that we do not require equality for *all* $t \in \mathbb{R}$ in the above definition.

To state sufficient conditions for perspectiveobservability we need the following concepts. A matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ is called *strongly cyclic* if $\text{Real } \lambda_i \neq \text{Real } \lambda_j$ for all different eigenvalues of \mathbf{A} . Thus there is only one Jordan block of \mathbf{A} corresponding to eigenvalues with a fixed real part. Let $\lambda_1, \dots, \lambda_r \in \mathbb{C}$ denote the distinct real or complex eigenvalues of $\mathbf{A} \in \mathbb{R}^{n \times n}$ and let $l \leq 2r$ be the number of non-real, complex-conjugate pairs of eigenvalues. Then, $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called *strongly regular* if $r = n$ and $\text{Real } \lambda_i \neq \text{Real } \lambda_j$ for $i \neq j$ and $\lambda_i \neq \overline{\lambda_j}$. Moreover, $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to satisfy the *irrationality condition* if two eigenvalues $i \neq j$, $\lambda_i \neq \overline{\lambda_j}$ and $\lambda_i, \lambda_j \notin \mathbb{R}$ a rationally independent.

Theorem 2.10. [101]

1) Assume that $\mathbf{A} \in \mathbb{C}^{n \times n}$ is strongly cyclic. Then the following statements are equivalent:

- (a) The pair (\mathbf{C}, \mathbf{A}) is *perspectively observable* on $\mathbb{P}(\mathbb{C}^n)$.
- (b) $\dim \mathbf{C}\mathcal{V} = \dim \mathcal{V}$ holds for each complex 2-dimensional \mathbf{A} -invariant subspaces \mathcal{V} .
- (c) For all $\alpha, \beta \in \mathbb{C}$ one has

$$\text{rank} \begin{pmatrix} \mathbf{A}^2 + \alpha\mathbf{A} + \beta\mathbf{I}_n \\ \mathbf{C} \end{pmatrix} = n.$$

2) Assume that $\mathbf{A} \in \mathbb{R}^{n \times n}$ is strongly regular and satisfies the irrationality condition. Then the following statements are equivalent:

- (a) The pair (\mathbf{C}, \mathbf{A}) is *perspectively observable* on $\mathbb{P}(\mathbb{R}^n)$.
- (b) For each 1- and 2-dimensional \mathbf{A} -invariant subspaces \mathcal{V} one has the equality $\dim \mathbf{C}\mathcal{V} = \dim \mathcal{V}$.
- (c) For all $\alpha, \beta \in \mathbb{R}$ one has

$$\text{rank} \begin{pmatrix} \mathbf{A}^2 + \alpha\mathbf{A} + \beta\mathbf{I}_n \\ \mathbf{C} \end{pmatrix} = n.$$

Note that the complex case in Theorem 2.10 covers earlier re-derives results from [134]. For recent results an efficient tests for perspective observability we refer to [264].

2.2.3 Sensitivity Analysis of Networks

In this section the focus is on the following question: How to measure the sensitivity of the behavior of the overall network with respect to variations in the network structure? This problem has been investigated in a subproject that has led to the publications [176], [177]. For simplicity we stick to the sensitivity of networks with identical SISO node systems. Although this theory can be extended to heterogenous networks [177]. In the sequel, we consider the input-to-state and the state-to-output interconnections since this leads to an interesting connection to model reduction based on balanced truncation.

Let the N identical SISO node systems be described by the matrices $\alpha \in \mathbb{F}^{n \times n}$, $\beta \in \mathbb{F}^n$ and $\gamma \in \mathbb{F}^{1 \times n}$. Further, the network structure is described the matrices $L \in \mathbb{F}^{N \times N}$ and $B \in \mathbb{F}^{N \times m}$ and $C \in \mathbb{F}^{p \times N}$. Recall that the network state-space representation

$$\begin{aligned} \mathbf{x}(k+1) &= \bar{\mathbf{A}}\mathbf{x}(k) + \bar{\mathbf{B}}\mathbf{u}(k) \\ \mathbf{y}(k) &= \bar{\mathbf{C}}\mathbf{x}(k) \end{aligned} \quad (2.27)$$

is given by

$$\bar{\mathbf{A}} = \mathbf{I}_N \otimes \alpha + \mathbf{L} \otimes \beta \gamma, \quad \bar{\mathbf{B}} = \mathbf{B} \otimes \beta \quad \text{and} \quad \bar{\mathbf{C}} = \mathbf{C} \otimes \gamma.$$

For homogenous networks the node transfer function is given by

$$g(z) = \gamma(z\mathbf{I}_n - \alpha)^{-1}\beta. \quad (2.28)$$

Throughout this section g is supposed to be bounded real and hence, $h(z) = 1/g(z)$ is real rational satisfying $h(\infty) = \infty$ and $|h(z)| \geq 1$ for all $|z| \geq 1$. In addition, it is assumed that the network transfer function

$$\phi(z) = \mathbf{C}(z\mathbf{I}_N - \mathbf{L})^{-1}\mathbf{B}$$

has all its poles in the unit open disc $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. Consequently, the global network transfer function of (2.27) can be written as

$$\Phi(z) = \bar{\mathbf{C}}(z\mathbf{I}_{nN} - \bar{\mathbf{A}})^{-1}\bar{\mathbf{B}} = \mathbf{C}(h(z)\mathbf{I}_N - \mathbf{L})^{-1}\mathbf{B} \quad (2.29)$$

and Φ is strictly proper.

Let $p = m$, then the Jacobians of $\Phi = \phi \circ h$ with respect to the input/output coupling parameters \mathbf{B}, \mathbf{C} are given by the stable transfer functions

$$\begin{aligned} \frac{\partial \Phi}{\partial \mathbf{B}}(z) &= (h(z)\mathbf{I}_N - \mathbf{L}^T)^{-1}\mathbf{C}^T, \\ \frac{\partial \Phi}{\partial \mathbf{C}}(z) &= (h(z)\mathbf{I}_N - \mathbf{L})^{-1}\mathbf{B}. \end{aligned} \quad (2.30)$$

Then we call

$$\mathcal{S}_g(\mathbf{L}, \mathbf{B}, \mathbf{C}) := \frac{1}{2\pi i} \int_{|z|=1} (\| (h(z)\mathbf{I}_N - \mathbf{L})^{-1} \mathbf{B} \|^2 + \| \mathbf{C} (h(z)\mathbf{I}_N - \mathbf{L})^{-1} \|^2) \frac{dz}{z} \quad (2.31)$$

the *network sensitivity* of Φ with respect to \mathbf{B} and \mathbf{C} . It assigns to each realization $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ of ϕ the sum of the L_2 -norms of the Jacobians (2.30). Note that the network sensitivity depends on the realization $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ of the network transfer function. It is thus of interest to find realizations so that the network sensitivity is minimal. A realization $(\mathbf{L}_*, \mathbf{B}_*, \mathbf{C}_*) = (\mathbf{T}\mathbf{L}\mathbf{T}^{-1}, \mathbf{T}\mathbf{B}, \mathbf{C}\mathbf{T}^{-1})$, or the associated state space coordinate transformation \mathbf{T} , is said to have *minimal sensitivity*, if among all realizations $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ of ϕ , it holds that

$$\mathcal{S}_g(\mathbf{L}_*, \mathbf{B}_*, \mathbf{C}_*) \leq \mathcal{S}_g(\mathbf{L}, \mathbf{B}, \mathbf{C}).$$

In order to investigate the existence and the derivation of realizations having minimal sensitivity we define the *network sensitivity Gramians*

$$\mathbf{W}_c^g := \frac{1}{2\pi i} \int_{|z|=1} (h(z)\mathbf{I}_N - \mathbf{L})^{-1} \mathbf{B} \mathbf{B}^T (\overline{h(z)}\mathbf{I}_N - \mathbf{L}^T)^{-1} \frac{dz}{z} \quad (2.32)$$

$$\mathbf{W}_o^g := \frac{1}{2\pi i} \int_{|z|=1} (\overline{h(z)}\mathbf{I}_N - \mathbf{L}^T)^{-1} \mathbf{C}^T \mathbf{C} (h(z)\mathbf{I}_N - \mathbf{L})^{-1} \frac{dz}{z}. \quad (2.33)$$

Using these Gramians the sensitivity of a network can be expressed as the sum of the traces of \mathbf{W}_c^g and \mathbf{W}_o^g , i.e.

$$\mathcal{S}_g(\mathbf{L}, \mathbf{B}, \mathbf{C}) = \text{tr}(\mathbf{W}_c^g + \mathbf{W}_o^g), \quad (2.34)$$

where $\text{tr}(\mathbf{X})$ denotes the trace of the matrix \mathbf{X} .

Lemma 2.1. *Suppose that the node transfer function g is strictly proper with minimal realizations. Let the network transfer function ϕ be stable. Then, the sensitivity Gramians are positive definite if and only if the realization $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ is controllable and observable.*

As we are looking for state space coordinate transformations that minimize sensitivity it is of interest how the sensitivity Gramians change under similarity transformations. Let \mathbf{T} describe a state space transformation $(\mathbf{L}, \mathbf{B}, \mathbf{C}) \mapsto (\mathbf{T}\mathbf{A}\mathbf{T}^{-1}, \mathbf{T}\mathbf{B}, \mathbf{C}\mathbf{T}^{-1})$. It is easy to see that the Gramians transform as

$$(\mathbf{W}_c^g, \mathbf{W}_o^g) \mapsto (\mathbf{T} \mathbf{W}_c^g \mathbf{T}^T, (\mathbf{T}^T)^{-1} \mathbf{W}_o^g \mathbf{T}^{-1}).$$

So, we assign to each state space transformation \mathbf{T} (or $\mathbf{P} := \mathbf{T}^T \mathbf{T}$, respectively) the network sensitivity cost as

$$\begin{aligned} \mathcal{S}_g(\mathbf{T}\mathbf{L}\mathbf{T}^{-1}, \mathbf{T}\mathbf{B}, \mathbf{C}\mathbf{T}^{-1}) &:= \text{tr}(\mathbf{W}_c^g \mathbf{T}^T \mathbf{T} + \mathbf{W}_o^g \mathbf{T}^{-1} (\mathbf{T}^T)^{-1}) \\ &= \text{tr}(\mathbf{W}_c^g \mathbf{P} + \mathbf{W}_o^g \mathbf{P}^{-1}). \end{aligned}$$

To characterize sensitivity optimal realizations we consider the optimization problem for the convex cost function

$$\Phi_g : \mathcal{P} \rightarrow \mathbb{R}, \quad \Phi_g(\mathbf{P}) := \text{tr}(\mathbf{W}_c^g \mathbf{P} + \mathbf{W}_o^g \mathbf{P}^{-1}),$$

where \mathcal{P} denotes the convex space of all real $N \times N$ positive definite matrices \mathbf{P} . The subsequent statement provides a characterization of sensitivity optimal realizations. Let $\mathbf{R}, \mathbf{O}, \mathbf{H}$ denote the reachability, observability and the Hankel matrix of $(\mathbf{L}, \mathbf{B}, \mathbf{C})$, respectively.

Theorem 2.11. [176] *Let $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ be an arbitrary, not necessarily minimal, realization of the stable strictly proper transfer function $\phi(z)$. Equivalent are:*

1. *There exists an invertible coordinate transformation \mathbf{T}_* that minimizes the sensitivity function $\mathbf{T} \mapsto S_g(\mathbf{T}\mathbf{L}\mathbf{T}^{-1}, \mathbf{T}\mathbf{B}, \mathbf{C}\mathbf{T}^{-1})$.*
2. *There exists \mathbf{T}_* such that $(\mathbf{L}_*, \mathbf{B}_*, \mathbf{C}_*) = (\mathbf{T}_* \mathbf{L} \mathbf{T}_*^{-1}, \mathbf{T}_* \mathbf{B}, \mathbf{C} \mathbf{T}_*^{-1})$ has equal Gramians $\mathbf{W}_c^g(\mathbf{L}_*, \mathbf{B}_*) = \mathbf{W}_o^g(\mathbf{C}_*, \mathbf{L}_*)$.*
3. *There exists a unique positive definite matrix $\mathbf{P}_* = \mathbf{T}_*^T \mathbf{T}_*$ that minimizes $\Phi_g(\mathbf{P}) = \text{tr}(\mathbf{W}_c^g \mathbf{P} + \mathbf{W}_o^g \mathbf{P}^{-1})$ on \mathcal{P} .*
4. *There exists $\mathbf{P}_* \in \mathcal{P}$ with $\mathbf{P}_* \mathbf{W}_c^g \mathbf{P}_* = \mathbf{W}_o^g$.*
5. $\text{rk } \mathbf{R} = \text{rk } \mathbf{O} = \text{rk } \mathbf{H}$.

Thus, if $\text{rk } \mathbf{R} = \text{rk } \mathbf{O} = \text{rk } \mathbf{H}$ holds, then a realization $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ of ϕ is sensitivity optimal if and only if $\mathbf{W}_c^g = \mathbf{W}_o^g$. Moreover, if $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ is controllable and observable, then any two sensitivity optimal realizations $(\mathbf{L}_, \mathbf{B}_*, \mathbf{C}_*), (\mathbf{L}_{**}, \mathbf{B}_{**}, \mathbf{C}_{**})$ are similar via a unique orthogonal coordinate transformation $\mathbf{T} \in O_N(\mathbb{R})$.*

In particular, this theorem shows that any network implementation of identical systems described by a bounded real transfer function so that input/output weights of the network are equal ($\mathbf{C}^T = \mathbf{B}$) has minimal sensitivity.

In [176] we propose some numerical results for the computation of sensitivity optimal realizations using a simple Newton algorithm which is fast (locally quadratically convergent) and globally convergent.

For any positive definite symmetric matrices $\mathbf{W}_c^g, \mathbf{W}_o^g$ and any $\mathbf{P} \in \mathcal{P}$ we denote by $\mathbf{Z} = \mathbf{Z}(\mathbf{P})$ the unique positive definite solution of the Lyapunov equation

$$\mathbf{Z}\mathbf{W}_c^g \mathbf{P} + \mathbf{P}\mathbf{W}_c^g \mathbf{Z} = \mathbf{P}\mathbf{W}_o^g \mathbf{P} - \mathbf{W}_o^g. \quad (2.35)$$

Furthermore, we consider the iteration

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \mathbf{Z}(\mathbf{P}_t), \quad t = 0, 1, \dots, \quad (2.36)$$

where $\mathbf{P}_0 \in \mathcal{P}$ is so that $\mathbf{P}_0 \mathbf{W}_c^g \mathbf{P}_0 - \mathbf{W}_o^g > 0$. Note that this is exactly the Riemannian Newton algorithm for optimizing Φ_g , with respect to a suitable

Riemannian metric on \mathcal{P} . Hence, the sequence $(\mathbf{P}_t)_{t \in \mathbb{N}}$ monotone decreasing in the sense that for any $t \in \mathbb{N}$ it holds that $\mathbf{P}_t > \mathbf{P}_{t+1}$. In addition, $(\mathbf{P}_t)_{t \in \mathbb{N}}$ converges locally quadratically to the sensitivity optimal \mathbf{P}_* .

In the sequel we compare the sensitivity Gramians to the classical discrete-time controllability and observability Gramians associated with a realization $(\mathbf{L}, \mathbf{B}, \mathbf{C})$. We denote the classical controllability and observability Gramians by $\mathbf{W}_c(\mathbf{L}, \mathbf{B})$ and $\mathbf{W}_o(\mathbf{C}, \mathbf{L})$, respectively, which are given by the unique solutions of the following Lyapunov equations

$$\mathbf{L}\mathbf{W}_c(\mathbf{L}, \mathbf{B})\mathbf{L}^\top - \mathbf{W}_c(\mathbf{L}, \mathbf{B}) + \mathbf{B}\mathbf{B}^\top = 0, \quad (2.37)$$

$$\mathbf{L}^\top\mathbf{W}_o(\mathbf{C}, \mathbf{L})\mathbf{L} - \mathbf{W}_o(\mathbf{C}, \mathbf{L}) + \mathbf{C}^\top\mathbf{C} = 0. \quad (2.38)$$

For model reduction based on balanced truncation the classical Gramians are of vital interest. To gain insight into sensitivity optimal model reduction and its limitations we show that the sensitivity Gramians satisfy the following Lyapunov-like equations. For brevity we use the notations $\mathbf{X}(z) := (h(z)\mathbf{I}_N - \mathbf{L})^{-1}\mathbf{B}$ and $\mathbf{Y}(z) := \mathbf{C}(h(z)\mathbf{I}_N - \mathbf{L})^{-1}$. Also we denote by $|\lambda_{\max}(\mathbf{A})|$ the spectral radius of the matrix \mathbf{A} .

Lemma 2.2. *Suppose that g is stable and satisfies $\max_{|z|=1} |g(z)| < \frac{1}{|\lambda_{\max}(\mathbf{L})|}$. Then,*

$$\mathbf{L}\mathbf{W}_c^g\mathbf{L}^\top - \mathbf{W}_c^g + \mathbf{B}\mathbf{B}^\top = \frac{1}{2\pi i} \int_{|z|=1} (|h(z)|^2 - 1) \mathbf{X}(z) \mathbf{X}(z)^* \frac{dz}{z},$$

$$\mathbf{L}^\top\mathbf{W}_o^g\mathbf{L} - \mathbf{W}_o^g + \mathbf{C}^\top\mathbf{C} = \frac{1}{2\pi i} \int_{|z|=1} (|h(z)|^2 - 1) \mathbf{Y}(z)^* \mathbf{Y}(z) \frac{dz}{z}.$$

The subsequent statement draws a comparison between the sensitivity Gramians $\mathbf{W}_c^g, \mathbf{W}_o^g$ and the classical Gramians $\mathbf{W}_c(\mathbf{L}, \mathbf{B}), \mathbf{W}_o(\mathbf{C}, \mathbf{L})$, respectively. To this end we write $\mathbf{X} < (\leq) \mathbf{Y}$ if the matrix $\mathbf{Y} - \mathbf{X}$ is positive (semi-)definite. Recall that g is called lossless if it is bounded real and satisfies $|g(z)| = 1$ for all $|z| = 1$.

Theorem 2.12. [176]

(a) *Suppose that g is bounded real with $|g(z)| \neq 1$ for $|z| = 1$. Then,*

$$\mathbf{W}_c^g < \mathbf{W}_c(\mathbf{L}, \mathbf{B}) \quad \text{and} \quad \mathbf{W}_o^g < \mathbf{W}_o(\mathbf{C}, \mathbf{L}).$$

(b) *The transfer function g is lossless if and only if*

$$\mathbf{W}_c^g = \mathbf{W}_c(\mathbf{L}, \mathbf{B}) \quad \text{and} \quad \mathbf{W}_o^g = \mathbf{W}_o(\mathbf{C}, \mathbf{L}).$$

(c) *Suppose that $1 \leq |g(z)| < \frac{1}{|\lambda_{\max}(\mathbf{L})|}$ for all $|z| = 1$. Then,*

$$\mathbf{W}_c^g \geq \mathbf{W}_c(\mathbf{L}, \mathbf{B}) \quad \text{and} \quad \mathbf{W}_o^g \geq \mathbf{W}_o(\mathbf{C}, \mathbf{L}).$$

On the other hand, the relation of the sensitivity Gramians to the classical Gramians associated with the realization $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$ is also of interest. In [268] it is shown that

$$\begin{aligned} (h(z)\mathbf{I}_N - \mathbf{L})^{-1}\mathbf{B} &= (\mathbf{I} \otimes \boldsymbol{\gamma}) (z\mathbf{I}_{n_N} - \bar{\mathbf{A}})^{-1}\bar{\mathbf{B}}, \\ \mathbf{C}(h(z)\mathbf{I}_N - \mathbf{L})^{-1} &= \bar{\mathbf{C}} (z\mathbf{I}_{n_N} - \bar{\mathbf{A}})^{-1}(\mathbf{I} \otimes \boldsymbol{\beta}). \end{aligned}$$

Therefore, the sensitivity Gramians can be expressed via the classical Gramians associated with $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$ by

$$\begin{aligned} \mathbf{W}_c^g &= (\mathbf{I} \otimes \boldsymbol{\gamma}) \mathbf{W}_c(\bar{\mathbf{A}}, \bar{\mathbf{B}}) (\mathbf{I} \otimes \boldsymbol{\gamma}^T), \\ \mathbf{W}_o^g &= (\mathbf{I} \otimes \boldsymbol{\beta}^T) \mathbf{W}_o(\bar{\mathbf{C}}, \bar{\mathbf{A}}) (\mathbf{I} \otimes \boldsymbol{\beta}). \end{aligned}$$

This implies that the sensitivity Gramians can be computed as

$$\mathbf{W}_c^g = (\mathbf{I} \otimes \boldsymbol{\gamma}) \mathbf{X} (\mathbf{I} \otimes \boldsymbol{\gamma}^T)$$

via the linear matrix inequality

$$\bar{\mathbf{A}}\mathbf{X}\bar{\mathbf{A}}^T - \mathbf{X} + \bar{\mathbf{B}}\bar{\mathbf{B}}^T = 0, \quad \mathbf{X} > 0.$$

Moreover, in [211] it is shown that, if the transfer function g is bounded real the classical controllability and observability Gramians for $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$ and $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ are related via

$$\mathbf{W}_c(\bar{\mathbf{A}}, \bar{\mathbf{B}}) \leq \mathbf{W}_c(\mathbf{L}, \mathbf{B}) \otimes \mathbf{P}^{-1} \quad \text{and} \quad \mathbf{W}_o(\bar{\mathbf{C}}, \bar{\mathbf{A}}) \leq \mathbf{W}_o(\mathbf{C}, \mathbf{L}) \otimes \mathbf{P},$$

where \mathbf{P} is the unique positive definite solution to the Riccati equation

$$\boldsymbol{\alpha}\mathbf{P}\boldsymbol{\alpha}^T - \mathbf{P} + \frac{\boldsymbol{\alpha}^T\mathbf{P}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{P}\boldsymbol{\alpha}}{1 - \boldsymbol{\beta}^T\mathbf{P}\boldsymbol{\beta}} + \boldsymbol{\gamma}^T\boldsymbol{\gamma} = 0$$

in terms of the matrices $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ describing the node systems. Since it holds that $\boldsymbol{\beta}^T\mathbf{P}\boldsymbol{\beta} \leq 1$ and $\boldsymbol{\gamma}^T\mathbf{P}^{-1}\boldsymbol{\gamma} \leq 1$ the sensitivity Gramians can be estimated by the classical Gramians corresponding to the network realization $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ and realization $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{C})$ of the node transfer function.

Corollary 2.4. *Suppose that the node transfer function g is bounded real and \mathbf{P} is as above. Then,*

$$\mathbf{W}_c^g \leq \boldsymbol{\gamma}^T\mathbf{P}^{-1}\boldsymbol{\gamma} \cdot \mathbf{W}_c(\mathbf{L}, \mathbf{B}), \quad \mathbf{W}_o^g \leq \boldsymbol{\beta}^T\mathbf{P}\boldsymbol{\beta} \cdot \mathbf{W}_o(\mathbf{C}, \mathbf{L}).$$

At the end of this subsection we examine the issue of balancing with respect to the sensitivity Gramians. A network realization $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ is said to be *g -sensitivity balanced* if the sensitivity Gramians coincide and are diagonal, i.e.

$$\mathbf{W}_c^g = \mathbf{W}_o^g = \boldsymbol{\Sigma}^g = \text{diag}(\sigma_1^g, \dots, \sigma_N^g).$$

The entries $\sigma_1^g \geq \dots \geq \sigma_N^g$ are called the *g -singular values* of the network. By Theorem 2.12 (b) it turns out that if the node transfer function is lossless then sensitivity balancing of the overall network is equivalent to balancing of the network structure.

Theorem 2.13. [176] *Suppose that the node transfer function g is lossless. Then the network is g -sensitivity balanced if and only if $(\mathbf{L}, \mathbf{B}, \mathbf{C})$ is balanced, i.e. $\mathbf{W}_c(\mathbf{L}, \mathbf{B}) = \mathbf{W}_o(\mathbf{C}, \mathbf{L}) = \mathbf{\Sigma}$. In particular, the g -singular values coincide with the singular values of $(\mathbf{L}, \mathbf{B}, \mathbf{C})$.*

2.3 Minimal Bit Rates and Entropy for Control Tasks

2.3.1 Motivation and Earlier Results

A prerequisite for the analysis of the communication demands within networked systems is a thorough analysis of the information needed for performing control tasks. This problem has been considered for more than twenty years, cf. the survey [271], however with limited success. Early landmarks are the paper [92] which considered quantized information for stabilization and proposed to use statistical methods from ergodic theory, and [390] discussing stabilization of linear systems via coding. From the wealth of literature on this topic we also cite [93, 114, 254, 270, 357, 396]. We propose an approach that is different from quantization of the state space. In particular, no information pattern is specified describing how the information on the states of the system is available for the controller.

In non-technical terms, the basic idea for the approach is the following. Consider a control task on the time interval $[0, \infty)$. For example, this may be the problem to make a subset of the state space invariant or the problem to stabilize the system about an equilibrium. Usually, a controller device is constructed which performs the control task based on measurements of the output of the system. If successful, the controller will generate control actions on the system such that the desired behavior is achieved for all initial values in a given set \mathcal{K} in the state space.

If continual measurement of the output is not possible due to data rate constraints (in a noiseless communication channel), the controller only has a finite amount of information available on any finite interval $[0, \tau]$. Hence, it may appear reasonable that the controller can only generate a finite number of time-dependent control functions $\mathbf{u}(t)$, $t \in [0, \tau]$, which are to guarantee the desired behavior on $[0, \tau]$ for all initial states in \mathcal{K} . If time increases, the amount of information for the controller increases, and hence it may generate more controls. Looking at this from the other side, the number of controls, which are necessary for accomplishing the control task on $[0, \tau]$, determines the minimal data rate. Thus, the growth rate of the minimal number of controls as time tends to infinity is a measure for the minimal data rate necessary to accomplish the control task on $[0, \infty)$ for all initial values in \mathcal{K} .

More specifically, we see that the combination of the aspects of control and information, which traditionally are separated, requires a modification of the familiar feedback paradigm. If the communication channel between the

system and the controller only admits a finite bit rate, then it is not justified that the controller generates infinitely many different control functions on a finite time interval, since only a finite amount of information is available. This excludes state-dependent controls $\mathbf{u}(\mathbf{x}(t, \mathbf{x}_0))$, where depending on the initial state \mathbf{x}_0 one obtains infinitely many control functions. Instead, the controller might use information on the system to compute open-loop controls on some time interval $[t_i, t_{i+1}]$; then, using updated information, a new control function might be computed and used on the next time interval $[t_{i+1}, t_{i+2}]$ (the t_i might depend on the initial value \mathbf{x}_0 .) The computations may be based on quantization of the state space (cf. Section 5.4), symbolic controllers or may be done via devices like model predictive control (MPC) (cf. Chapter 4). In any case, this results in time-dependent control functions $\mathbf{u}(t)$ defined on $[0, \infty)$.

Instead of concentrating on the algorithmic question how to generate these controls, we discuss the minimal bit rate needed to discern the time-dependent control functions $\mathbf{u}(t)$ on any time interval $[0, \tau]$, $\tau > 0$. This, in fact, is the point of view taken in [357], p. 1057, that estimates the minimal bit rate for stabilization of discrete-time linear systems from below (see Proposition 3.2 in [357]).

This basic idea can be formalized in different ways. We will discuss in some detail invariance entropy for the problem to keep the system within a subset of the state space, and an entropy for the stabilization problem.

In this problem formulation, a close analogy to the notion of topological entropy in the theory of dynamic systems becomes apparent; the monographs [110, 197, 376] contain expositions of this theory. Here one counts, how fast trajectories of a dynamic system move apart, and hence one has to look at initial points. In control, the decisive parameter which determines the behavior of trajectories is the control function. Hence, we will count control functions and then we use rather analogous mathematical constructions.

In Section 2.3.2 the problem to render a compact subset of the state space invariant will be discussed in detail; cf. [66, 70, 71, 73, 198–201]. Section 2.3.3 presents results on exponential stabilization.

Notation. The closure of a set \mathcal{A} is $\text{cl } \mathcal{A}$ and the cardinality of a set \mathcal{A} is $\#\mathcal{A}$; thus if \mathcal{A} is a finite set, $\#\mathcal{A}$ is the number of its elements and otherwise $\#\mathcal{A} = \infty$.

2.3.2 Invariance Entropy

Overview of New Results. The concept of *invariance entropy* has its origins in the problem to determine data rates needed in order to achieve desired control objectives. NAIR, EVANS, MAREELS and MORAN introduced ideas from topological dynamics into this field. They defined and studied *topological feedback entropy* for the problem of stabilizing a discrete-time system

into a compact subset of the state space and *local topological feedback entropy* for stabilization at an equilibrium. We modify this approach by introducing the notion of *invariance entropy* for continuous-time systems which measures how fast the number of open-loop control functions grows which are needed to keep the system inside a compact set for longer and longer times. More precisely, for a compact and controlled invariant subset \mathcal{Q} of the state space, and a compact set $\mathcal{K} \subset \mathcal{Q}$ of allowed initial states, the minimal number of open-loop control functions necessary to keep the system in \mathcal{Q} up to time $\tau > 0$, from an arbitrary initial state in \mathcal{K} , is denoted by $r_{\text{inv}}^*(\tau, \mathcal{K}, \mathcal{Q})$. Then the exponential growth rate

$$h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q}) := \limsup_{\tau \rightarrow \infty} \frac{1}{\tau} \ln r_{\text{inv}}^*(\tau, \mathcal{K}, \mathcal{Q})$$

is called the *strict invariance entropy of $(\mathcal{K}, \mathcal{Q})$* . Requiring that the solutions only stay in an ε -neighborhood of \mathcal{Q} , another quantity $h_{\text{inv}}(\mathcal{K}, \mathcal{Q})$, simply called the *invariance entropy of $(\mathcal{K}, \mathcal{Q})$* , is defined similarly, now also taking the limit for ε going to zero. We started a thorough investigation of these notions with an emphasis on the derivation of upper and lower bounds.

In this section, we give an overview of the main results obtained so far. After introducing the basic notions and explaining the elementary properties of invariance entropy, we show that the strict invariance entropy $h_{\text{inv}}^*(\mathcal{Q})$ can be characterized as the minimal data rate for the control objective of rendering the set \mathcal{Q} invariant by a coder-controller device. Also relations between invariance entropy and Lyapunov exponents are addressed. In particular, the following two results are highlighted:

For linear systems, the invariance entropy equals the sum of the unstable eigenvalues counted according to their multiplicities. For nonlinear systems, under additional controllability assumptions on the set \mathcal{Q} , the strict invariance entropy $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$ is upper bounded by the sum of the unstable Lyapunov exponents of any regular periodic solution in the interior of \mathcal{Q} .

The latter result can be used to obtain a formula for the invariance entropy of a control set for a scalar control-affine system with a single control vector field. An application to a linearized inverted pendulum is given.

Definition of Invariance Entropy. We consider a continuous-time control system

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \quad \mathbf{u} \in \mathcal{U} \quad (2.39)$$

on a smooth manifold \mathcal{M} , endowed with a metric d (not necessarily a Riemannian distance). The set of admissible control functions is given by

$$\mathcal{U} = \{\mathbf{u} : \mathbb{R} \rightarrow \mathbb{R}^m : \mathbf{u} \text{ measurable with } \mathbf{u}(t) \in \mathcal{U} \text{ a.e.}\}$$

with a compact control range $\mathcal{U} \subset \mathbb{R}^m$, and the right-hand side $\mathbf{f} : \mathcal{M} \times \mathbb{R}^m \rightarrow T\mathcal{M}$ is a continuous function which is continuously differentiable in

the first argument. Then, for each initial value $\mathbf{x} \in \mathcal{M}$ and for each $\mathbf{u} \in \underline{\mathcal{U}}$, there exists a unique solution (in the sense of Carathéodory), which we denote by $\varphi(t, \mathbf{x}, \mathbf{u})$. Let $\mathcal{Q} \subset \mathcal{M}$ be a compact and *controlled invariant* set, that is, for every $\mathbf{x} \in \mathcal{Q}$ there is $\mathbf{u} \in \underline{\mathcal{U}}$ with $\varphi(t, \mathbf{x}, \mathbf{u}) \in \mathcal{Q}$ for all $t \geq 0$. Furthermore, let $\mathcal{K} \subset \mathcal{Q}$ be another compact set.

Definition 2.3. For $\tau > 0$, a set $\mathcal{S} \subset \underline{\mathcal{U}}$ is called $(\tau, \mathcal{K}, \mathcal{Q})$ -spanning if for every $\mathbf{x} \in \mathcal{K}$ there is $\mathbf{u} \in \mathcal{S}$ with

$$\varphi(t, \mathbf{x}, \mathbf{u}) \in \mathcal{Q} \text{ for all } t \in [0, \tau].$$

Let $r_{\text{inv}}^*(\tau, \mathcal{K}, \mathcal{Q})$ be the minimal cardinality of such a set. The strict invariance entropy of $(\mathcal{K}, \mathcal{Q})$ is defined as the exponential growth rate of $r_{\text{inv}}^*(\tau, \mathcal{K}, \mathcal{Q})$ for $\tau \rightarrow \infty$, that is,

$$h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q}) := \limsup_{\tau \rightarrow \infty} \frac{1}{\tau} \ln r_{\text{inv}}^*(\tau, \mathcal{K}, \mathcal{Q}).$$

If $\mathcal{K} = \mathcal{Q}$, we simply write $r_{\text{inv}}^*(\tau, \mathcal{Q})$ and $h_{\text{inv}}^*(\mathcal{Q})$.

One problem with this definition is that the numbers $r_{\text{inv}}^*(\tau, \mathcal{K}, \mathcal{Q})$ need not be finite. To overcome this problem, we also work with another version of invariance entropy.

Definition 2.4. For $\tau, \varepsilon > 0$, a set $\mathcal{S} \subset \underline{\mathcal{U}}$ is called $(\tau, \varepsilon, \mathcal{K}, \mathcal{Q})$ -spanning if for every $\mathbf{x} \in \mathcal{K}$ there is $\mathbf{u} \in \mathcal{S}$ with

$$\text{dist}(\varphi(t, \mathbf{x}, \mathbf{u}), \mathcal{Q}) < \varepsilon \text{ for all } t \in [0, \tau].$$

Let $r_{\text{inv}}(\tau, \varepsilon, \mathcal{K}, \mathcal{Q})$ be the minimal cardinality of such a set. The invariance entropy of $(\mathcal{K}, \mathcal{Q})$ is defined as

$$h_{\text{inv}}(\mathcal{K}, \mathcal{Q}) := \lim_{\varepsilon \searrow 0} \limsup_{\tau \rightarrow \infty} \frac{1}{\tau} \ln r_{\text{inv}}(\tau, \varepsilon, \mathcal{K}, \mathcal{Q}).$$

If $\mathcal{K} = \mathcal{Q}$, we simply write $r_{\text{inv}}(\tau, \varepsilon, \mathcal{Q})$ and $h_{\text{inv}}(\mathcal{Q})$.

In the preceding definition, we always have $r_{\text{inv}}(\tau, \varepsilon, \mathcal{K}, \mathcal{Q}) < \infty$, which easily follows from compactness of \mathcal{K} and continuous dependence on the initial value. Existence of the limit in the definition of $h_{\text{inv}}(\mathcal{K}, \mathcal{Q})$ follows from monotonicity of $r_{\text{inv}}(\tau, \varepsilon, \mathcal{K}, \mathcal{Q})$ in ε . Some elementary properties of $h_{\text{inv}}(\mathcal{K}, \mathcal{Q})$ and $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$ are the following.

1. $0 \leq h_{\text{inv}}(\mathcal{K}, \mathcal{Q}) \leq h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$ and $h_{\text{inv}}(\mathcal{K}, \mathcal{Q}) < \infty$.
2. $h_{\text{inv}}(\mathcal{K}, \mathcal{Q})$ is independent of the choice of the metric on \mathcal{M} .
3. $h_{\text{inv}}(\mathcal{K}, \mathcal{Q})$ and $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$ are invariant under C^0 -state equivalence.
4. If $\mathcal{K} = \mathcal{K}_1 \cup \dots \cup \mathcal{K}_n$ (\mathcal{K}_i compact), then

$$h_{\text{inv}}^{(*)}(\mathcal{K}, \mathcal{Q}) = \max_{i=1}^n h_{\text{inv}}^{(*)}(\mathcal{K}_i, \mathcal{Q}).$$

5. For the system with right-hand side $s \cdot F$ ($s > 0$) instead of F , the set \mathcal{Q} is also controlled invariant and

$$h_{\text{inv}}^{(*)}(\mathcal{K}, \mathcal{Q}; s \cdot F) = s \cdot h_{\text{inv}}^{(*)}(\mathcal{K}, \mathcal{Q}; F).$$

In general, the equality $h_{\text{inv}}(\mathcal{K}, \mathcal{Q}) = h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$ does not hold, since there are cases when $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q}) = \infty$. However, in some special situations equality can be shown.

Minimum Transmission Data Rate. We relate the strict invariance entropy $h_{\text{inv}}^*(\mathcal{Q})$ to minimal transmission data rates in feedback loops with a digital channel between coder and controller. Consider system (2.39) and suppose that a sensor, which is connected to a controller via a digital noiseless channel, measures its states at sampling times $k\tau$, $k \geq 0$, for some fixed time step $\tau > 0$. The state at time $k\tau$ is coded using a finite alphabet \mathcal{S}_k of (time-varying) size μ_k (Fig. 2.1). The coder transmits a symbol $s_k \in \mathcal{S}_k$, which may depend on the present state and on all past states. The corresponding coder mapping is denoted by $\gamma_k : \mathcal{M}^{k+1} \rightarrow \mathcal{S}_k$. At time $k\tau$ the controller has $k + 1$ symbols s_0, s_1, \dots, s_k available and generates a finite-time control function $\mathbf{u}(k) : [0, \tau] \rightarrow \mathcal{U}$. We denote the corresponding controller mapping by $\delta_k : \mathcal{S}_0 \times \mathcal{S}_1 \times \dots \times \mathcal{S}_k \rightarrow \underline{\mathcal{U}}_\tau := \{\mathbf{u}_{|[0, \tau]} : \mathbf{u} \in \underline{\mathcal{U}}\}$.

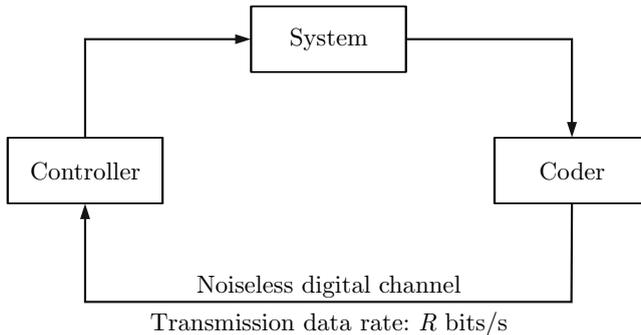


Fig. 2.1 Feedback system with coder and controller

We call the quadruple $\mathcal{H} := (\mathcal{S}, \gamma, \delta, \tau)$, where $\mathcal{S} = (\mathcal{S}_k)_{k \geq 0}$, $\gamma = (\gamma_k)_{k \geq 0}$, and $\delta = (\delta_k)_{k \geq 0}$, a *coder-controller*, and we define the (asymptotic, average) *transmission data rate* of \mathcal{H} by

$$R(\mathcal{H}) := \liminf_{k \rightarrow \infty} \frac{1}{k\tau} \sum_{j=0}^{k-1} \log_2 \mu_j.$$

Additionally, we require that this number is finite. (This definition is essentially taken from [270].) We say that \mathcal{H} renders \mathcal{Q} invariant if for all $\mathbf{x}(0) \in \mathcal{Q}$ the sequence $(\mathbf{x}(k))_{k \geq 0}$, defined recursively by

$$\mathbf{x}(k) := \varphi(\tau, \mathbf{x}(k-1), \mathbf{u}(k-1))$$

with

$$\mathbf{u}(k-1) = \delta_{k-1}(\gamma_0(\mathbf{x}(0)), \gamma_1(\mathbf{x}(0), \mathbf{x}(1)), \dots, \gamma_{k-1}(\mathbf{x}(0), \dots, \mathbf{x}(k-1)))$$

satisfies $\varphi(t, \mathbf{x}(k), \mathbf{u}(k)) \in \mathcal{Q}$ for all $k \geq 0$ and $t \in [0, \tau]$. Then we have the following result, cf. [198, 202].

Theorem 2.14. *Let \mathcal{Q} be a compact and controlled invariant set for system (2.39). Then*

$$h_{\text{inv}}^*(\mathcal{Q}) = \inf_{\mathcal{H}} \frac{R(\mathcal{H})}{\log_2 e}, \quad (2.40)$$

where the infimum is taken over all coder-controllers \mathcal{H} that render \mathcal{Q} invariant.

To prove this theorem, one first shows that every (finite) (τ, \mathcal{Q}) -spanning set $\{u_1, \dots, u_n\}$ defines coding regions $\mathcal{C}_1, \dots, \mathcal{C}_n$ with $\mathcal{Q} = \bigcup_{i=1}^n \mathcal{C}_i$, such that a coder-controller \mathcal{H} can be defined as follows: The coder sends the symbol i to the controller if the measured state is in \mathcal{C}_i . Upon receiving the symbol i , the controller generates the control function \mathbf{u}_i . Then it is easy to see that the corresponding transmission data rate is $\log_2(n)/\tau$. This proves the existence of a sequence \mathcal{H}_k of coder-controllers whose transmission data rates come arbitrarily close to $h_{\text{inv}}^*(\mathcal{Q})$. On the other hand, every coder-controller $\mathcal{H} = (\mathcal{S}, \gamma, \delta, \tau)$ gives $(k\tau, \mathcal{Q})$ -spanning sets for all $k \geq 1$ by concatenation of the control functions generated by the controller at times $i\tau$, $(i = 0, 1, \dots, k)$. The cardinalities of these sets are $\prod_{i=0}^{k-1} \mu_i$, which implies the inequality “ \leq ” in (2.40).

Relation between Invariance Entropy and Lyapunov Exponents.

Next, we present two results which show that the invariance entropy is related to the Lyapunov exponents of the given control system. For linear systems, where the Lyapunov exponents coincide with the real parts of the eigenvalues of the dynamic matrix of the system, we have the following result.

Theorem 2.15. *For the linear control system*

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad \mathbf{u} \in \underline{\mathcal{U}},$$

with $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times m}$ it holds that

$$h_{\text{inv}}(\mathcal{K}, \mathcal{Q}) = \sum_{\lambda \in \text{spec}(A)} \max\{0, n_\lambda \text{Real}(\lambda)\}, \quad (2.41)$$

if \mathcal{K} has positive Lebesgue measure. Here, n_λ denotes the algebraic multiplicity of the eigenvalue λ .

One may wonder why the value of the invariance entropy does only depend on the matrix \mathbf{A} , but not on \mathbf{B} . As one sees in the proof of Theorem 2.15 in [70], this is basically a consequence of the specific structure of the solutions of a linear system. In particular, the distance of two solutions $\varphi(t, \mathbf{x}, \mathbf{u})$ and $\varphi(t, \mathbf{y}, \mathbf{u})$ is independent of the matrix \mathbf{B} , and the same is true for the volume growth factor under application of $\varphi(t, \cdot, \mathbf{u})$. Both of these quantities are strongly related to the numbers $r_{\text{inv}}(\tau, \varepsilon, \mathcal{K}, \mathcal{Q})$. However, we should remark that the assumptions of Theorem 2.15 cannot be satisfied for arbitrary pairs of matrices (\mathbf{A}, \mathbf{B}) . The existence of a compact controlled invariant set \mathcal{Q} with nonempty interior (and hence positive Lebesgue measure) can be guaranteed if the pair (\mathbf{A}, \mathbf{B}) is controllable, the matrix \mathbf{A} is hyperbolic, and the control range \mathcal{U} is compact and convex with nonempty interior. Then there exists a unique control set \mathcal{D} whose closure $\mathcal{Q} = \text{cl } \mathcal{D}$ has the desired properties, cf. [70].

Remark 2.1. For inhomogeneous bilinear systems on \mathbb{R}^n a similar result holds in form of an estimate from below. Here, the real parts of the eigenvalues are replaced by minimal Lyapunov exponents of the associated homogeneous system on invariant subbundles of $\underline{\mathcal{U}} \times \mathbb{R}^n$, cf. [198, 202].

For nonlinear systems we have a finite set of Lyapunov exponents for each solution, that is, for each pair $(\mathbf{x}, \mathbf{u}) \in \mathcal{M} \times \underline{\mathcal{U}}$. Hence, we cannot expect a formula for the invariance entropy as simple as the one in Theorem 2.15. In fact, it seems improbable that a general formula can be given here. However, under additional controllability assumptions on the system, an upper bound, similar to the expression in (2.41), can be derived.

Recall that a *control set* $\mathcal{D} \subset \mathcal{M}$ for system (2.39) is a maximal set of controlled invariance and *approximate controllability* (for each two states $\mathbf{x}, \mathbf{y} \in \mathcal{D}$ and $\varepsilon > 0$ there are $\mathbf{u} \in \underline{\mathcal{U}}$ and $t \geq 0$ with $d(\varphi(t, \mathbf{x}, \mathbf{u}), \mathbf{y}) < \varepsilon$.) For \mathcal{Q} being the closure of a control set, the following result holds, cf. [198, 199, 202].

Theorem 2.16. *Assume that the right-hand side \mathbf{f} of system (2.39) is continuously differentiable, and let \mathcal{D} be a control set with compact closure $\mathcal{Q} = \text{cl } \mathcal{D}$ for this system. Let $(\varphi(\cdot, \mathbf{x}_0, \mathbf{u}_0), \mathbf{u}_0(\cdot))$ be a controlled periodic solution in $\text{int } \mathcal{D} \times \text{int } \mathcal{U}$ which is regular, that is, the linearization along this solution is controllable. Then for each compact set $\mathcal{K} \subset \mathcal{D}$ it holds that*

$$h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q}) \leq \sum_{\lambda} \max\{0, n_{\lambda} \lambda\},$$

where the sum is taken over the different Lyapunov exponents of $\varphi(\cdot, \mathbf{x}_0, \mathbf{u}_0)$ and n_{λ} denotes the multiplicity of λ .

Remark 2.2. Assuming local accessibility for system (2.39) on the interior of a control set \mathcal{D} , there is a periodic solution in $\text{int } \mathcal{D}$ with initial value \mathbf{x} for every $\mathbf{x} \in \text{int } \mathcal{D}$ (cf. Proposition 4.3.3 in [72]). In general, it is not clear if such periodic solutions are regular. However, for real-analytic systems, existence of regular periodic solutions through every point $\mathbf{x} \in \text{int } \mathcal{D}$ can be guaranteed if the system is strongly accessible, which follows from results of [344, 345].

For scalar control-affine systems of the form

$$\dot{x}(t) = f(x(t)) + u(t)g(x(t)), \quad u \in \underline{\mathcal{U}}, \quad (2.42)$$

with C^1 -vector fields $f, g : \mathbb{R} \rightarrow \mathbb{R}$ the estimate of Theorem 2.16 can be used together with other results to prove the following result, cf. [198, 199, 202].

Theorem 2.17. *Let \mathcal{D} be a bounded control set of system (2.42). Further assume that local accessibility holds on $\mathcal{Q} := \text{cl } \mathcal{D}$. Then for every compact set $\mathcal{K} \subset \mathcal{D}$ with nonempty interior it holds that*

$$h_{\text{inv}}(\mathcal{K}, \mathcal{Q}) = h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q}) = \max \left\{ 0, \min_{x \in \mathcal{Q}} \left(f'(x) - \frac{f(x)}{g(x)} g'(x) \right) \right\}. \quad (2.43)$$

The expression to be minimized in (2.43) is the derivative of the right-hand side of (2.42) at the equilibrium pair (x, u_x) with $u_x = -f(x)/g(x)$, which coincides with the Lyapunov exponent for this pair. Since in dimension one (under appropriate regularity assumptions), the full Lyapunov spectrum is already attained by the Lyapunov exponents of equilibria, Eq. (2.43) can also be written as

$$h_{\text{inv}}(\mathcal{K}, \mathcal{Q}) = h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q}) = \max \{ 0, \min \Sigma_{\text{Ly}}(\mathcal{Q}) \},$$

where $\Sigma_{\text{Ly}}(\mathcal{Q})$ denotes the Lyapunov spectrum over \mathcal{Q} , that is, the set of all limits

$$\limsup_{\tau \rightarrow \infty} \frac{1}{\tau} \ln \left| \frac{\partial \varphi}{\partial x}(\tau, x, u) \right|,$$

such that $\varphi(t, x, u) \in \mathcal{Q}$ for all $t \geq 0$.

Example 2.1 The inverted pendulum

Theorem 2.17 can be applied to a controlled linearized inverted pendulum. The pendulum is represented by

$$\ddot{y}(t) + 2b\dot{y}(t) - (1 + u(t))y(t) = 0, \quad u \in \underline{\mathcal{U}}$$

with $b > 0$ and control range $\mathcal{U} = [-\rho, \rho]$, where $0 < \rho < b^2 + 1$. This equation describes the linearization of a controlled damped mathematical pendulum at the unstable position. The corresponding state-space representation is:

$$\dot{\mathbf{x}}(t) = \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & -2b \end{pmatrix}}_{=: \mathbf{A}_0} \mathbf{x}(t) + u(t) \underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}}_{=: \mathbf{A}_1} \mathbf{x}(t), \quad u \in \underline{\mathcal{U}}.$$

The eigenvalues of the matrix \mathbf{A}_0 are given by

$$\lambda_{\pm} = -b \pm \sqrt{b^2 + 1}.$$

Since $b > 0$, λ_- is negative and λ_+ is positive. Hence, the uncontrolled system has one stable and one unstable direction. We can project the bilinear system to the unit circle \mathcal{S}^1 . An easy computation yields the equations for the projected system:

$$\dot{\varphi} = (-2b \sin(\varphi) \cos(\varphi) - \sin^2(\varphi) + \cos^2(\varphi)) + u(t) \cos^2(\varphi), \quad u \in \underline{\mathcal{U}}.$$

From Theorem 8.1.1 in [72] it follows that the control sets on \mathcal{S}^1 consist of equilibria. Hence, in order to determine these sets, we have to find the zeros of the right-hand side. Therefore, we divide by $\cos^2(\varphi)$ (possible for $2\varphi \notin \{\pi, 3\pi\}$) to get

$$\tan^2(\varphi) + 2b \tan(\varphi) - (1 + u) = 0 \quad \Leftrightarrow \quad \tan(\varphi) = -b \pm \sqrt{b^2 + 1 + u}.$$

Hence, we obtain the solutions

$$\varphi_{\pm} = \arctan\left(-b \pm \sqrt{b^2 + 1 + u}\right) \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

and two other solutions in $(\pi/2, (3\pi)/2)$. The solutions φ_{\pm} are real numbers, since

$$b^2 + 1 + u \in [b^2 + 1 - \rho, b^2 + 1 + \rho] \subset (0, 2(b^2 + 1)).$$

Hence, in $(-\pi/2, \pi/2)$ we obtain the following two intervals of equilibria (cf. Fig. 2.2), which are the closures of control sets:

$$\begin{aligned} \mathcal{Q}_- &= \left[\arctan\left(-b - \sqrt{b^2 + 1 + \rho}\right), \arctan\left(-b - \sqrt{b^2 + 1 - \rho}\right) \right], \\ \mathcal{Q}_+ &= \left[\arctan\left(-b + \sqrt{b^2 + 1 - \rho}\right), \arctan\left(-b + \sqrt{b^2 + 1 + \rho}\right) \right]. \end{aligned}$$

Using the formula of Theorem 2.17, we find

$$h_{\text{inv}}(\mathcal{K}, \mathcal{Q}_{\pm}) = \max \left\{ 0, \min_{\varphi \in \mathcal{Q}_{\pm}} (-2b - 2 \tan(\varphi)) \right\}.$$

Hence, we obtain

$$\begin{aligned} h_{\text{inv}}(\mathcal{K}, \mathcal{Q}_-) &= \max \left\{ 0, \min_{u \in [-\rho, \rho]} \left(2\sqrt{b^2 + 1 - u} \right) \right\} = 2\sqrt{b^2 + 1 - \rho}, \\ h_{\text{inv}}(\mathcal{K}, \mathcal{Q}_+) &= 0. \end{aligned}$$

We can interpret this result as follows: The control set $\mathcal{D} = \text{int } \mathcal{Q}_-$ contains $\varphi_0 := \arctan(-b - \sqrt{b^2 + 1})$, which is an equilibrium for the control $u = 0$, that is, the vector $(\cos(\varphi_0), \sin(\varphi_0))^T$ is an eigenvector of the matrix \mathbf{A}_0 corresponding to the stable eigenvalue $\lambda_- = -b - \sqrt{b^2 + 1}$. On \mathcal{D} , the projected system is controllable. This implies that the cone $\pi^{-1}(\mathcal{D}) \subset \mathbb{R}^2$ over \mathcal{D} is the maximal subset of \mathbb{R}^2 where it is possible to steer to the stable axis (that is, to the one-dimensional eigenspace corresponding to λ_-) with the bilinear system. Thus, we have computed the invariance entropy of the maximal subset of \mathbb{R}^2 , where the system can be stabilized to the equilibrium $(0, 0)$. The control set \mathcal{Q}_+ is easily seen to be invariant and hence its invariance entropy is zero. \square

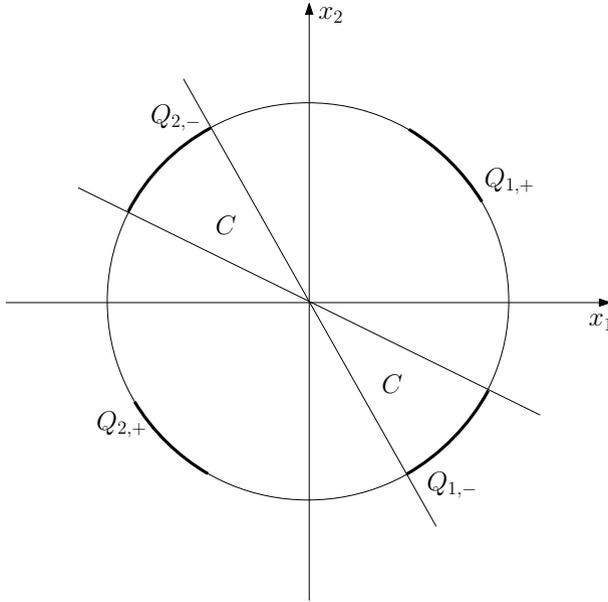


Fig. 2.2 Control sets on the unit circle

Open Problems and Extensions. The most important open problem in the theory of invariance entropy concerns the relation between the two quantities $h_{\text{inv}}(\mathcal{K}, \mathcal{Q})$ and $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$. We always have $h_{\text{inv}}(\mathcal{K}, \mathcal{Q}) \leq h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$ and we know that $h_{\text{inv}}(\mathcal{K}, \mathcal{Q}) < \infty$, while $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q}) = \infty$ is possible. In the one-dimensional control-affine case, Theorem 2.16 shows that under appropriate controllability assumptions, the two quantities coincide. An analogous result holds for linear control systems in arbitrary dimensions, which is shown in [198]. In general, however, this problem remains completely open.

There are other open problems which are somewhat similar. For example, if \mathcal{Q} is the closure of a control set \mathcal{D} , we can only give an upper bound for $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$ when $\mathcal{K} \subset \mathcal{D}$, but not for $h_{\text{inv}}^*(\mathcal{Q})$. While $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$ in this case is independent of \mathcal{K} as long as \mathcal{K} has nonempty interior and $\mathcal{K} \subset \mathcal{D}$, it is not clear whether $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q}) = h_{\text{inv}}^*(\mathcal{Q})$. This question is of importance, since $h_{\text{inv}}^*(\mathcal{Q})$, as explained above, is a measure for the minimal data rate for rendering \mathcal{Q} invariant, while we cannot show something similar for $h_{\text{inv}}^*(\mathcal{K}, \mathcal{Q})$ with $\mathcal{K} \neq \mathcal{Q}$.

The upper estimate of Theorem 2.16 can be improved by imposing stronger regularity assumptions on the given control system. Indeed, assuming real analyticity and strong accessibility plus some weak hyperbolicity condition, one can show that the assumptions of regularity and periodicity of the controlled solution can be omitted. This result can be found in the forthcoming monograph [202].

The results for invariance entropy presented above are restricted to compact controlled invariant sets \mathcal{Q} for finite-dimensional systems. They can be extended in several directions. In [68, 69] the set \mathcal{Q} is replaced by a controlled invariant subspace and techniques from geometric control are used to derive estimates for a corresponding notion of invariance entropy. In [183], invariance entropy for certain linear infinite-dimensional systems (including parabolic partial differential equations and delay equations) is characterized.

2.3.3 Entropy for Exponential Stabilization

Section Overview. Using an approach which is similar to invariance entropy, we consider the time-dependent controls which have to be generated by a controller in order to achieve exponential stability, cf. [67]. Again, we consider continuous-time systems.

There are two ways to measure the information needed for exponential stabilization for a given set \mathcal{Q} of initial states: Either one can look at sets of control functions on $[0, \infty)$ which admit exponential stabilization and at the required bits to discern them on any finite interval $[0, \tau]$ and then analyze the bit rate for $\tau \rightarrow \infty$; finally, one can minimize the bit rate over such sets of controls. Alternatively, one may look at sets of control functions defined on a bounded interval $[0, \tau]$ admitting exponential decay on $[0, \tau]$ and then look at the minimal number of bits for such a set of control functions. Then, again, one can analyze what happens for $\tau \rightarrow \infty$. In the first case we speak about minimal bit rates and in the second case, we speak about entropy, since this approach is close to entropy in the theory of dynamic systems.

A discouraging example (Example 2.2) shows that finitely many controls are not sufficient for an exponential estimate on a finite interval. Hence, we impose somewhat weaker conditions. Then the stabilization entropy can be shown to be finite. In the linear case, a formula in terms of eigenvalues holds. It is worth to emphasize that this formula pertains to a fixed exponential decay rate (Theorem 2.19). Thus, it also applies to cases, where the control goal is to increase the exponential decay rate for a stable system, a situation where finite communication channels might more easily be tolerated. On the other hand, the stabilization entropy provides an upper bound for the minimal bit rate. In particular, the minimal bit rate is finite (Theorem 2.18), and in the linear case, an explicit estimate is available.

Stabilization with Limited Bit Rates. Consider a nonlinear control system of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \quad \mathbf{u} \in \mathcal{U}, \quad (2.44)$$

where $\mathbf{f} : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$ is continuous and Lipschitz continuous with respect to the first argument. The admissible controls are given by

$$\underline{\mathcal{U}} = \{ \mathbf{u} : [0, \infty) \rightarrow \mathbb{R}^m : \mathbf{u} \text{ measurable with } \mathbf{u}(t) \in \mathcal{U} \text{ a.e.} \},$$

where we assume that the controls \mathbf{u} are integrable on every bounded interval and the control range \mathcal{U} is a subset of \mathbb{R}^m . We assume that (i) unique global solutions $\varphi(t, \mathbf{x}_0, \mathbf{u})$, ($t \geq 0$) of the differential equation with initial condition $\mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}^d$ and control $\mathbf{u} \in \underline{\mathcal{U}}$ exist and (ii) on compact intervals, the solutions depend continuously on the initial value.

We consider the bit rates for the problem to stabilize an equilibrium $\mathbf{x}^* \in \mathbb{R}^d$ corresponding to an admissible control value $\mathbf{u}^* \in \mathcal{U}$, that is, $\mathbf{0} = \mathbf{f}(\mathbf{x}^*, \mathbf{u}^*)$. Without loss of generality, the equilibrium is given by $\mathbf{x}^* = \mathbf{0}$ corresponding to the control value $\mathbf{u}^* = \mathbf{0} \in \mathcal{U}$. Suppose that the system is exponentially controllable to the equilibrium for all \mathbf{x}_0 in a neighborhood of $\mathbf{x}^* = \mathbf{0}$, that is, there are constants $M, \alpha > 0$ such that for all $\mathbf{x}_0 \neq \mathbf{0}$ there is $\mathbf{u} \in \underline{\mathcal{U}}$ with

$$\|\varphi(t, \mathbf{x}_0, \mathbf{u})\| < M e^{-\alpha t} \|\mathbf{x}_0\| \text{ for all } t \geq 0. \quad (2.45)$$

This holds, for example, if the linearized system is stabilizable by a linear feedback; (see, for instance, [346] (Section 5.8) for asymptotic stability and [151] (Satz 9.8) for exponential stability). For linear control systems, an estimate of the form (2.45) holds iff the unstable part is controllable.

Since we are interested in problems where the controller does not have continual access to the present state $\mathbf{x}(t)$ (or to observed values $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t))$), we discuss the minimal bit rate needed to discern the required time-dependent control functions \mathbf{u} on any time interval $[0, \tau]$, $\tau > 0$.

The following example (the simplest controllable system) shows that a direct approach is not possible. Here, on a finite time interval $[0, \tau]$ one cannot find finitely many controls such that for every point in a neighborhood of the origin the exponential estimate (2.45) is satisfied.

Example 2.2 *A discouraging example*

Consider the following scalar system:

$$\dot{x}(t) = u(t), \quad u(t) \in \mathcal{U} = \mathbb{R}. \quad (2.46)$$

Let $\alpha > 0$, $M > 1$ and fix $\tau > 0$. We claim that for τ large enough there is no finite set \mathcal{S} of control functions on $[0, \tau]$, such that for every $0 \neq x_0 \in \mathcal{K} := [-1, 1]$ there is $u \in \mathcal{S}$ such that the exponential estimate

$$|\varphi(t, x_0, u)| = \left| x_0 + \int_0^t u(s) ds \right| < M e^{-\alpha t} |x_0| \text{ for all } t \in [0, \tau] \quad (2.47)$$

holds. We proceed by contradiction: Suppose that a finite set $\mathcal{S} = \{u_1, \dots, u_n\}$ with the stated properties exists and define

$$\mathcal{K}_j := \{x_0 \in \mathcal{K} : |\varphi(t, x_0, u_j)| < M e^{-\alpha t} |x_0| \text{ for all } t \in [0, \tau]\}.$$

Observe that $0 \notin \mathcal{K}_j$. For the control $u_0(t) \equiv 0$, every point x_0 is an equilibrium. Thus, for $x_0 \neq 0$ the control u_0 does not satisfy (2.47) if τ is large enough such

that $Me^{-\alpha\tau} < 1$ and we may assume that $u_0 \notin \mathcal{S}$. Hence, for every j one finds $t_j \in [0, \tau]$ with

$$c_j := \max_{t \in [0, \tau]} \left| \int_0^t u_j(s) ds \right| = \left| \int_0^{t_j} u_j(s) ds \right| > 0.$$

Consider an initial point $x_0 \in \mathcal{K}$ with $|x_0| < \min_j (c_j / (2M))$. Observe that then $|x_0| < M|x_0| < (1/2) \min_j c_j$, since $M > 1$. We claim that there is no control $u_j \in \mathcal{S}$ satisfying (2.47). In fact, one computes

$$\begin{aligned} \left| x_0 + \int_0^{t_j} u_j(s) ds \right| &\geq \left| \int_0^{t_j} u_j(s) ds \right| - |x_0| \geq c_j - \frac{c_j}{2} \\ &= \frac{c_j}{2} > e^{-\alpha t_j} \frac{c_j}{2} > Me^{-\alpha t_j} |x_0|, \end{aligned}$$

which proves the claim. \square

In contrast to the linear example (2.46), the scalar bilinear system

$$\dot{x}(t) = (1 + u(t))x(t), \quad u(t) \in \mathcal{U} = \mathbb{R}$$

can be stabilized by the single constant control $u(t) \equiv -2$. Thus, no transfer of bits is necessary. Reference [83] describes other situations where finitely many bits are sufficient. While it might be worthwhile to study bilinear control systems in this context, we follow a different path and relax the exponential stability property (2.45) by introducing a small additive term. The following simple, but basic lemma shows that then only finitely many bits are required on a finite interval. It follows by using continuous dependence on initial values and compactness.

Lemma 2.3. *Consider control system (2.44) and let \mathcal{K} be a compact subset of \mathbb{R}^n . Assume that there are constants $M > 1$ and $\alpha > 0$ such that for all $0 \neq \mathbf{x} \in \mathcal{K}$ there is a $\mathbf{u} \in \underline{\mathcal{U}}$ with*

$$\|\varphi(t, \mathbf{x}, \mathbf{u})\| < Me^{-\alpha t} \|\mathbf{x}\| \text{ for all } t \geq 0. \quad (2.48)$$

Let $\varepsilon > 0$. Then for every $\tau > 0$ there is a finite set $\mathcal{S} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset \underline{\mathcal{U}}$ such that for every $\mathbf{x} \in \mathcal{K}$ there is a $\mathbf{u}_j \in \mathcal{S}$ with

$$\|\varphi(t, \mathbf{x}, \mathbf{u}_j)\| < e^{-\alpha t} (\varepsilon + M \|\mathbf{x}\|) \text{ for all } t \in [0, \tau].$$

Stabilization Entropy. In view of the preceding lemma, we will consider weakened versions of the estimate (2.48). There are the two ways to measure the information needed for stabilization as indicated above and we begin with an entropy-like notion. Consider a compact set $\mathcal{K} \subset \mathbb{R}^n$ of initial states, and let $\alpha > 0$, $M > 1$ and $\varepsilon > 0$. For a time $\tau > 0$ we call a subset $\mathcal{S} \subset \underline{\mathcal{U}}$ a (τ, ε) -spanning set of controls if for every $\mathbf{x} \in \mathcal{K}$ there is $\mathbf{u} \in \mathcal{S}$ with

$$\|\varphi(t, \mathbf{x}, \mathbf{u})\| < e^{-\alpha t} (\varepsilon + M \|\mathbf{x}\|) \text{ for all } t \in [0, \tau]. \quad (2.49)$$

The minimal cardinality of all these sets is

$$s_{\text{stab}}(\tau, \varepsilon) := \min \{ \#\mathcal{S} : \mathcal{S} \text{ is } (\tau, \varepsilon)\text{-spanning} \}.$$

If there is no finite set \mathcal{S} of controls with this property (in particular, if there is no such set at all), we let $s_{\text{stab}}(\tau, \varepsilon) := \infty$. Lemma 2.3 shows that the exponential controllability condition (2.48) implies the existence of finite (τ, ε) -spanning sets.

We want to determine which information has to be transmitted through a digital communication channel in order to identify a control function in such a finite set \mathcal{S} . The elements can be encoded by symbols given by finite sequences of 0's and 1's in the set

$$\Sigma_k := \{ (s_0 s_1 s_2 \dots s_{k-1}) : s_i \in \{0, 1\} \text{ for } i = 0, 1, \dots, k-1 \},$$

where $k \in \mathbb{N}$ is the least integer greater than or equal to $\log_2 \#\mathcal{S}$. Thus, $\#\mathcal{S}$ is bounded above by 2^k . Equivalently, the number of bits determining an element of \mathcal{S} is $\log_2(2^k) = k$. It will be convenient to use here the natural logarithm instead of the logarithm with base 2. Now we consider what happens for time tending to infinity and then ε tending to 0 to obtain the following notion describing an entropy property for exponential stabilization.

Definition 2.5. *Let \mathcal{K} be a compact set in \mathbb{R}^n and $\alpha > 0$, $M > 1$. Then the (α, M) -stabilization entropy $h_{\text{stab}}(\alpha, M, \mathcal{K})$ is defined by*

$$h_{\text{stab}}(\alpha, M, \mathcal{K}) := \lim_{\varepsilon \searrow 0} \limsup_{\tau \rightarrow \infty} \frac{1}{\tau} \ln s_{\text{stab}}(\tau, \varepsilon).$$

In the following, we drop the argument \mathcal{K} in this and in similar notions if the choice of \mathcal{K} is clear or if its specification is irrelevant in the corresponding context. The existence of the limit for $\varepsilon \searrow 0$ is obvious, since it equals the supremum over all $\varepsilon > 0$. (The value $+\infty$ is allowed.) Furthermore, the inequality $h_{\text{stab}}(\alpha', M') \leq h_{\text{stab}}(\alpha, M)$ holds for $\alpha \geq \alpha' > 0$ and $M' \geq M > 1$.

Remark 2.3. If one would consider $\alpha = 0$, condition (2.49) implies that every trajectory starting in \mathcal{K} remains in the ball around the origin with radius $\varepsilon + M \max_{x \in \mathcal{K}} \|x\|$.

Minimal Bit Rate. A second way of counting bits is the following. Consider a set of control functions defined on $[0, \infty)$ which allow us to steer the system asymptotically to the equilibrium $\mathbf{x}^* = \mathbf{0}$ satisfying the following conditions. Let $M > 1$, $\alpha > 0$, $\varepsilon > 0$ and let γ be a decreasing function on $[0, \infty)$ with $\gamma(0) = \varepsilon$ and $\lim_{t \rightarrow \infty} \gamma(t) = 0$. For brevity, we call γ an \mathcal{L}_ε -function (note that continuity of γ is not required.)

Let $\mathcal{R}(\gamma, \varepsilon) \subset \underline{\mathcal{U}}$ be a set of control functions such that for every $\mathbf{x} \in \mathcal{K}$ there is $\mathbf{u} \in \mathcal{R}(\gamma, \varepsilon)$ with

$$\|\varphi(t, \mathbf{x}, \mathbf{u})\| < \gamma(t) + M e^{-\alpha t} \|\mathbf{x}\| \text{ for all } t \geq 0. \quad (2.50)$$

Then $\mathcal{R}(\gamma, \varepsilon)$ is called (γ, ε) -stabilizing for \mathcal{K} . Thus, in the ε -neighborhood of the equilibrium, the decay given by the exponential rate α may slow down, but still convergence holds for $t \rightarrow \infty$. Let

$$\mathcal{R}_\tau(\gamma, \varepsilon) := \{\mathbf{u}|_{[0, \tau]} : \mathbf{u} \in \mathcal{R}(\gamma, \varepsilon)\}$$

be the corresponding restrictions of the controls in $\mathcal{R}(\gamma, \varepsilon)$. Suppose that the cardinality $\#\mathcal{R}_\tau(\gamma, \varepsilon)$ is finite. The bit rate on the time interval $[0, \tau]$ is defined as $(1/\tau) \log_2 \#\mathcal{R}_\tau(\gamma, \varepsilon)$. If there is no set $\mathcal{R}(\gamma, \varepsilon)$ with (2.50) or if $\mathcal{R}_\tau(\gamma, \varepsilon)$ contains infinitely many elements, we set $\#\mathcal{R}_\tau(\gamma, \varepsilon) := \infty$. The required bit rate for stabilization using controls in $\mathcal{R}(\gamma, \varepsilon)$ is

$$b(\mathcal{R}(\gamma, \varepsilon)) := \liminf_{\tau \rightarrow \infty} \frac{1}{\tau} \log_2 \#\mathcal{R}_\tau(\gamma, \varepsilon).$$

Definition 2.6. *With the notions introduced above, the minimal bit rate for (α, M) -stabilization at $\mathbf{x}^* = \mathbf{0}$ of a compact set $\mathcal{K} \subset \mathbb{R}^n$ is*

$$b_{\text{stab}}(\alpha, \mathcal{K}, M) := \lim_{\varepsilon \searrow 0} \inf_{\gamma \in \mathcal{L}_\varepsilon} \inf_{\mathcal{R}(\gamma, \varepsilon)} b(\mathcal{R}(\gamma, \varepsilon)),$$

where the inner infimum is taken over all (γ, ε) -stabilizing sets $\mathcal{R}(\gamma, \varepsilon) \subset \underline{\mathcal{U}}$ of controls and the outer infimum is taken over all \mathcal{L}_ε -functions γ .

The limit for $\varepsilon \searrow 0$ is the supremum for $\varepsilon > 0$. Note also the inequality $b_{\text{stab}}(\alpha', M') \leq b_{\text{stab}}(\alpha, M)$ for $\alpha \geq \alpha' > 0$ and $M' \geq M > 1$.

An example for an allowed \mathcal{L}_ε -function is $\gamma(t) = \varepsilon e^{-\alpha t}$, $t \geq 0$. However, for this function, we cannot prove our main result for the stabilization bit rate (Theorem 2.20) showing that the stabilization entropy provides an upper bound. Also, this theorem will only give a result for the limit inferior for $\tau \rightarrow \infty$ (not for the limit superior). Consideration of the limit inferior may be justified by the fact that we are interested in the minimal bit rate for $\tau \rightarrow \infty$, hence the times may be chosen appropriately.

Comparison of Stabilization Entropy and Minimal Bit Rate. The stabilization entropy indicates how much the number of required control functions increases, when time increases. Here minimization is performed on each interval $[0, \tau]$ separately. If one wants to enlarge the time interval where the exponential decay holds, one may have to consider controls which, when restricted to the smaller interval, are different from the earlier ones. This is in contrast to minimal bit rates, where restrictions to $[0, \tau]$ are considered for control functions defined on $[0, \infty)$. Thus, while stabilization entropy certainly merits its own interest, the minimal bit rate might appear more appealing from this point of view.

The difference between these two concepts can also be seen by looking at them from a quantization point of view. Let $\mathcal{R}(\gamma, \varepsilon)$ be a (γ, ε) -stabilizing set such that for every $\tau > 0$ the set $\mathcal{R}_\tau(\gamma, \varepsilon)$ of restrictions to $[0, \tau]$ is finite. Then define for every $\mathbf{u} \in \mathcal{R}_T(\gamma, \varepsilon)$

$$\mathcal{K}(\mathbf{u}, \tau) := \{ \mathbf{x} \in \mathcal{K} : \|\varphi(t, \mathbf{x}, \mathbf{u})\| < \gamma(t) + Me^{-\alpha t} \|\mathbf{x}\| \text{ for all } t \in [0, \tau] \}.$$

The sets $\mathcal{K}(\mathbf{u}, \tau)$ form an open cover of \mathcal{K} which may be viewed as a finite quantization. For $\tau' > \tau$, the same construction for $\mathcal{R}_{\tau'}(\gamma, \varepsilon)$ again yields a finite quantization of \mathcal{K} which is obtained by refining the quantization at time τ , since both are obtained by restrictions of controls in $\mathcal{R}(\gamma, \varepsilon)$. In contrast, the quantization for $\tau' > \tau$ obtained by a (τ', ε) -spanning set of controls used for defining the entropy is not related to the quantization associated with a (τ, ε) -spanning set.

Results on the Stabilization Entropy. The next theorem gives upper and lower bounds for the stabilization entropy, which is easier to analyze than the minimal bit rate.

Theorem 2.18. *Consider control system (2.44) and let $\mathbf{x}^* = \mathbf{0}$ be an equilibrium with $\mathbf{0} = \mathbf{f}(\mathbf{0}, \mathbf{0})$, $\mathbf{0} \in \mathcal{U}$. Suppose that the derivative $\mathbf{f}_x(\mathbf{x}, \mathbf{u})$ of \mathbf{f} with respect to \mathbf{x} exists and is continuous in (\mathbf{x}, \mathbf{u}) . Let $\mathcal{K} \subset \mathbb{R}^n$ be a compact neighborhood of the origin and denote $\kappa := \max_{\mathbf{x} \in \mathcal{K}} \|\mathbf{x}\|$.*

1. *Assume further that $\operatorname{div}_x \mathbf{f}(\mathbf{x}, \mathbf{u})$ is bounded below for $\|\mathbf{x}\| \leq \kappa + 1$ and $\mathbf{u} \in \mathcal{U}$. Then for $\alpha > 0$, $M > 1$ the stabilization entropy of \mathcal{K} satisfies the estimate*

$$h_{\text{stab}}(\alpha, M) \geq \alpha d + \min_{\mathbf{u} \in \mathcal{U}} \operatorname{div}_x \mathbf{f}(\mathbf{0}, \mathbf{u}).$$

2. *Assume that for every $\mathbf{0} \neq \mathbf{x} \in \mathcal{K}$ there is a control $\mathbf{u} \in \underline{\mathcal{U}}$ with*

$$\|\varphi(t, \mathbf{x}, \mathbf{u})\| < Me^{-\alpha t} \|\mathbf{x}\| \text{ for all } t \geq 0$$

and that there is a Lipschitz constant $L > 0$ such that for all $\mathbf{x}_1, \mathbf{x}_2$ in an open set containing \mathcal{K}

$$\|\mathbf{f}(\mathbf{x}_1, \mathbf{u}) - \mathbf{f}(\mathbf{x}_2, \mathbf{u})\| \leq L \|\mathbf{x}_1 - \mathbf{x}_2\| \text{ for all } \mathbf{u} \in \mathcal{U}.$$

Then the stabilization entropy of \mathcal{K} satisfies

$$h_{\text{stab}}(\alpha, M) \leq (L + \alpha)d.$$

Next we consider the stabilization entropy for linear control systems in \mathbb{R}^n of the form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad \mathbf{u} \in \underline{\mathcal{U}}, \quad (2.51)$$

with matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times m}$ and control range $\mathcal{U} \subset \mathbb{R}^m$ containing the origin.

Theorem 2.19. *Consider a linear control system of the form (2.51) with $\mathbf{0} \in \mathcal{U}$. Assume that there are $\alpha > 0$, $M > 1$ such that for every initial value $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ there is a control $\mathbf{u} \in \underline{\mathcal{U}}$ with*

$$\|\varphi(t, \mathbf{x}, \mathbf{u})\| < Me^{-\alpha t} \|\mathbf{x}\| \text{ for all } t \geq 0.$$

For every compact neighborhood \mathcal{K} of the origin in \mathbb{R}^n the (α, M) -stabilization entropy of system (2.51) satisfies

$$h_{\text{stab}}(\alpha, M, \mathcal{K}) = \sum_{\text{Real}(\lambda) > -\alpha} (\alpha + \text{Real}(\lambda)); \tag{2.52}$$

here summation is over all eigenvalues λ of \mathbf{A} , counted according to their algebraic multiplicities, with $\text{Real}(\lambda_i) > -\alpha$. In particular, $h_{\text{stab}}(\alpha, M, \mathcal{K})$ is independent of \mathcal{K} .

The next theorem establishes the announced relation between the stabilization bit rate and the stabilization entropy.

Theorem 2.20. *Consider a control system of the form (2.44) and suppose that $\mathcal{K} \subset \mathbb{R}^d$ is a compact neighborhood of the equilibrium $\mathbf{x}^* = \mathbf{0}$. Assume that there are constants $M > 1$ and $\alpha^* > 0$ such that for all $\mathbf{0} \neq \mathbf{x} \in \mathcal{K}$ there is $\mathbf{u} \in \underline{\mathcal{U}}$ with*

$$\|\varphi(t, \mathbf{x}, \mathbf{u})\| < Me^{-\alpha^* t} \|\mathbf{x}\| \text{ for all } t \geq 0. \tag{2.53}$$

Then for $\alpha \in (0, \alpha^*)$ the stabilization bit rate and the stabilization entropy satisfy

$$b_{\text{stab}}(\alpha, M) \leq \log_2 e \cdot h_{\text{stab}}(\alpha^*, M) < \infty.$$

For linear control systems, one gets an easy corollary.

Corollary 2.5. *Suppose that the linear control system (2.51) satisfies the assumptions of Theorem 2.20. Then for all $\alpha \in (0, \alpha^*)$ the stabilization bit rate satisfies*

$$\begin{aligned} (\log_2 e)^{-1} \cdot b_{\text{stab}}(\alpha, M) &\leq h_{\text{stab}}(\alpha, M) \\ &= \sum_{\text{Real}(\lambda) > -\alpha} (\alpha + \text{Real}(\lambda)) \end{aligned}$$

and

$$\begin{aligned} (\log_2 e)^{-1} \cdot \inf_{\alpha > 0} b_{\text{stab}}(\alpha, M) &= \inf_{\alpha > 0} h_{\text{stab}}(\alpha, M) \\ &= \sum_{\text{Real}(\lambda) > 0} \text{Real}(\lambda). \end{aligned}$$

For smooth nonlinear control systems and arbitrarily small control range, we conjecture that a formula analogous to (2.52) holds, now for the Jacobian at the equilibrium (cf. [270], Theorem 3, for such a claim in the context of local uniform asymptotic stabilization).

2.4 Dynamic Quantization for Feedback Stabilization with Delayed Data-Rate Limited Communication

2.4.1 Problem Statement

In this section we consider networked control systems over a digital channel with limited data rate. An inherent property of such channels is that data needs to be quantized. Here we employ dynamic quantization for stabilization. The term dynamic quantization refers to an approach where the size of the quantization region as well as the center of the region change in time according to the input received from system measurements. This information from the system allows to estimate the region in which quantization is required. With this approach limited bandwidth can be employed to communicate increasingly accurate information, which will eventually be sufficient to achieve stabilization.

The class of systems under consideration is

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{k}(\mathbf{x}(t) + \mathbf{e}(t))), \quad (2.54)$$

where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$. The aim is to stabilize the system at the equilibrium $(\mathbf{x}^*, \mathbf{u}^*) = (\mathbf{0}, \mathbf{0})$. We assume that a controller \mathbf{k} has been designed which renders system (2.54) input-to-state stable (ISS) with respect to the quantization error \mathbf{e} . For a discussion of the ISS property, see Section 5.5. We assume that \mathbf{f} is sufficiently smooth to guarantee existence and uniqueness of solutions. In particular \mathbf{f} is Lipschitz in the first component with Lipschitz constant L .

We will assume that the state of the system can be measured and that this measured information needs to be transmitted to the controller via a digital channel. We will present an encoder/decoder scheme that will result in a stable system, given the information on \mathbf{f} and \mathbf{k} in (2.54). A particular problem in networked control systems are delays and packet loss. Here we describe how dynamic quantization can be implemented to cope with these phenomena.

2.4.2 Dynamic Quantization

In a first pass we ignore delay and packet loss and explain the idea behind dynamic quantization with the help of Fig. 2.3. In our setup communication attempts occur at the time instances $t_0 < t_1 < t_2 < \dots$. At every communication instant the quantization region is a hypercube that is determined by its center \mathbf{C} and the edge length ℓ . This hypercube is partitioned into N^n smaller hypercubes of equal size; recall that n is the dimension of the state space. The smaller hypercubes are called subregions. At the communication event t_k the encoder sends the number of the subregion in which the state $\mathbf{x}(t_k)$ of the system lies. At the time of transmission the estimate of the state $\hat{\mathbf{x}}$ is the center of the subregion. In this case \mathbf{e} represents the error between

the actual value of the system and the midpoint of the subregion, which is used as an approximation thereof.

The problem that has to be solved by the scheme lies in the fact that encoder and decoder have to agree on center and length of the subregion that is determined by the transmitted code word, so that the transmitted information can be used to compute the control value $\mathbf{k}(\mathbf{x} + \mathbf{e})$.

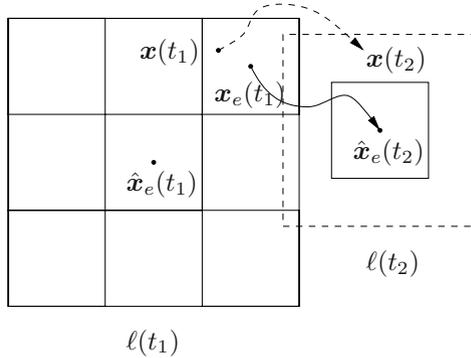


Fig. 2.3 Dynamic quantization

At a time t_k a certain quantization region, i.e. its center and edge length, is determined by the previous dynamics and it has to be ensured that the state $\mathbf{x}(t_k)$ is within the quantization region at time t_k . Given that encoder and decoder agree on the value $\hat{\mathbf{x}}_e(t_k)$, they perform identical numerical estimates of the evolution of the system

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{k}(\mathbf{x}(t))), \quad \mathbf{x}(t_k) = \hat{\mathbf{x}}(t_k),$$

and the solution $\hat{\mathbf{x}}(t_{k+1}) := \mathbf{x}(t_{k+1}; t_k, \mathbf{x}(t_k))$ of this initial value problem at time t_{k+1} is the center of the quantization region at time t_{k+1} .

In other words, the sensor determines the actual subregion in which the state lies and calculates the center $\mathbf{x}_e(t_k)$. We will refer to this process as encoding. The decoder on the other end of the channel has a copy of these values, i.e. the decoder knows the center and length of the region and the values n and N . Thus if we transmit the number of the subregion in which the state lies, the decoder is able to reconstruct the value $\mathbf{x}_e(t_k)$.

If both encoder and decoder let the center \mathbf{x}_e of the subregion follow the closed-loop dynamics until time t_2 , the error between the estimate \mathbf{x}_e and the state \mathbf{x} can grow by a certain factor. If we let the subregion grow by the same amount (the augmented region is the dashed box in Fig. 2.3), we are sure that the state at time t_2 is still within the subregion. This subregion becomes the new quantization region with the new center $\hat{\mathbf{x}}_e(t_2) = \mathbf{x}_e(t_2^-)$. Now we are in the same situation as previously, namely, we know a hypercube containing the state and we can now repeat the same steps.

If the quotient between the growth of the quantization region and the reduction of the error due to N is smaller than 1, the quantization error converges to 0 as the number of communication events increases.

During continuous flow of the system, the growth of the uncertainty of the state (i.e., the quantization error) may be bounded with the help of the Lipschitz constant L of the system (depicted by the dotted box in Fig. 2.3). This bound ensures that the state is at any time in some region with known size and center.

The growth of the error is countered by a reduction of the size of the quantization region by a defined number N (in Fig. 2.3: $N = 3$), which is the number of smaller regions (per dimension) within the quantization range (depicted in Fig. 2.3 as the jump from the center of the large region to the center of the smaller subregion in which the actual state lies).

Delays and Packet Loss. As sketched above it is important that encoder and decoder agree on certain values. This is easily achieved, if there is no delay in the channel, because the decoder can copy the behavior of the encoder exactly.

If delays and packet loss are present then the information available at the decoder and encoder side may be different. To account for this fact, we will in the sequel use variables available at the encoder by $\mathbf{x}_e, \hat{\mathbf{x}}_e$, etc., and those available at the decoder by $\mathbf{x}_d, \hat{\mathbf{x}}_d$, etc. The scheme described in the following relies crucially on time stamps and on acknowledgments (acks). The time stamps provide the information at which time certain information was available at the other side and acks provide the encoder with the information that data has been received by the decoder. Note that we assume acks to be time-stamped as well.

For the communication channel we assume the following properties:

Assumption 2.1

1. All packets are time stamped with the current time they are sent.
2. Only packets sent from encoder to decoder are lost.
3. There exists a minimal delay from encoder to decoder, given by τ_{\min} , i.e., $t_k^* - t_k \geq \tau_{\min}$ and $t_k - t_{k-1}^* \geq \tau_{\min}$.
4. The channel is able to transmit packets containing a value from a set of N^n (N odd) discrete values within τ_{\min} units of time.
5. If τ_{\max} time elapses without receiving an ack, the packet sent last time is considered lost and a new packet will be sent.

These assumptions have the following consequences:

- According to Part 1 we have to send the actual time together with the encoded state information. It is not reasonable to be able to transmit the state information quantized and the time information not. For the sake of simplicity we omit details on time quantization but refer to [326] for a discussion.

- Part 2 is a major restriction on the channel used. As the acks are much smaller than the state information the decoder could send many acks to ensure that at least one arrives at the encoder. Without this assumption we could not guarantee that the encoder and the decoder agree on their states.
- Part 3 is in general not a restrictive one. In every real communication channel such a minimal delay exists.
- Part 4 states that the bandwidth of the channel B must be large enough to transmit the state information within τ_{\min} units of time. For instance, if binary encoding is used we require

$$B \geq \frac{n \log_2 N}{\tau_{\min}}. \quad (2.55)$$

If this condition is not met, the decoder could introduce an artificial delay by waiting to ensure that τ_{\min} is large enough to fulfill the bandwidth constraint.

The choice for N to be odd guarantees that the center of the quantization region lies in the interior of one of the subregions. Although this is not needed in general, it eases the presentation.

- By Part 5, the values τ_{\min} and τ_{\max} may be regarded as design parameters. While choosing τ_{\min} too small can violate the bandwidth constraint, a larger value can degrade the performance of the overall system. Similar statements hold for τ_{\max} . If τ_{\max} is chosen too small, no ack will arrive at the encoder before a new packet will be sent and again choosing τ_{\max} too large may have a negative effect on the performance.

Further assumptions that will be of importance are the following:

Assumption 2.2. *The clocks of encoder and decoder are synchronized and the time $t_1 = 0$ when the encoder sends the first packet is known by the encoder and the decoder.*

Assumption 2.3. *Both the encoder and the decoder know the same bound of the initial state of the system (i.e., encoder and decoder agree on a constant $R > 0$ such that $|x(0)| \leq R$).*

We consider TCP-like packet-based transmissions over a noiseless, error-free channel with delay and packet loss.

The encoder encodes the state and sends a symbol from a finite alphabet to the decoder together with the time when the state was encoded (time stamping). As soon as a packet arrives, the decoder reconstructs the encoded state and sends an acknowledgment back to the encoder. If this ack arrives at the encoder or a predefined time elapses without receiving one, it repeats the encoding. Denote by t_k the k -th time instance the encoder received an ack.

The time when the k -th information sent by the encoder is received by the decoder is denoted by t_k^* . Note that we assume that there is no time delay between the arrival of an information and the sending of the next packet, i.e. t_k and t_k^* are also the time instances when the encoder sends information and the decoder sends an ack, respectively.

We now detail the complete description of encoder and decoder. A fundamental element of the encoder is the map φ , which yields the subregion in which the state \mathbf{x} lies as a function of the center of the quantization region \mathbf{x}_e and the edge length ℓ . It is given by

$$\varphi(\mathbf{x}_e, \mathbf{x}, \ell) = \left\lfloor \frac{N}{\ell} (\mathbf{x} - \mathbf{x}_e) + \frac{1}{2} [1 \dots 1]^\top \right\rfloor, \quad (2.56)$$

where the floor function $\lfloor \cdot \rfloor$ is applied in each component.

Encoder and the decoder are initialized trivially by

$$\begin{aligned} k = 1, t_0 = 0, t_0^* = 0, t_1 = 0, t_1^* = 0 \text{ and } \hat{\mathbf{x}}_d(0) = \hat{\mathbf{x}}_e(0) = 0 \\ \mathbf{x}_e(0^-) = \mathbf{x}_d(0^-) = 0 \text{ and } \ell_e(0^-) = \ell_d(0^-) = 2R. \end{aligned}$$

Encoder Dynamics. For the encoder three cases have to be treated: (i) times at which an acknowledgment of a previous packet is received by the encoder, (ii) times at which the maximal admissible time span τ_{\max} passes without the receipt of an acknowledgment, at which time the previous data will be considered lost, and (iii) intermediate times. The required actions for each of these cases are given as follows. In both cases ((i) and (ii)) t_s and $s(t_s)$ are auxiliary variables, describing the data payload of the packets sent from encoder to decoder.

(i) Every time an ack arrives at the encoder ($t = t_k$) the following operations are invoked:

$$t_s := t \quad (2.57)$$

$$\ell_e(t_k) := \ell_e(t_{k-1}) e^{L(t_k - t_{k-1})} / N \quad (2.58)$$

$$\begin{aligned} \mathbf{x}_e(t_k^-) := \mathbf{x}_e(t_{k-1}) + \int_{t_{k-1}}^{t_{k-1}^*} \mathbf{f}(\mathbf{x}_e(\tau), \mathbf{k}(\hat{\mathbf{x}}_e(\tau))) \, d\tau \\ + \int_{t_{k-1}^*}^{t_k} \mathbf{f}(\mathbf{x}_e(\tau), \mathbf{k}(\mathbf{x}_e(\tau))) \, d\tau \end{aligned} \quad (2.59)$$

$$\mathbf{s}(t_s) := \varphi(\mathbf{x}_e(t_k^-), \mathbf{x}(t_k), \ell_e(t_k)) \quad (2.60)$$

$$\mathbf{x}_e(t_k) := \mathbf{x}_e(t_k^-) + \mathbf{s}(t_s) \frac{\ell_e(t_k)}{N} \quad (2.61)$$

$$\hat{\mathbf{x}}_e(t_k) = \mathbf{x}_e(t_k^-). \quad (2.62)$$

Accordingly, the encoder updates the length of the quantization region according to the growth of the error on the last interval (2.58). The center

of the quantization region is updated via (2.59). Both integrals are needed to account for the change in the control action on the decoder side at time t_{k-1}^* which is known to the encoder by the time stamp of the ack. The subregion in which the state lies is determined by (2.60). This information will be sent to the decoder together with the actual time (2.57). The jump from the center to the subregion is done by Eq. (2.61). The value of the old quantization region is copied by (2.62) in order to have access to this information in case that the current communication event fails.

- (ii) If τ_{\max} time instants elapse without receipt of an ack the following simplified actions need to be taken:

$$t_s := t \quad (2.63)$$

$$\ell_e(t) := \ell_e(t_k) e^{L(t-t_k)} \quad (2.64)$$

$$\mathbf{x}_e(t^-) := \hat{\mathbf{x}}_e(t) \quad (2.65)$$

$$\mathbf{s}(t_s) := \varphi(\mathbf{x}_e(t^-), \mathbf{x}(t), \ell_e(t)) \quad (2.66)$$

$$\mathbf{x}_e(t) := \mathbf{x}_e(t^-) + \mathbf{s}(t_s) \frac{\ell_e(t)}{N}. \quad (2.67)$$

Accordingly, the packet sent last time is considered lost and a new one is sent. Similarly to the case of no loss, the encoder updates the length of the quantization region (2.64). Note that there is no division by N , as the reduction effect provided by the transmitted information is not present. Equation (2.65) cancels the jump from the center to the subregion performed in the last encoding step. Equations (2.66) and (2.67) follow the same reasoning as in the case of no loss.

- (iii) In between transmission events only the center of the quantization region has to be updated by

$$\dot{\hat{\mathbf{x}}}_e(t) = \mathbf{f}(\hat{\mathbf{x}}_e(t), \mathbf{k}(\hat{\mathbf{x}}_e(t))). \quad (2.68)$$

We need (2.68) to know the trajectory which will be used to close the loop on the decoder side. It is also needed to treat the case of packet loss in (2.65).

Decoder Dynamics. For the decoder only two cases have to be distinguished: (i) the arrival of information at times t_k^* and (ii) the times in between arrivals.

- (i) Every time a packet arrives at the decoder ($t = t_k^*$) the following operations are carried out:

$$\ell_d(t_s) := \ell_d(t_{k-1}) e^{L(t_s-t_{k-1})} / N \quad (2.69)$$

$$\mathbf{x}_d(t_s^-) := \hat{\mathbf{x}}_d(t_{k-1}^*) + \int_{t_{k-1}^*}^{t_s} \mathbf{f}(\mathbf{x}_d(\tau), \mathbf{k}(\mathbf{x}_d(\tau))) d\tau \quad (2.70)$$

$$\mathbf{x}_d(t_s) := \mathbf{x}_d(t_s^-) + s(t_s) \frac{\ell_d(t_s)}{N} \quad (2.71)$$

$$\hat{\mathbf{x}}_d(t_k^*) := \mathbf{x}_d(t_s) + \int_{t_s}^{t_k^*} \mathbf{f}(\mathbf{x}_d(\tau), \mathbf{k}(\hat{\mathbf{x}}_d(\tau))) d\tau. \quad (2.72)$$

(ii) Otherwise

$$\dot{\hat{\mathbf{x}}}_d(t) = \mathbf{f}(\hat{\mathbf{x}}_d(t), \mathbf{k}(\hat{\mathbf{x}}_d(t))). \quad (2.73)$$

The decoder copies the behavior of the encoder with the help of (2.69)-(2.71). Equation (2.72) compensates for the delay between encoder and decoder.

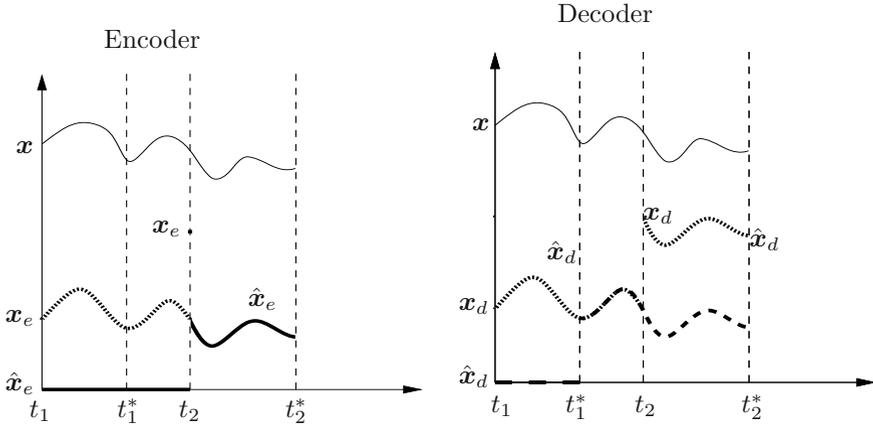


Fig. 2.4 Snapshot of the different trajectories at time t_2^*

A sketch of the evolution of the different trajectories is depicted in Fig. 2.4. The overall effect of this encoder/decoder scheme is that if at time t the state lies within the quantization region, the error between the state and the estimate shrinks by N because of the jump from the center of the region to the center of a subregion. Hence we guarantee that

$$|\mathbf{x}(t) - \mathbf{x}_e(t^-)| \leq \frac{\ell}{2} \Rightarrow |\mathbf{x}(t) - \mathbf{x}_e(t)| \leq \frac{\ell}{2N} \quad (2.74)$$

holds, which can be seen from (2.56) and (2.61). Overall, the evolution of the closed-loop system is given by

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{k}(\hat{\mathbf{x}}_d(t))). \quad (2.75)$$

For this system it is possible to show the following stabilization result.

Theorem 2.21. *Consider a control system with a time-stamped dynamic quantization encoder/decoder scheme. If*

$$N > e^{L\tau^*},$$

where τ^ is the long time average between successful transmissions, then the closed-loop system is asymptotically stable.*

It is worth mentioning that the quantity N is directly related to the bandwidth of the communication channel.

2.4.3 Markovian Communication Models

One shortcoming of the results presented so far lies in the deterministic nature of the model of the communication channel. For many practical examples the appropriate models for the dynamic behavior are of stochastic nature. This can be seen for example for TCP and for the wireless case, where packet loss frequently is due to the influence of an environment which can be suitably modeled as random.

In this section we derive conditions for the stabilization of system (2.54) under the condition that the communication channel can be described by a (time-homogeneous) Markov process.

In the following we assume as given a communication channel in which external perturbations such as average load of other users is stationary. We consider a Markov chain $\{X(k)\}_{k \in \mathbb{N}}$, where the state X lies in a state space of the chain S . This state space has to be specified with the concrete situation in mind. As we only send information at discrete time instances, a Markov process which is discrete in time is justified.

We assume as given two continuous maps describing the communication, namely

$$T : S \rightarrow [T_{\min}, \infty), \quad g : S \rightarrow \mathbb{N}, \quad (2.76)$$

where $T(X)$ denotes the length of the next communication interval depending on the state X of the channel and $g(X)$ denotes the number of bits that can be sent in that interval. Thus if communication starts at a time $t_1 \in \mathbb{R}$ and $l := n \log_2 N + b$ bits have to be sent, where b is the acknowledgment size in bits, we define a stopping time of the Markov chain by

$$k_1^* = \min \left\{ k \left| \sum_{j=0}^k g(X(j)) \geq l \text{ or } \sum_{j=0}^k T(X(j)) \geq \tau_{\max} \right. \right\} \quad (2.77)$$

and until k_1^*

$$\sum_{j=0}^{k_1^*} T(X(j)) =: \tau(1)$$

units of time elapse. If $\tau(1) \geq \tau_{\max}$ (with $\tau_{\max} > 0$ as a design parameter), we consider the information to be lost. To ensure the Markovian property of our description, we assume that $X(k)$ has the strong Markov property, i.e. the evolution of the process only depends on the state of the chain at the stopping time k^* . The next time we want to send information, i.e. at time $t_1 + \tau(1)$ we define the next stopping time k_2^* by

$$k_2^* = \min \left\{ k \left| \sum_{j=k_1^*}^k g(X(j)) \geq l \text{ or } \sum_{j=k_1^*}^k T(X(j)) \geq \tau_{\max} \right. \right\}$$

and the duration from k_1^* until k_2^* by $\sum_{j=k_1^*}^{k_2^*} T(X(j)) =: \tau(2)$. If we repeat this procedure a sequence of time instances is given by $\{\tau(j)\}_{j \in \mathbb{N}}$.

We now assume ergodicity of the Markov chain, which ensures that almost surely

$$\tau_M^* := \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{j=0}^k \tau(j) \mathbb{1}\{\tau(j) < \tau_{\max}\} \quad (2.78)$$

exists, where $\mathbb{1}$ denotes the indicator function.

With the above considerations we are able to state a stochastic version of Theorem 2.21, where the long time average is given by τ_M^* instead of τ^* .

Theorem 2.22. *Consider a control system with a time-stamped dynamic quantization encoder/decoder scheme. Assume that the communication channel is given by a strongly ergodic Markov process X , and maps T and g given in (2.76) and with stopping times as in (2.77). If*

$$N > e^{L\tau_M^*}, \quad (2.79)$$

where τ_M^ is given by (2.78), then the closed loop system is asymptotically stable with probability 1.*

Again this result relates the error growth due to the systems dynamics given by the Lipschitz constant L with the information that can be sent on average using the communication channel and the refinement of the quantization.

As this point we see the relation to the results in Section 2.3: In general, with the information provided by entropy considerations finer bounds on the necessary information can be obtained. However, information on the entropy of the process is frequently hard to obtain, so that in the results presented in this section we have settled for a coarse estimate as given by the Lipschitz constant.

Example 2.3 *Dynamic quantization for the inverted pendulum*

We discuss the results of the previous section for the example of a pendulum on a cart that is stabilized in the upright position. The dynamic equations are given by

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{mg}{M} & 0 & 0 \\ 0 & -\frac{g(m+M)}{lM} & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{M} \\ -\frac{1}{lM} \end{bmatrix} u \tag{2.80}$$

where x_1, x_3 are angle and angular velocity of the pendulum and x_2 resp. x_4 denote the position and velocity of the cart.

In the simulations the values $m = 0.329, M = 3.2, l = 0.44, g = 9.81$ have been used. The shaded grey area in Figs. 2.5 and 2.6 represents the quantization region as a function of time. Between communication events this region grows according to the Lipschitz constant of the system. It shrinks if new information is received.

In the simulation a stochastic model has been used for the generation of delays, but so that the conditions of Theorem 2.22 are satisfied. It can be seen that the quantization region quickly approximates the trajectory, so that the information at the controller becomes increasingly accurate. The discontinuous lines in both figures represent the value that is used for controlling the system. Because of the delay in the system the discontinuities of these control values do not occur at the communication times; the latter are indicated by dots on the horizontal axis.

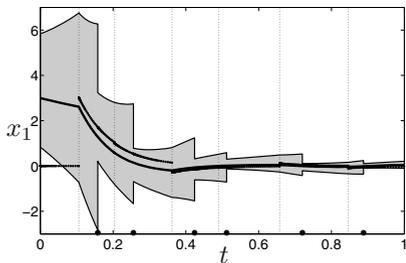


Fig. 2.5 Trajectory of angle (x_1)

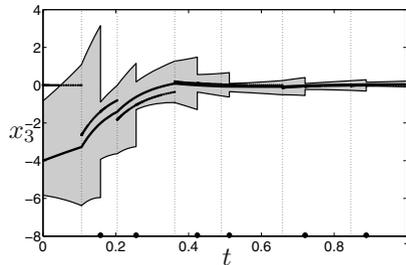


Fig. 2.6 Trajectory of angular velocity (x_3)

The error between system state and estimate used by the controller is shown in Figs. 2.7 and 2.8. □

Bibliographical Notes

The approach of dynamic quantization we are using here, was introduced by [41] and has been extended to the nonlinear case in [231]. However, the notion of non-static quantization is not new. It was first mentioned within the control community in [359] or even earlier within the communication community (c.f. [137]).

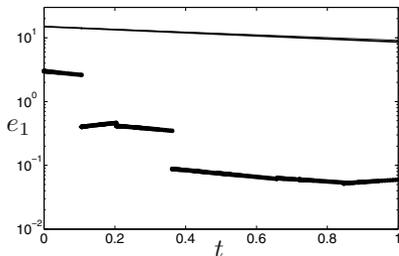


Fig. 2.7 Error e_1

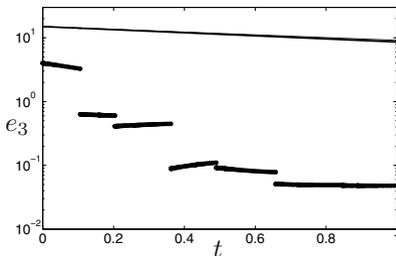


Fig. 2.8 Error e_3

One of the main problems in this area is that the amount of information that can be sent from the sensing device to the controller is limited and can be corrupted in several ways. This type of problems has received considerable attention, see [84–86, 163, 178, 276, 326] as a starting point. First steps in this area considered communication constraints such as limited bandwidth or data rate, but issues as delays and packet loss were not treated. Also the communication channel was treated as static [86, 178, 276], whereas many realizations of communication channels use protocols which define internal dynamics of the channel. Examples of this are given by TCP and certain wireless protocols.

The results presented here are amenable to more dynamic communication channels. They were first published in [325, 326].